HealthCyberMap: Mapping the Health Cyberspace Using Hypermedia GIS and Clinical Codes

Maged Nabih Kamel Boulos

MBBCh, MSc (Derm & Vener), MSc (Medical Informatics)

A Thesis Submitted for the Degree of Philosophy Doctor (PhD)

City University, London, UK

Centre for Measurement and Information in Medicine (MIM),
School of Informatics

December 2002

Contents

List of Tables

List of Illustrations

Acknowledgements

The author wishes to acknowledge and thank all those who have assisted him in the development of this work. In particular, I wish to thank my supervisors, Dr. Abdul V. Roudsari, Senior Lecturer in Medical Informatics, MIM Centre, School of Informatics, City University, and Professor Ewart R. Carson, Director of MIM Centre, School of Informatics, City University, for their constant support and constructive feedback; without their help this thesis would never have been completed.

I would also like to thank Dr. Natalya F. Noy and the Protégé Team in Stanford Medical Informatics laboratory at Stanford University, US, for providing the *UMLS tab* for Protégé-2000 free of charge for the purpose of this research, and also for their critical comments and support[†]. Special thanks are also extended to Dr. Christopher Austin president of GeoHealth, Inc., US, who supplied *BodyViewer v2.1 Extension Version (ICD-9)* free of charge for this research. I also extend my thanks to Dr. David Hunt president of Yaki Technologies, US, for providing me with their proprietary *ICD-9-CM search technology* as a research grant to build HealthCyberMap Medical Semantic Subject Search Engine.

[†] At the time this research was done, a licence for the UMLS tab was required and the plugin was not available for public download; the Open Source edition of this tab was not yet released.

Declaration

Peer-reviewed Publications Arising out of this Project

Journal Papers

1. **Kamel Boulos MN**, Roudsari AV, Carson ER. Towards a semantic medical Web: HealthCyberMap's tool for building an RDF metadata base of health information resources based on the Qualified Dublin Core Metadata Set. *Medical Science Monitor*. 2002 Jul;8(7):MT124-36. Available from: <http://www.medscimonit.com/pub/vol_8/no_7/2615.pdf>
2. **Kamel Boulos MN**, Roudsari AV, Carson ER. A simple method for serving Web hypermaps with dynamic database drill-down. *International Journal of Health Geographics*. 2002 Aug;1:1. Available from: <http://www.ij-healthgeographics.com/content/pdf/1476-072X-1-1.pdf>
3. **Kamel Boulos MN**, Roudsari AV, Carson ER. HealthCyberMap: A Semantic Visual Browser of Medical Internet Resources Based on Clinical Codes and the Human Body Metaphor. *Health Information and Libraries Journal*. 2002 Dec;19(4) (Accepted 17 May 2002 - in press)
4. **Kamel Boulos MN**, Roudsari AV, Carson ER. A Dynamic Problem to Knowledge Linking Semantic Web Service Based on Clinical Codes. *Medical Informatics & The Internet in Medicine*. 2002;27 (Accepted 15 April 2002 - in press)

Related Journal Papers (Also by the Author):

5. **Kamel Boulos MN**, Roudsari AV, Carson ER. Health Geomatics: An Enabling Suite of Technologies in Health and Healthcare (Methodolical Review). *Journal of Biomedical Informatics*. 2001 Jun;34(3):195-219. Available from: <http://www.idealibrary.com/links/doi/10.1006/jbin.2001.1015>
6. **Kamel Boulos MN**, Roudsari AV, Gordon C, Muir Gray JA. The Use of Quality Benchmarking in Assessing Web Resources for the Dermatology Virtual Branch Library of the National electronic Library for Health (NeLH). *Journal of Medical Internet Research*. 2001 Mar;3(1):e5. Available from: <http://www.jmir.org/2001/1/e5/>

Conference Papers/ Abstracts

7. **Kamel Boulos MN**, Roudsari AV, Carson ER. HealthCyberMap's Dublin Core Ontology in Protégé-2000. Presented at the Fifth International Protégé Workshop, SCHIN, July 2001, Newcastle, UK. Available from: <http://protege.stanford.edu/ontologies/dublincore/hcm_dc_in_protege.zip> (ontology) and <http://www.schin.ncl.ac.uk/protege2001/presentations/newcastle_20010718.ppt>
8. **Kamel Boulos MN**, Roudsari AV, Carson ER. HealthCyberMap: Mapping the Health Cyberspace Using Hypermedia GIS and Clinical Codes. In: *Proceedings of the First European Conference for Geographic Information Sciences in Public Health, September 2001, Sheffield, UK*. Available from: <http://gis.sheffield.ac.uk/newconferencepages/abstracts.htm#other3mb>
9. **Kamel Boulos MN**, Roudsari AV, Carson ER. A Dynamic Problem-Knowledge Coupling Semantic Web Service. In: Della Mea V, Beltrami CA, Woodall J, Arvanitis TN (editors). *Proceedings of the 6th World Congress on the Internet in Medicine, December 2001, Udine, Italy/ Technology and Healthcare*. 2001;9(6):477-479. Amsterdam: IOS Press. Available from: <http://mednet2001.drmm.uniud.it/proceedings/paper.php?id=44>

There are other submitted journal papers by the author (still under peer-reviewing). Please refer to <http://healthcybermap.semanticweb.org/publications/> for the latest list of publications arising out of this project.

1 Abstract

HealthCyberMap (<http://healthcybermap.semanticweb.org>) is a Semantic Web service for healthcare professionals and librarians, patients and the public in general that aims at mapping parts of medical/ health information resources in cyberspace in novel ways to improve their retrieval and navigation. The Semantic Web (<http://www.w3.org/2001/sw/> and <http://www.semanticweb.org>) aims to be the next-generation World Wide Web by giving machine-readable semantics and context to the currently presentation-based Web pages.

HealthCyberMap features an unconventional use of GIS (Geographic Information Systems) to map conceptual spaces occupied by collections of medical/ health information resources. Besides mapping the semantic and non-geographical aspects of these resources using suitable spatial metaphors, HealthCyberMap also collects and maps the geographical provenance of these resources. Some of HealthCyberMap Web interfaces are visual (maps for browsing resources by clinical/ health topic, by provenance and by type), while others are textual (multilingual interfaces for browsing resources by language, and a directory of topical resource categories, besides HealthCyberMap Semantic Subject Search Engine that goes beyond conventional free-text and keyword-based search engines, and supports synonyms, disease variants, subtypes, as well as some semantic relationships between terms).

HealthCyberMap adopts a clinical metadata framework built upon a clinical coding scheme (vocabulary or ontology—ICD-9-CM[‡] clinical classification in the current pilot service). Clinical coding schemes serve as a reliable common backbone for topical resource indexing, automated topical classification, topical visualisation and navigation of coded resource pools (using suitable metaphors), and enhanced information retrieval and linking. A resource metadata base based on Dublin Core metadata set with HealthCyberMap's own extensions holds information about selected high-quality resources. HealthCyberMap then uses GIS spatialisation methods to generate interactive navigational cybermaps from the metadata base. These visual cybermaps are based on familiar metaphors for image-word association to give users a broad overview and understanding of what is available in this complex conceptual space of medical/ health Internet resources and help them navigate it more efficiently and effectively.

HealthCyberMap cybermaps can be considered as semantically-spatialised, ontology-based browsing views of the underlying resource metadata base. Using a clinical coding scheme as a metric for spatialisation ("semantic distance") is unique to HealthCyberMap and is very much suited for the semantic categorisation and navigation of medical/ health Internet information resources. HealthCyberMap also introduces a useful form of cyberspatial analysis for the detection of topical coverage gaps in its resource pool using choropleth (shaded) maps of human body systems. The project features a cost-effective method for serving Web hypermaps with dynamic metadata base drill-down functionality. It also demonstrates the feasibility of Electronic Patient Record to Online Information Services (like HealthCyberMap) Problem to Knowledge Linking using clinical codes as crisp problem-knowledge linkers or knowledge hooks.

The Semantic Subject Search Engine queries the same HealthCyberMap resource metadata base. Explicit concepts in resource metadata map onto a brokering domain ontology (ICD-9-CM) allowing the search engine to infer implicit meanings (synonyms and semantic relationships) not directly mentioned in either the resource or its metadata. Similarly, user queries would map to the same ontology allowing the search engine to infer the implicit semantics of user queries and use them to optimise retrieval.

A formative evaluation study of HealthCyberMap pilot service using an online user evaluation questionnaire, in addition to analysis of HealthCyberMap server transaction log, has been conducted during the period from 18 April 2002 to 1 June 2002 with very encouraging results. This two-method evaluation approach was guided by methodologies described in NIH[*] Web Site Evaluation and Performance Measures Toolkit among other resources.

Many exciting future possibilities have been also investigated by the author, including the further development of HealthCyberMap as a customisable, location-based medical/ health information service.

[‡] International Classification of Diseases, 9th Revision, Clinical Modification [*] National Institutes of Health, US

2 Aims and Objectives

"The application of GIS is limited only by the imagination of those who use it."
—Jack Dangermond, President, ESRI

2.1 Background and Motivation

The Internet is a system with unimaginable complexity. It has been said the Web is its own map. But as we surf the Web, we can only appreciate a very small part of it at any given time. We cannot for example figure out the relations of the page/ site we are visiting to the rest of the Web, how it measures compared to the rest of the Web, or how the rest of the Web looks like. We can have difficulties finding our way back to a resource we visited just few hours ago. We cannot easily locate medical/ health Web resources covering narrower, broader, related or similar topics as the one we are currently looking at, and except when we have no particular goal(s) in mind, we cannot plan our Web journeys ahead. We need a map or set of maps for this purpose. Ideally, we should be also able to link our health-related and clinical questions to such maps, and use them (or the underlying information) to find the appropriate answers to these questions in the right place and at the right time.

Maps are well known as powerful graphic tools that can be used to classify, visualise, communicate and navigate spatial and/ or spatialised relations (relations projected into some conceptual space for mapping) in worlds that are too large and too complex to be seen directly. Maps build on humans' powerful spatio-cognitive abilities. Good Web navigational maps should be also based on sound cartographic principles and variables (to ensure their spatial legibility and utility), and adapted to the unique nature of the Web.

GIS (Geographic Information Systems) take simple cartography one step further by providing contextual links between maps and underlying databases (where attributes of features on the maps are stored). On the Web these links can be implemented as sensitive clickable maps (hypermaps). The quality and utility of such maps will obviously depend on the quality of the information in the underlying databases.

2.2 Aims and Mission Statement

HealthCyberMap (<http://healthcybermap.semanticweb.org>) is a Web-based service that aims at mapping medical/ health information resources in cyberspace in unique and novel ways to deliver a semantically superior experience to resource providers and consumers (the general public, patients, healthcare professionals and librarians), helping them plan, manage, search and navigate this complex "virtual" space more efficiently and effectively. This is achieved through semantic indexing, "intelligent" categorisation, and interactive hypermedia visualisation of the medical/ health information cyberspace using metadata, clinical codes and GIS technologies. HealthCyberMap only attempts to map *parts* of available health and healthcare Internet resources and services.

2.3 Objectives

(with measurable outcomes)

A navigational hypermap for browsing collections of Internet information resources should be ideally driven by an underlying metadata base that stores meta-information (information about information) describing these resources. HealthCyberMap comprises two main arms or layers. The top-level visualisation/ navigation arm (interface layer) is founded upon a robust semantic layer.

2.3.1 Semantic Arm Objectives

a. To define and model a suitable framework that tailors existing generic metadata standards/ recommendations to suit the description of medical/ health information resources.
b. To improve resource metadata quality and semantics through proper use of clinical codes from a controlled vocabulary as descriptors of medical/ health resource contents.
c. To propose and examine ways of reasoning semantically with resource metadata (including clinical codes) beyond simple queries, and combining metadata with other related and useful knowledge bases (ontologies or vocabularies) for the purpose of customising the language, interface and content of material delivered to consumers and offering them location-specific information.
d. To build a pilot metadata base describing a small part of medical/ health Internet resources based on the above framework, with the application of suitable resource quality benchmarking measures.
e. To develop a pilot Medical Semantic Subject Search Engine that outperforms conventional free-text and keyword-based search engines, and supports synonyms,

disease variants, subtypes, as well as some semantic relationships between medical/ health terms.

f. To develop a pilot clinical Problem to Knowledge Linking solution based on clinical codes to contextually link the Electronic Patient Record to the resource metadata base.

2.3.2 Visualisation/ Navigation Arm Objectives

a. To find an effective semantic metric based on clinical codes that can serve as a basis for the topical mapping of medical/ health information resources in cyberspace.
b. To find and use suitable spatialisation methods and familiar metaphors to visually browse and navigate medical/ health information resources on the Internet in different (but complementary) ways based on clinical codes and other metadata elements in the resource metadata base. This objective will be complemented by carrying a review/ critique of some existing and related cybermapping projects.
c. To use GIS to automate the classification and generation of spatialised browsing views (navigational cybermaps) of the resource metadata base based on the above cybermapping methods.
d. To develop a method for identifying topical coverage gaps (infogaps) in mapped resources.
e. To develop a cost-effective solution for serving hypermaps with dynamic database links/ drill-down functionality on the Web and use it for publishing the GIS-generated navigational maps on the Web while maintaining their links with the underlying resource metadata base (these maps form the main part of HealthCyberMap pilot service on the Web). It should be also possible to use the same solution to publish other interactive GIS-driven maps on the Web, e.g., maps of real-world health problems.
f. To develop a strategy for maintaining the currency of the generated hypermaps and dealing with problems related to Web resource link persistence.

2.3.3 Evaluation Objectives

a. To identify the need and reasons for evaluating HealthCyberMap pilot service interfaces on the Web, and investigate and develop/ refine suitable methods for conducting this evaluation and analysing its results.

b. To conduct a small-scale formative evaluation of HealthCyberMap research pilot using suitable methods (developed in the above objective), analyse the results, and identify any useful lessons that could be learned and possible future directions that could help improving HealthCyberMap if the service is further developed beyond the current project plan and timeline.

Part I: Literature Review

3 Towards a Semantic Medical Web: Metadata, Ontologies and Web Services

3.1 Background: Limitations of the Current Web and the Semantic Web Initiative

The current Web is primarily designed for (but not optimised for) human consumption. Web content is only intended for human readers to understand, leaving formatting and presentation to machines. However, humans commonly report difficulties locating what they want on the Web in a timely manner (sometimes it is like trying to find a needle in a haystack). Because machines are unaware of the actual *context* and content *meaning* of different Web resources, they cannot offer us much help in this regard. Current Web search engines do not really distinguish officially approved medical guidelines from experts' opinions on the same topics, and cannot easily tell a personal home page from an academic Web site [1, 2].

Free text word-based (or phrase-based) search engines like Google (<http://www.google.com>), AltaVista (<http://www.altavista.com>), Lycos (<http://www.lycos.com>) and Excite (<http://www.excite.com>) typically return innumerable completely irrelevant "hits", requiring much manual weeding by the user [3, 4] and might miss important resources. Free text search is not always efficient and effective for the following reasons [5]:

- the sought page might be using a different term (synonym) that points to the same concept. "Myocardial infarction" and "coronary thrombosis" cannot be matched, although they are the same;
- spelling mistakes and variants are considered as different terms. For example, "psoriasis" (correct spelling) and "psoriaisis" (typographical error) cannot be matched. Similarly, "anaemia" (correct UK spelling) and "anemia" (correct US spelling) cannot be matched; and
- search engines cannot process HTML (HyperText Mark-up Language) intelligently and are unaware of the actual context and content meaning of different Web resources. For example, searching for resources on "psoriasis" will retrieve all the documents containing this word, but many of these resources might not be relevant ("psoriasis" was just mentioned by the way in these documents, e.g., under a "See also" heading, and is not their actual topic).

Even when sophisticated statistical techniques are used for information retrieval, e.g., PubMed's RELATED ARTICLES function, the results are not that good (regarding their relevance) [6]. Liu and Altman [7] developed a program for incremental updates of a bibliography based on PubMed's RELATED ARTICLES function and could only demonstrate a recall of 75%, a strict retrieval precision (perfect relevance) of 32% and a partial precision (incomplete relevance) of 42%.

Many people have proposed using Natural Language Processing (NLP) to figure out Web pages. Unfortunately, NLP is still immature, and has not yet overcome many tremendous obstacles like the interpretation of Web graphics and diagrams designed for a human reader [3].

On the other hand, human-made Web catalogues or topical directories like Yahoo! and the Open Directory Project (<http://dmoz.org/>), where resources are classified under different categories, are not much better. The Web is very rapidly growing and changing, with so much information that humans at these directory services cannot possibly keep up [3]. Moreover, there is frequently an inevitable overlap between the categories in a Web catalogue and imprecision regarding the definition of their scope, leading to confusion as to what to expect under a given category [8].

With all current systems, irrespective of the methods they use, there will always be someone who cannot find what they want for *semantic* reasons. These problems can only be alleviated if search engines no longer search for matching words, but search on the semantic concepts underlying the information in Web pages and their relationships. The Semantic Web initiative (<http://www.w3.org/2001/sw/> and <http://www.semanticweb.org>) aims at creating the next-generation World Wide Web where information semantics (or meaning) are represented in a form that can be "understood" by machines as well as by humans (by giving machine-readable semantics to the currently presentation-based Web pages). This will pave the way for more "intelligent" machine-to-machine communication and information agent interoperability, and should ultimately empower human Web readers and solve many of the information management and retrieval problems they experience today with the current Web [1, 2]. Metadata, ontologies and ontology representation languages are pivotal ingredients of the forthcoming Semantic Web [4].

3.2 Ontologies Defined

Originally a philosophical discipline, ontologies are now also a hot topic in computer science in such diverse areas as knowledge representation, natural language processing, machine learning, databases, information brokering and retrieval, knowledge discovery and management (transforming document repositories into proper knowledge repositories), multi-agent systems and Semantic Web research [1, 9]. Knowledge Engineers are now called Ontologists. Ontologies can enhance Web searches, relate the information on a page to associated knowledge structures and inference rules, and help us develop agents that can address complicated questions whose answers do not reside on a single Web page [4].

An ontology is a consensual, shared and formal description of the concepts that are important in a given domain and their properties (attributes) and relations between them, i.e., it is a conceptual knowledge *model* or a specification of a conceptualisation [1, 10]. Ontologies also help establishing common information exchange grounds between members of a community of interest. These members can be human or automated agents [11].

Stoffel and colleagues [12] distinguish between two types of ontologies:

- Traditional ontologies consisting of only the definitions (correspond to Resource Description Framework (RDF) schema—see below).
- Hybrid ontologies combining both ontological relations and the instances defined thereon (correspond to RDF schema + RDF instances—see below).

Merging multiple ontologies (knowledge domains and vocabularies) is usually necessary to solve real world problems. Sharing and re-using ontologies can save time, make systems more flexible and facilitate many maintenance tasks. A good example is MELISA (MEdical Literature Search Agent), an ontology based information retrieval agent that merges a custom-built medical ontology with MeSH (Medical Subject Headings) ontology. MELISA demonstrates how ontologies can be very useful in enhancing Web searches [13].

3.3 RDF as a Semantic Web (Ontology Representation) Language

In his original 1989 proposal that gave rise to the World Wide Web, Tim Berners-Lee mentioned some ideas very closely related to those formalised nearly a decade later as RDF (Resource Description Framework). In particular, he suggested a directed, labelled graph model with link and node types that encompass metadata applications, in addition to simpler document linking [14].

RDF is a W3C (World Wide Web Consortium) recommendation for metadata. RDF can be serialised as XML (the eXtensible Markup Language), and builds on a well-known branch of mathematics: graph theory, plus the experiences of the knowledge acquisition and representation community (see <http://www.w3.org/TR/rdf-mt/> for details of RDF Model Theory). RDF can represent relationships, while raw XML cannot. Object-Attribute-Value triples (or Subject-Predicate-Object triples) form the basis of RDF. The Object (Subject) is usually a resource identified by a URI (Universal Resource Identifier). The Attribute (Predicate) is a property of that resource. The Value of this Attribute (Object) can be either atomic or other resources (URIs) or even metadata instances [5, 15].

RDF offers a simple format for a very small subset of predicate logic (basic RDF has no support for variables, rules, quantifiers, and n-ary predicates with n>2), making it possible to use it for modelling ontologies and drawing conclusions by generalising from assertions or from combining several assertions. Unlike conventional predicate logic, the syntax of RDF is declared in an RDF schema (RDFS), which means it is specific to the application at hand instead of being general, like predicate logic. The RDF schema is used to define the set of resources that may be used by a model, including constraints for resource (e.g., range and domain) and literal values (constants or string values). It creates the structure which the user later fills with his/her description (instances) and which can be used for consistency checks (that the actual RDF triples are following the defined constraints) [15, 16].

RDF schema enables us to express relations like `rdfs:subClassOf` and `rdfs:subPropertyOf` that standard relational database management systems (RDBMS) cannot easily cope with (though storing the base RDF Object-Attribute-Value triples in an RDBMS is not a problem, querying them remains very inefficient because triples belonging to each other are scattered over the whole database). XML namespaces are used in RDF to unambiguously define the context of statements [15]. However, XML namespaces have shown to be too weak to fulfil this purpose, and that is why some RDF-based languages and processors like Cwm (Closed world machine—<http://www.w3.org/2000/10/swap/doc/cwm.html>) and TRIPLE (<http://triple.semanticweb.org/>) extend RDF with a notion of contexts or models, respectively.

Newer languages, namely OIL (Ontology Inference Layer) and DAML+OIL (US Defense Advanced Research Projects Agency—DARPA Agent Markup Language merger with OIL), which build on RDFS, offer full support for features that cannot be represented in basic RDF, like axioms [16, 17, 18, 19, 20]. The W3C Web Ontology Working Group (WebOnt—

<http://www.w3.org/2001/sw/WebOnt/>) is currently developing a Web Ontology Language (OWL) to replace DAML+OIL. RDF also forms the basis of other emerging W3C standards, e.g., the Composite Capability/ Preference Profile (CC/PP) uses RDF to communicate profiles of mobile device properties to servers in order to deliver suitable content to these devices [1, 15, 21].

3.4 Metadata for the Semantic Web: Dublin Core Metadata Initiative (DCMI)

Metadata are documentation about documents and objects, or structured information about information. When properly implemented, metadata can crisply and unambiguously describe information resources, enhancing information retrieval and enabling accurate matches to be done, while being totally transparent and invisible to the user. Search specificity is increased (noise reduction/ only good matches) and search sensitivity is boosted (i.e., silence or missed matches are decreased and signal-to-noise ratio increased/ all good matches) [5].

The Dublin Core (DC – named after Dublin, Ohio, US) is a 15-element metadata set or vocabulary intended to aid discovery of electronic resources. The current DC version 1.1 (<http://dublincore.org/documents/dces/>) describes the following elements covering resource content, intellectual property and instantiation: title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage and rights. Most of the elements have commonly understood semantics of roughly the complexity of a library catalogue card [22, 23].

A list of health projects using Dublin Core, including HealthCyberMap, is available from: <http://www.chu-rouen.fr/documed/dc.html>.

Dublin Core Metadata Initiative (DCMI) defines two interrelated types of metadata [23]:

1. metadata embedded within the resource itself (peripheral metadata; this type helps automating the acquisition of the second type); and
2. stand-alone metadata separate from the resource it points to (central catalogue or index metadata).

Metadata of the first type can be inserted in the header of HTML documents (resources) using `meta` and `link` tags [5]. Some offline and online tools already exist that will assist the creators of Web resources in compiling and inserting DC metadata mark-up into their HTML documents, e.g., ReadTag (offline – [5]), and Reggie and UKOLN (online – [24, 25]).

Stand-alone metadata can exist in any kind of database (including RDF), and provide a link to the described peripheral (external) resource. HealthCyberMap relies on stand-alone (central) metadata, coupled with a controlled healthcare vocabulary for crisp subject descriptions and to allow semantic reasoning about the indexed resources and their relations with each other (see Chapters 7 and 8).

Unlike other approaches like SHOE (Simple HTML Ontology Extension – [26, 27]), HealthCyberMap does not impose any structural changes like embedded metadata or ontology instance mark-up on the peripheral resources or their hosting servers. However, Web authors should be encouraged to annotate their resources with appropriate (peripheral) metadata tags using some standard set like DC. As more and more Web authors begin to embed suitable metadata tags in peripheral Web resources, central cataloguing might one day become straightforward and nearly automated, except for the process of ascribing quality rating information, which will always depend on a human assessor for quality appraisal.

3.5 Clinical Codes for the DC Subject Field

Metadata can greatly enhance information retrieval, but this depends on the *quality of the metadata* we are using. Using keywords in the DC subject field to describe the content of a resource is not the optimal solution [14, 28]. Using clinical codes to describe the subjects of medical Web resources can boost metadata quality and hence offer superior (and automated) topical categorisation or classification of these resources, especially when combined with the power of a terminology server/ semantic search engine (see Chapters 7 and 8).

Appleyard and Malet [29] were among the first to mention that "the incorporation of nomenclatures, such as UMLS (Unified Medical Language System) and SNOMED (Systematised Nomenclature of Medicine), into meta-tags will allow the linking of Web-based knowledge sources into electronic medical record systems". In 1997, Detmer and colleagues presented MedWeaver, a Web application integrating decision support functions, a literature searching system, and a clinical Web searching system, using the UMLS Metathesaurus for vocabulary translation [30]. Reuters Health (<http://www.reutershealth.com/>) currently uses SNOMED for the topical categorisation of its health news.

Rector [31] mentions the following tasks relevant to digital libraries and management of knowledge resources among other tasks that a clinical terminology can fulfil:

- navigating and browsing through information – either locally or on the Web;
- authoring knowledge – either static knowledge for browsing or dynamic decision support; and
- indexing knowledge – both general medical knowledge and information about individual patients (this can be the basis of clinical problem to knowledge linking—see Chapter 10).

A clinical terminology or classification is a kind of ontology by definition and should preserve (and "understand") the semantic relationships between the thousands of terms in it or else it would become a mere dictionary (or at best a thesaurus). By labelling or tagging a resource (or a metadata record of it) with clinical codes, we are automatically establishing the relationships (as defined by the coding scheme used to tag the resource) between this resource and related (tagged) resources in the medical Web and also the (similarly coded) Electronic Patient Record (for clinical problem to knowledge linking).

A complete controlled healthcare vocabulary solution should ideally take care of all concept synonyms[§] and be able to reason with the multi-hierarchical relationships between concepts[**] [28, 32]. Users could start asking questions like "Get pages describing the *complications of* diabetes mellitus" and retrieve relevant pages, say on "peripheral neuropathy", because a knowledge-based clinical terminology represents the semantic relationship between "diabetes mellitus" and "peripheral neuropathy" ("peripheral neuropathy" is not a synonym or variant of "diabetes mellitus").

DCMI recommended best practice for populating the DC subject element is to select a value or code from a "controlled vocabulary or formal classification scheme".

3.6 Semantic Web Services

McIlraith et al [33] and Hendler [34] provide an extensive discussion of Semantic Web Services. Hendler [34] mentions a service advertise/ discover mechanism. Microsoft has already implemented closely related ideas in its .NET/ Web Services

---

[§] Example: treat "myocardial infarction" and "coronary thrombosis" as synonyms.

[**] Examples:

- recognise that "myocardial infarction" *is a type of* "ischaemic heart disease".
- recognise that "surgical operation" acts as parent for "transplant operation" and "kidney operation", which are both in turn parents of "kidney transplant", so that a search for either "transplant operation" or "kidney operation" would find "kidney transplant".

framework (<http://msdn.microsoft.com/theshow/Episode013/>), which is starting to evolve into an industry-wide standard (see <http://www.w3.org/2002/ws/>).

A WSDL (Web Service Description Language) contract in XML (eXtensible Mark-up Language) is automatically published by the Internet Information Server (IIS, a Microsoft server) of Web Service provider and is used to prepare the necessary Proxy Class DLL (Dynamic Link Library) on the remote client. A Proxy Class allows us to code locally against a remote Web Service by proxy. SOAP (Simple Object Access Protocol) allows information exchange between service and clients in XML via standard Web protocols like HTTP (HyperText Transfer Protocol). A repository for Universal service Description, Discovery and Integration (UDDI—<http://uddi.microsoft.com/>) allows service providers to "advertise" their services (e.g., a medical information service), and service consumers to locate these services (e.g., an Electronic Patient Record client seeking contextual information), much like telephone Yellow Pages.

The philosophy behind Microsoft's implementation and other similar implementations like IBM WebSphere is that systems should be able to talk to each other instantly and reliably without the "headache" of developing *ad hoc* information exchange protocols. It is about developing an information architecture that is oriented around people and their data, instead of around specific devices, applications, or locations. Systems should advertise and provide their services in a reusable and flexible form to all disparate clients (service consumers) who might want to use these services and to integrate them in their interfaces.

3.7 Conclusion

By making the context and semantics of resources, not merely their raw text and formatting, amenable to computer "understanding", we can build a Semantic Web that is more useful to humans than the current Web. This requires proper use of metadata and ontologies (vocabularies and resource descriptions). Clinical codes can reliably describe the subjects of medical resources, establish the semantic relationships (as defined by underlying coding vocabulary) between related resources, automate the topical categorisation of resources, and ensure highly relevant information retrieval. RDF/RDFS (and related languages) and Web Services provide the "grammar" necessary for reliable information exchange across services and transparent service interoperability using shared vocabularies.

4 Cyberspace and Its Maps

4.1 Introduction

The term “cyberspace”, first used by William Gibson [35], literally means “navigable space” (from the Greek word *kyber*—to steer or navigate). The “soft” (information) part of the Internet consists of many interrelated spaces or services: the World Wide Web, FTP (File Transfer Protocol), e-mail, mailing lists, bulletin boards, newsgroups, online shopping (usually hosted on the Web), etc.[36] Considerable amounts of health-related activities and information are exchanged in these spaces every day.

Humans, by virtue of their spatio-cognitive abilities, are able to navigate through geographic space. Those cognitive skills also have similar value in the exploration and understanding of non-geographic information and spaces [37].

Maps are powerful graphic tools that classify, represent and communicate spatial and/ or spatialised relations (relations projected into some conceptual space for mapping—see below). Maps are also a method to visualise and navigate a world that is too large and complex to be seen directly. A map will show us more in less space, and consequently we will be able to plan our route in advance [38].

4.2 What Can We Map in Cyberspace?

There are three distinct but interrelated aspects of cyberspace that can be mapped [39]:

– Networks of multimedia resources and links, e.g., health information resources as mapped by HealthCyberMap based on their subject topics. The resultant maps can be classified as conceptual information space maps and can be used as a visual navigational aid for browsing mapped resources. The geographic aspects of information resources (coverage and provenance) can be also mapped and are sometimes very useful as an index to information resources [40, 41]. They are not necessarily the same as the physical location of hosting servers, e.g., a clinic in London, UK might have its Web site hosted on a server in California, US; however, the site remains more relevant to people living in London, UK.
– Networks of machines and communication links (the “hard” part of the Internet).
– Demographics of Internet users and cyber-societies. Such information can help us tailor Web-based information services like HealthCyberMap to users’ needs.

4.3 The Need for Maps to Navigate the Web

In Chapter 3, metadata descriptions of Web resources were introduced as the foundation for the next-generation Semantic Web. But as these metadata descriptions accumulate, it becomes necessary to develop effective ways for visualising and navigating the resultant metadata repositories as well as the different semantic relationships and attributes of described Web resources.

It has been said the Web is its own map [42]. But as we surf the Web, we can only appreciate a very small part of it at any given time. We cannot for example figure out the relations of the site we are visiting to the rest of the Web, how it measures compared to the rest of the Web, or how the rest of the Web looks like. We can even have difficulties finding our way back to a resource we have visited few hours ago ("lost in cyberspace"). We cannot easily locate health and medical Web resources covering narrower, broader or similar topics as the one we are currently looking at, and except when we have no particular goal(s) in mind, we cannot plan our Web journeys ahead and maximise their utility. We need a map or set of maps for this purpose.

4.4 Typology of Cybermaps

In their book "Atlas of Cyberspace" [39] and related Web site <http://www.cybergeography.org/atlas/atlas.html>, Martin Dodge and Rob Kitchin present at least fifteen categories of cybermaps (maps of cyberspace), some of them with apparently overlapping scopes. Two of these categories are very relevant to HealthCyberMap and are presented below.

4.4.1 Information Space Maps

These cybermaps represent information cyberspaces as two-dimensional maps using sophisticated information indexing and classification methods (cf. conventional landuse maps used in city planning). The aim of these maps is to give users a sense of the mapped information domains and to assist searching and information retrieval. HealthCyberMap maps belong to this category (see Chapter 9). Other examples of maps belonging to this group are presented below and include, Visual Net (<http://map.net/start> and <http://pubmed.antarcti.ca/start>—Figures 4.1 and 4.2), WebMap (<http://www.webmap.com/>—the public InternetMap demo is no longer available—Figure 4.3) and Kartoo (<http://www.kartoo.com/>—Figure 4.4).

Visual Net mapping technology has been developed by Antarcti.ca (<http://antarcti.ca/>) and aims at rendering computer networks in the form of 2D and 3D maps. Antarcti.ca used Visual Net technology in Map.Net (<http://map.net/start>—Figures 4.1 and 4.7) to provide multilevel (hierarchical/ categorical) information maps for browsing over two million Web sites from the Open Directory Project (<http://dmoz.com/>). Rather than using conventional search engine technology to navigate the Web (or indeed any other hierarchical information space), Visual Net creates a landscape that spatially represents relationships between data, though in a very abstract, geometric fashion.

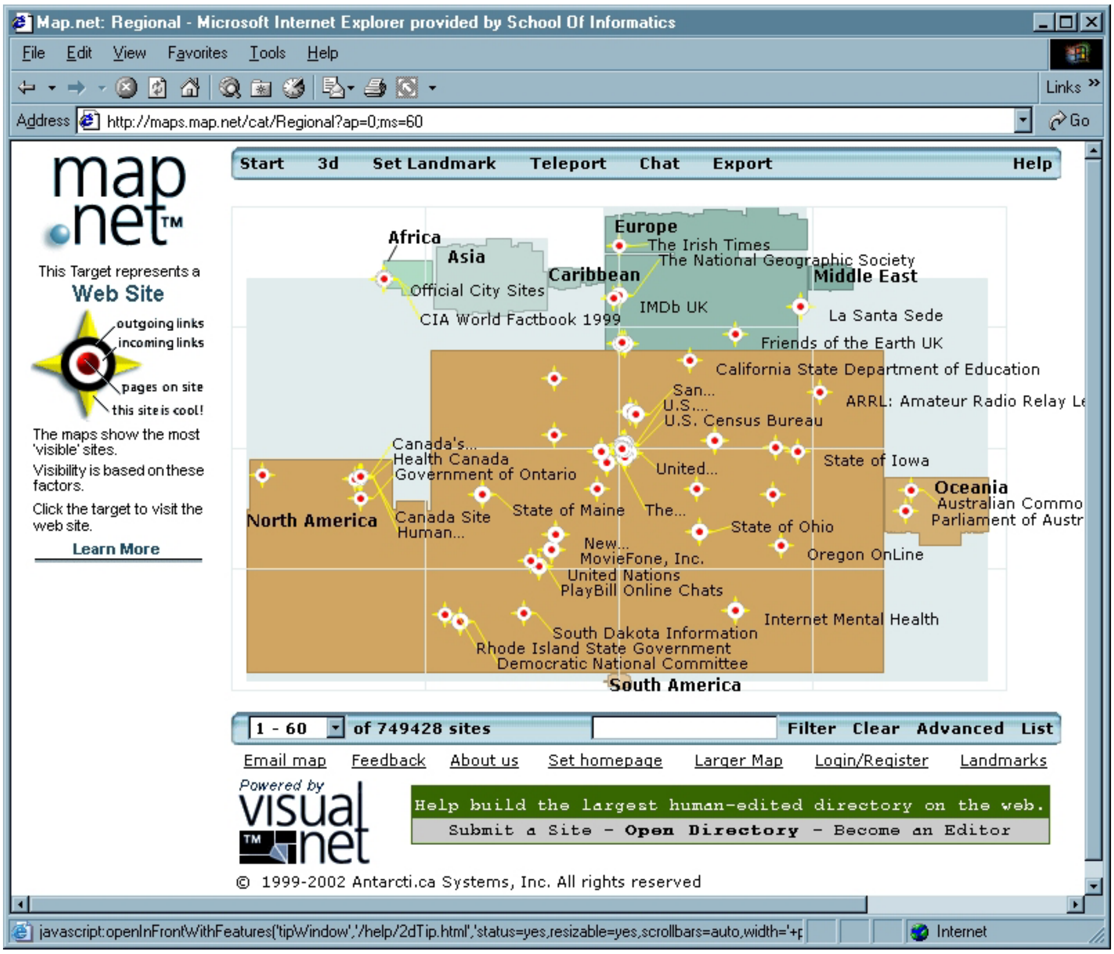


Figure 4.1. This Map.Net map (<http://maps.map.net/cat/Regional?ap=0;ms=60>) is supposed to be a regional category map of Web resources, but because it uses an *ad hoc* representation, some users might be lost, finding that Caribbean resources now lie between Asia and Europe, and above North America! Relative (geographic) position is not preserved in this representation for a topic that is explicitly geographical (continents/ regions of the world).

Antarcti.ca also applied Visual Net technology to PubMed, the US National Library of Medicine (NLM) well-known database of citations (<http://pubmed.antarcti.ca/start>—Figure 4.2—see also Figures 12.2 and 12.3).

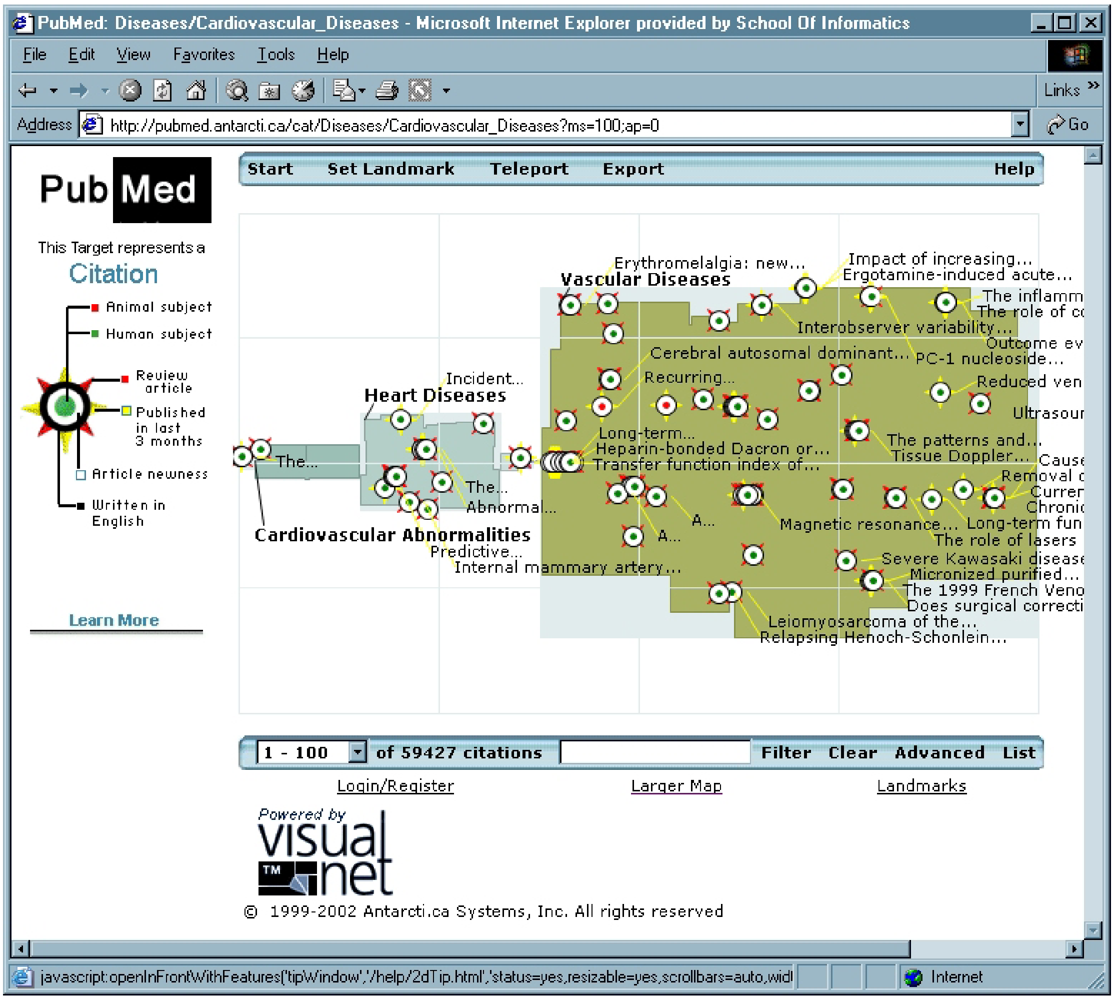


Figure 4.2. Screenshot of a map from Visual Net PubMed interface (<http://pubmed.antarcti.ca/cat/Diseases/Cardiovascular_Diseases?ms=100;ap=0>). The map appears overcrowded with overlapping citation symbols and their incomplete labels.

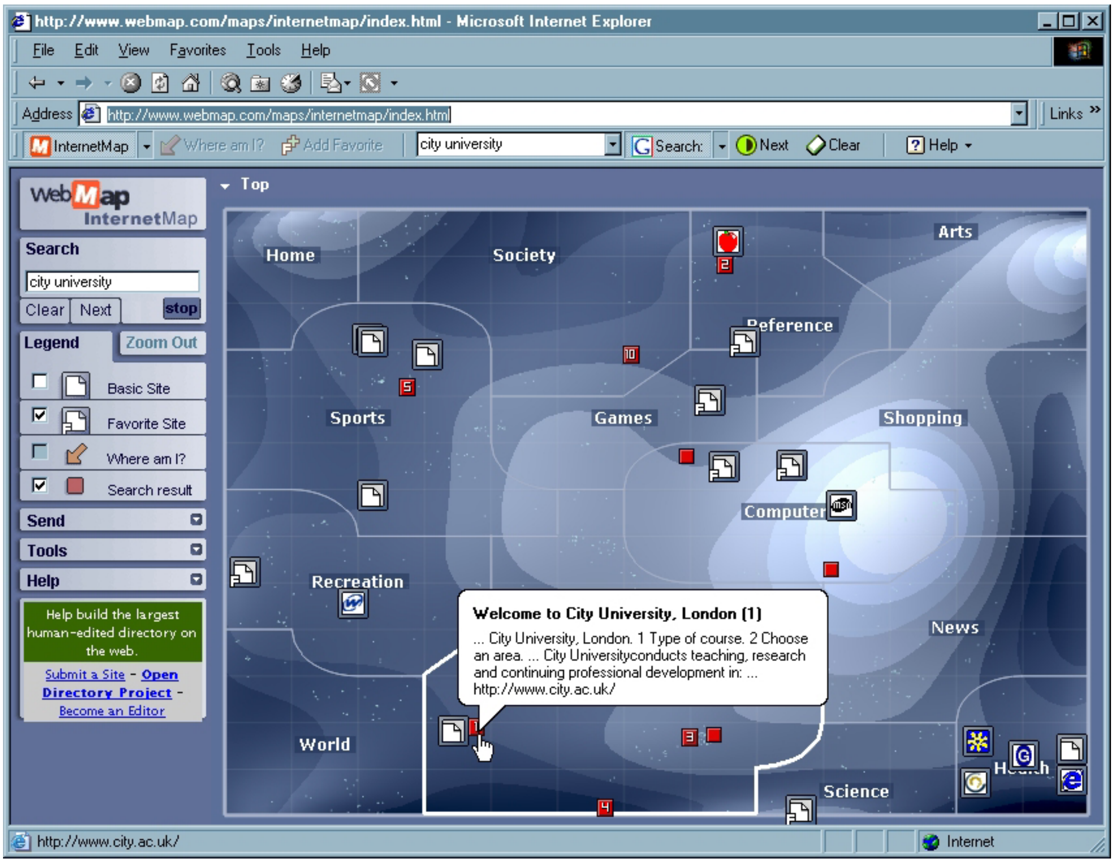


Figure 4.3. WebMap's InternetMap is an interactive, hierarchical visual directory that maps over two million categorised Web sites from the Open Directory Project (<http://dmoz.com/>). WebMap is a browser plug-in that represents hierarchical categories as irregular shaped polygons or territories. A smooth zoom brings more detailed categories into view. Each of these territories contains a different subject category. Within each territory, individual Web sites are shown by small mountain symbols and related sites are positioned close to one another. The concentric blue-coloured shapes represent WebMap's topography. The lighter areas are "peaks", while the dark blue areas are "valleys". With this topography, WebMap visually differentiates between sites. Sites that experience more traffic are placed higher on the topographic scale.

Figure 4.4. Kartoo (<http://www.kartoo.com/>) is a cartographic, visual meta-search engine that relies on the network metaphor as a visual interface to results pooled from external search engines. The cartographic interface of Kartoo uses Macromedia Flash. Labels of links between map nodes attempt to give an idea about the kind of relationship between two connected nodes (sites), but are very frequently confusing and incorrect.

Self-organising maps of information cyberspaces are also classified under this category of information space maps. Girardin [43] describes a Web resource mapping approach using Kohonen Self-Organising Maps (SOM). The self-organising map algorithm is an unsupervised neural network composed of an input layer and a competitive/ output neural layer. The most interesting property of this network is that the feature map preserves the topology of stimuli according to their similarity. Other related self-organising Web maps, include ET-Map at the University of Arizona (<http://ai2.bpa.arizona.edu/ent/entertain1/>—Figure 4.5) and WEBSOM at Helsinki University of Technology (<http://websom.hut.fi/websom/>). The latter has been used to map collections of Usenet newsgroups. SOM neural net output is sometimes semantically confusing. The flat landuse visualisation of ET-Map is also semantically poor compared to other forms of classification and visualisation.

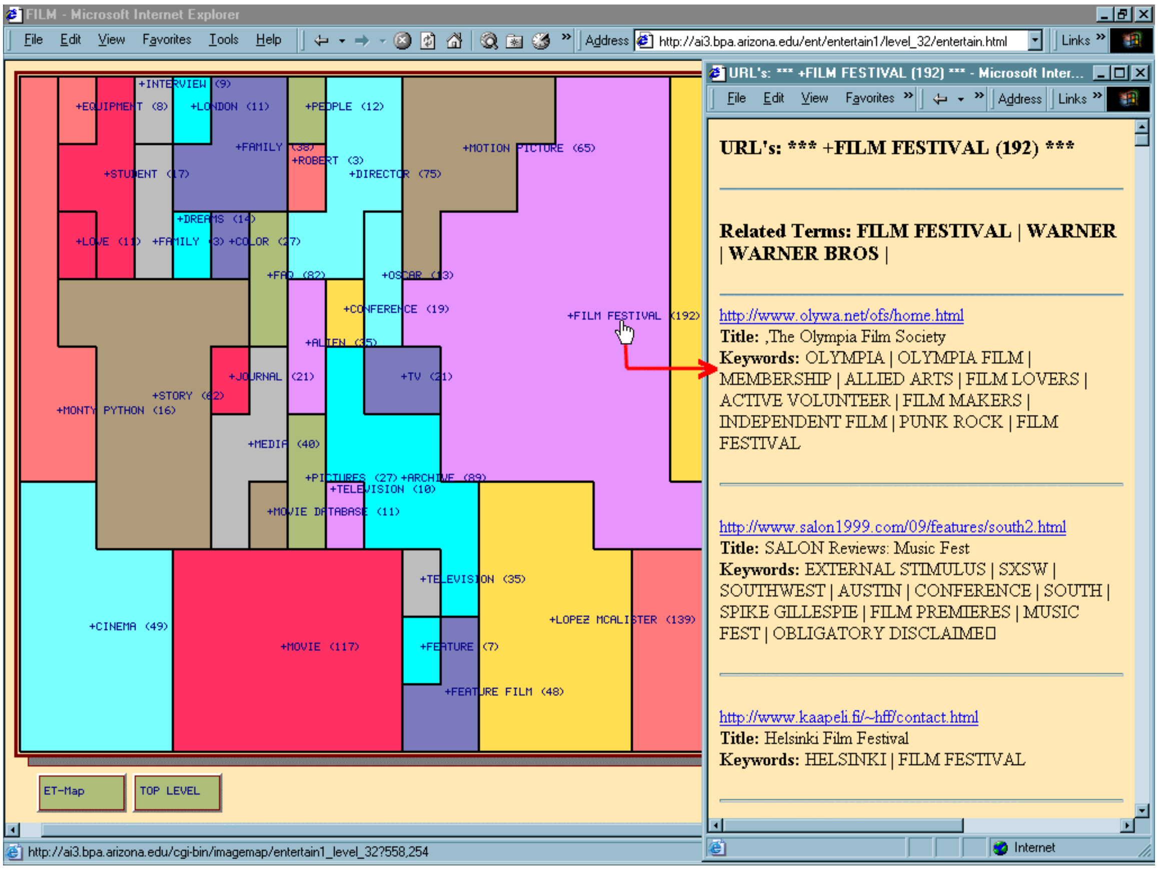


Figure 4.5. ET-Map (<http://ai3.bpa.arizona.edu/ent/entertain1/level_32/entertain.html>), a hierarchical category map set representing over 100,000 entertainment-related Web pages listed by Yahoo! ET-Map has been developed by researchers at the Artificial Intelligence Lab, University of Arizona, US.

Information resources can be also organised and navigated based on their geographic attributes, e.g., geographic scope of resource content or location of the hosting server or author(s)/ publisher(s). An example of such maps is the UK Academic Map, maintained by the University of Wolverhampton, UK, which went online in 1994 (Figure 4.6). Sites are shown on the map as dots, which act as hotlinks to the different Web sites of the academic institutions in the UK. However, by using a separate map symbol (dot) for each mapped Web site, symbols quickly become cluttered as many Web sites are obviously clustered in large cities that cover a relatively small geographic (and map) area in relation to the amount of information they contain [44].

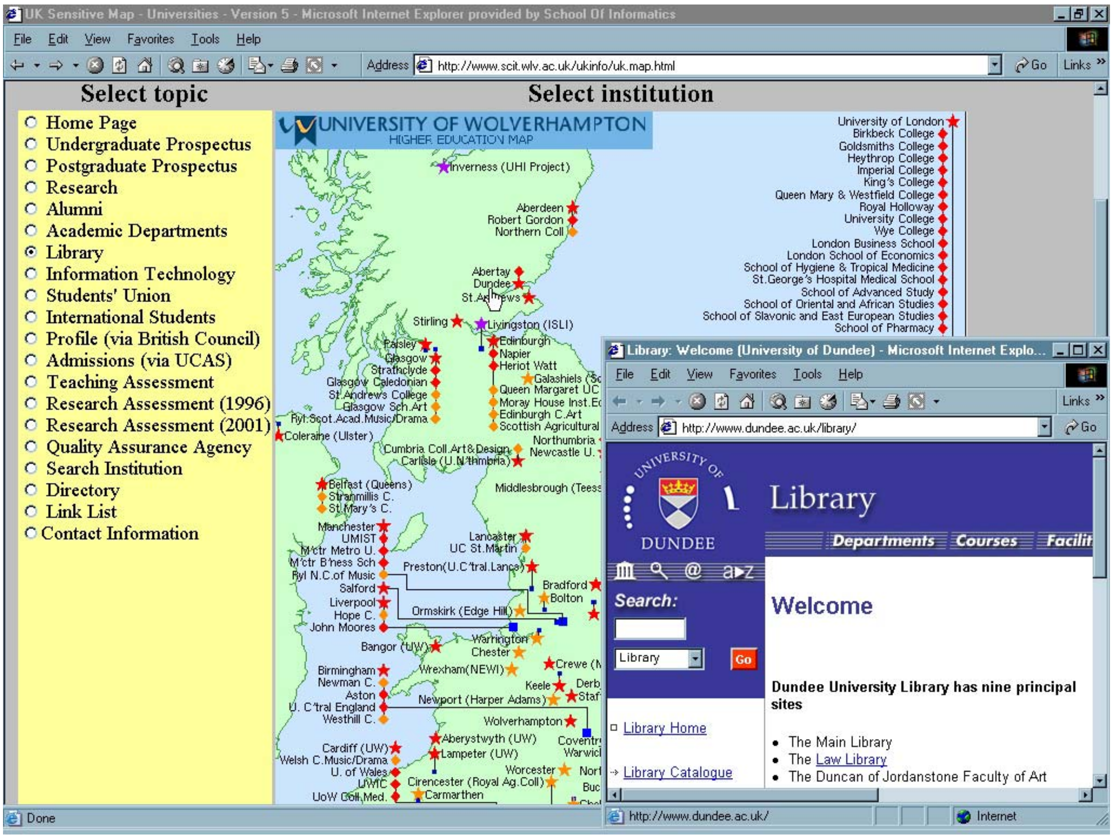


Figure 4.6. Screenshot of the University of Wolverhampton UK academic sensitive map (<http://www.scit.wlv.ac.uk/ukinfo/uk.map.html>). The "Library" topic was selected (yellow left pane) and "Dundee" clicked to launch "Dundee University Library and Information Services" Web site.

4.4.2 Information Landscapes

Maps under this class represent cyberspaces as three-dimensional (3D) landscapes. The 3D-cityscape view of the Web generated by Visual Net's Map.Net (Figure 4.7) is one example of such maps. It allows the user to fly-through the Web, with individual Web sites represented by different buildings. The large skyscrapers are the most popular and important sites on the Web. Another example of 3D information landscapes is StarWalker [45]—Figure 4.8.

Andrews [46] argues that visual representation of information spaces using technologies like Visual Net, although appealing, is unwieldy as a navigation method. He mentions that people usually complained of technical limitations, vertigo and confusion when presented with fly-through metaphors for data navigation. He believes that systems intended for *casual* or *untrained* Web surfers will likely have the best chance of success by focusing on increased *simplicity*, involving *universal iconography* on the order of road signs, not new dimensions of navigation.

Figure 4.7. Screenshot showing some UK health resources in Map.Net 3D (<http://map.net/start>). Users are essentially "walking" on the ground, among buildings that represent Web sites. Map.Net 3D requires a special plug-in to be downloaded and installed first.

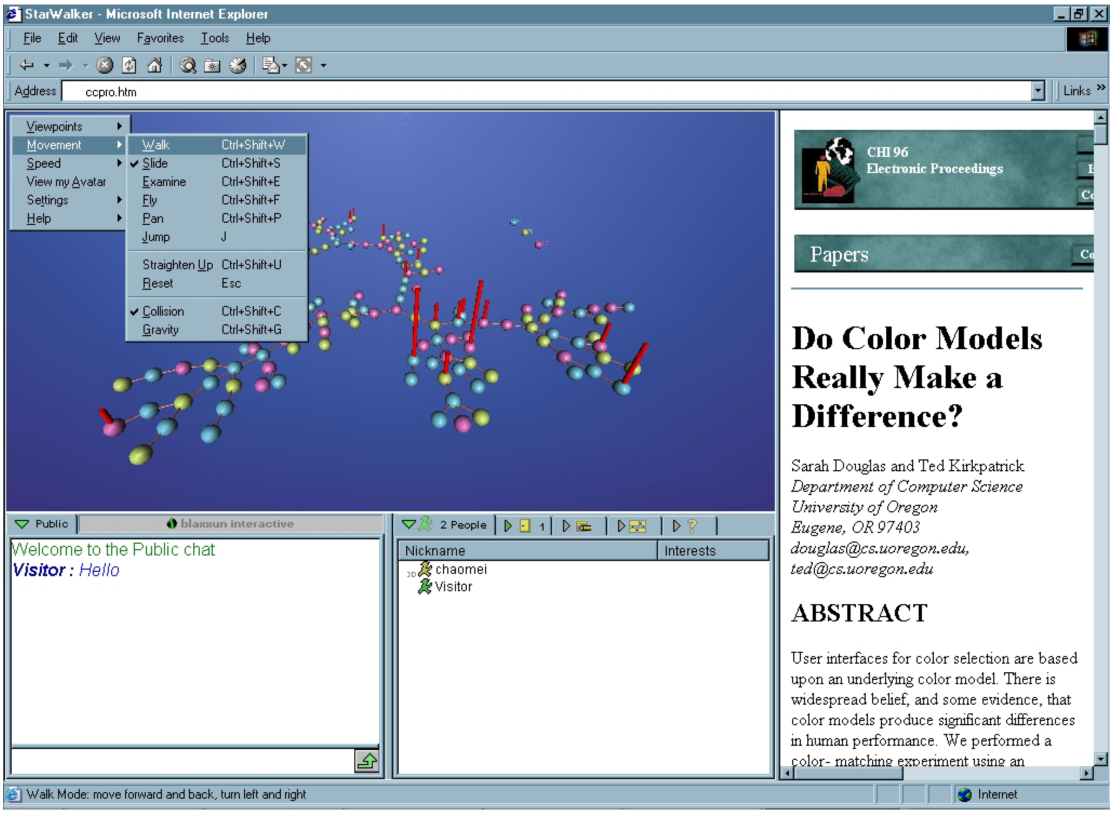


Figure 4.8. StarWalker interface (previously available from: <http://www.brunel.ac.uk/~cssrccc2/vrml2/starwalker/>). The linked stars are icons representing papers from three years conference proceedings (1995-1997) of ACM CHI (Association for Computing Machinery—Special Interest Group on Computer-Human Interaction). When the user clicks one of these icons, the full text version of the corresponding paper appears in the right frame. Links between spheres indicate that the corresponding papers have salient semantic connections as derived through a Generalised Similarity Analysis modelling procedure, which is used to build the semantic space that is then rendered into a Virtual Reality (VR) model in VRML (Virtual Reality Modelling Language). Vertical bars in red are some papers concerning social interaction and discourse structures in the context of Human-Computer Interaction. Users can move around within this VRML-based virtual world, e.g., walk, fly, etc. StarWalker does not tell the user what resource a star is pointing to until the star is actually clicked and the resource loaded. Also, there is no map legend, but the developers of StarWalker mention that articles (stars) are colour-coded by the year of publication [45].

4.5 Bibliographic/ Cybergraphic Uses of GIS

The conventional use of GIS (Geographic Information Systems) to analyse and map real-world health and healthcare data is well documented in the scientific literature [47]. However, research literature on the bibliographic/ cybergraphic uses of GIS to map semantic (information) spaces is scarce and includes the work done by Fabrikant

[48], Old [49] and Terpstra [50]. (It is noteworthy that none of the examples of cybermaps presented above uses GIS.)

Old [49] describes two main steps when using GIS to map conceptual information spaces (as in HealthCyberMap):

- First, information in non-spatial data is spatialised (see below), analysed, browsed, and processed using (desktop) GIS and cartographic methods; then
- The resultant information maps and their connections to the underlying data are shared on the Web as sensitive clickable maps for Internet browsing and navigation of mapped spaces.

4.6 Spatialisation and Metaphors

Spatialisation is the process by which information with no inherent spatial attributes (no geographic referent) is mapped onto a defined spatial framework using a variety of spatial metaphors. Information attributes are transformed into a spatial structure (conceptual space) that can be visualised and navigated through the application of concepts like hierarchy, proximity and similarity; e.g., a topical hierarchy based on the subjects of a collection of Web resources. The goals of spatialisation are to increase the spatial legibility and comprehension of information spaces, improve navigation through them, and enable people to find the information they are searching for in these spaces more easily [36].

A metaphor's primary function is to provide a partial understanding of one kind of experience in terms of another kind of experience. Spatial metaphors act as fundamental sense makers for abstract domains. Familiar metaphors taken from users' everyday life are usually much easier to understand [48]. In cybermaps, metaphor comprehension can be enhanced with appropriate use of visual variables (e.g., colour) and application of sound cartographic design principles [51].

It is noteworthy that a graphical browser already exists for visualising Read Codes using the familiar human body metaphor [52]. However, no one has yet used such visual interfaces that are based on the human body metaphor as the author did in HealthCyberMap, to categorise and browse Internet information resources indexed using a clinical coding scheme.

4.7 Conclusion

The term "cyberspace" as used in this manuscript refers to the navigable "soft" (information) part of the Internet, mainly the Web. As online information services accumulate metadata descriptions of Web resources, it becomes necessary to develop effective ways for visualising and navigating the resultant metadata repositories as well as the different semantic relationships and attributes of described Web resources. Maps are a good method to visualise and navigate a world that is too large and complex to be seen directly like the Web.

At least fifteen categories of cybermaps (maps of cyberspace) have been described, of them information space maps and information landscapes are the two categories most relevant to HealthCyberMap. HealthCyberMap maps can be classified as conceptual information space maps and can be used as a visual navigational aid for browsing mapped resources. The very abstract and geometric Visual Net maps (<http://pubmed.antarcti.ca/start>) are another example of information space maps. Information resources can be also organised and navigated based on their geographic attributes.

Research literature on the bibliographic/ cybergraphic uses of GIS (Geographic Information Systems) to map semantic (information) spaces is scarce. None of the cybermap examples presented in this chapter uses GIS to classify Web resource data and render the maps, with the exception of HealthCyberMap.

Spatialisation is the process by which information with no inherent spatial attributes is mapped onto a navigable visual spatial framework based on concepts like hierarchy, proximity and similarity, and using spatial metaphors to provide an understanding of one kind of experience in terms of another kind of experience. Spatialisation aims at increasing the comprehension of information spaces, and improving navigation and searching through them. Familiar metaphors taken from users' everyday life are much easier to understand and are the key to successful cybermaps.

5 Principles of Web Cartography

5.1 Introduction

Maps published in cyberspace could be either maps covering topics related to cyberspace itself, e.g., cybermaps of Web resources as outlined in Chapter 4, or maps using cyberspace, usually the Web, as a publishing medium, e.g., MARA/ARMA maps of malaria risk in Africa (<http://www.mara.org.za/>). The same principles and opportunities of Web cartography apply to both types.

5.1.1 Technical Classification of Web Maps

Web maps are classified into static maps and dynamic maps according to whether or not they are animated [38]. Each group is further subdivided into view only maps and interactive maps (Figure 5.1).

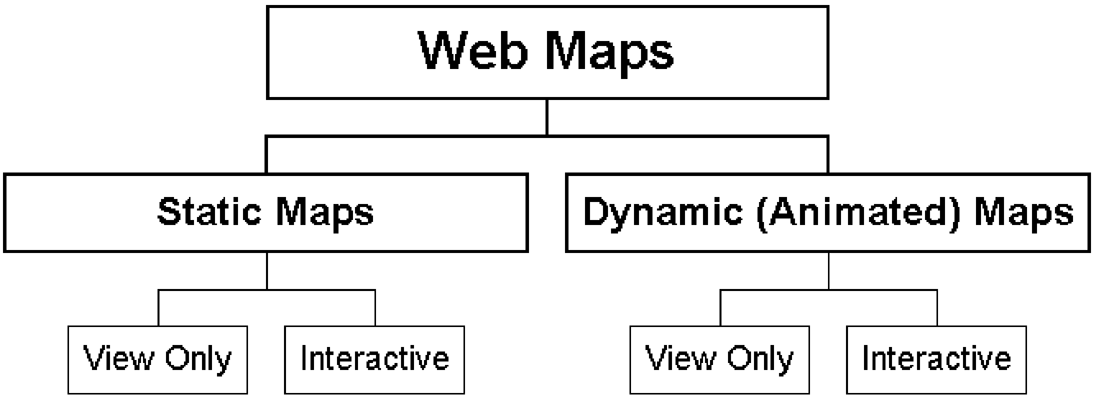


Figure 5.1. Technical classification of Web maps. HealthCyberMap's maps fit into the Interactive Static Maps category. (Adapted from Kraak and Brown [38].)

Non-animated (static) interactive Web maps, also called clickable maps, imagemaps or hypermaps, can be served either as client-side imagemaps or as server-side imagemaps depending on where mouse-click co-ordinates are resolved.

Interacting with a map can stimulate a user's (visual) thinking and encourage exploration. Clicking an object on such maps can lead to other Web resources, including other Web maps, and can even trigger a query against an underlying database and display the results. It is possible to put all kinds of additional information behind the map image thus reducing map clutter and size. Mouse events such as mouse-over (e.g., ToolTips for map feature labelling) and mouse clicking of map objects can be associated with this extra information. Interactivity could also mean users have the option to define map contents by switching layers on and off.

The map interface itself can be made interactive by providing the user with control options like panning, zooming in and out, and a smaller interactive overview map to highlight/ select the area covered by the currently displayed map tile in relation to a bigger map [38].

Figure 5.2 provides a good example of many of these interactivity options and features (see <http://healthcybermap.semanticweb.org/world_map/>). Note the different map interface buttons on the left for zooming, panning and other functions. A help window is displayed by clicking the button with '**?**' question mark symbol. Also note the overview map with a red positional square on the right; this helps users know where they are within the larger world map that cannot be displayed in full detail in one screen. The overview map is also clickable and can be used to select a different area for viewing.

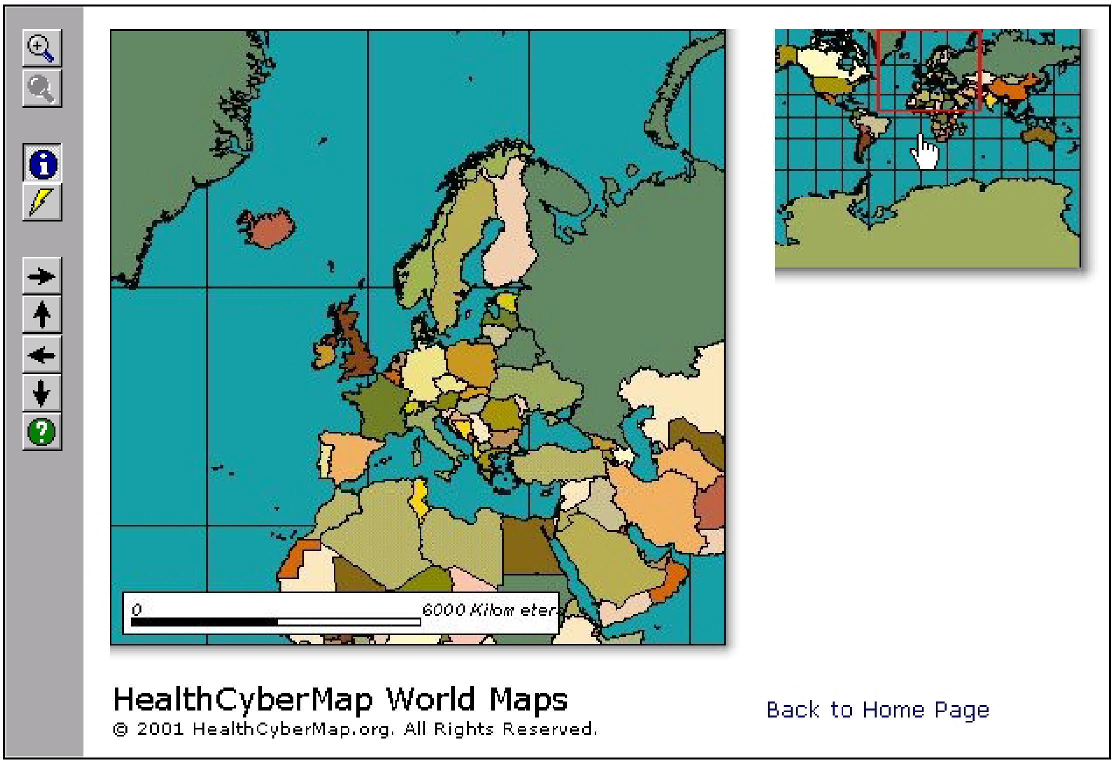


Figure 5.2. A good overview map with a red positional square like the one shown on the right in this screenshot from HealthCyberMap can help users get an indication of where they are within a large map like the world map which cannot be displayed in full detail in one screen. The overview map is also clickable and can be used to select a different area for viewing.

5.2 Cartographic Grammar and Guidelines for Web Maps

Cartographic methods and techniques are a kind of visual grammar. They allow for the optimal design and production of effective maps (including Web maps) that are well suited to the application at hand. The ultimate goal is always to visually reveal the spatial (or spatialised) objects, relations and patterns in the underlying data, so that map user can visually locate spatial (or spatialised) objects, while the colour and shape of symbols representing them on the map inform him or her about their characteristics.

5.2.1 Bertin's Visual Variables

To find the proper symbology for a map, a cartographic data analysis must be performed to assess the characteristics of the data and find out how they can be visualised. In his "*Sémiologie Graphique*", Bertin (1967—cited in [38]) distinguished six categories, which he called the visual variables. They are size, texture (grain), orientation, shape (form), value and colour (Table 5.1). These visual variables can be used to make one symbol (point, line or area symbol—see below) look different than another one. The choice of the correct variables for the data at hand is of utmost importance in the cartographic visualisation and communication process; if misused, it can lead to misleading results, cause loss of information, and convey a wrong impression.

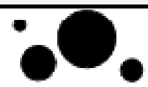
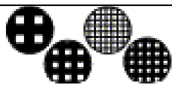
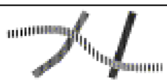
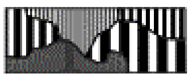
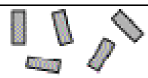
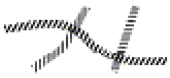
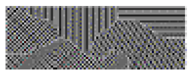
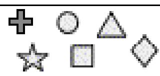
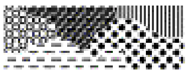
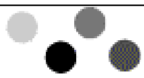
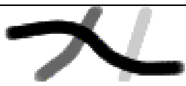
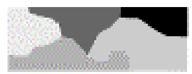
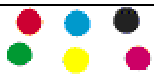
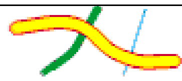
Visual variable	Q	S	O	A	Point symbols	Line symbols	Area symbols
Size	Yes	Yes	Yes				
Grain		Yes	Yes				
Orientation				Yes			
Form				Yes			
Value		Yes	Yes				
Colour		Yes		Yes			

Table 5.1. Bertin's visual variables — Q = Quantitative; S = Selective; O = Ordered; A = Associative. (Adapted from Kraak and Brown [38].)

The grouping in these six categories is based on the perceptual behaviour(s) that each category stimulates in the map user. These behaviours are quantitative, selective, ordered and associative. For example, differences in values (grey values or lighter/

darker shades of the same colour) convey differences in order, or relative quantity, e.g., maps of population or resource density. This variable (value) forms the basis of choropleth maps. The variables colour, shape and orientation have associative properties, as they can give the impression of difference among things and are the basis of chorochromatic or mosaic maps used to represent data such as landuse (cf. ET-Map, Figure 4.5) [38].

Contour maps and climate maps are probably the most well known representatives of the isoline maps [38]. Each line in an isoline map has the same value, e.g., of terrain height or temperature or Web site traffic/ popularity (cf. WebMap, Figure 4.3).

5.2.2 Cartographic Symbol Design for Web Maps

5.2.2.1 Point symbols

Point symbols are mainly used to represent spatial features occupying a very small area on the map at a given scale. They can be also used to provide shorthand information, e.g., a set of symbols next to a village on a tourist map to point to tourist facilities like restaurants, hotels, hospital, post office, etc. Three categories of point symbols exist: pictorial (e.g., aeroplane, tree, body organ symbol like the lung, etc.), geometric (polygonal symbols, squares, circles, etc.), and alphanumeric (letters and numbers). On interactive Web maps a point symbol can also function as a Web object, i.e., as an area that can respond to mouse events like mouse-over and mouse clicks, triggering JavaScript functions and hyperlinks.

Pictorial point symbols are usually very easy to understand by inexperienced map-readers, sometimes even without the use of a legend. The main problem in designing these symbols for Web maps is that the essential characteristics of the features they represent must be visualised without much clutter within a small screen area with a limited number of pixels (cf. designing icons for Windows). The symbols may also need to be larger than on equivalent paper maps to aid legibility and must be less complex regarding detail and colours to suit the lower resolution constraints of Web maps. However, this should not be considered a drawback since Web objects can be used to access the second level of information content that can cover a smaller area in more detail. For best results, a clear metaphorical relation should exist between the symbol and the user's real world knowledge.

Geometric or abstract symbols do not attempt to resemble the real feature represented. On different maps the same symbol can have a different meaning. Therefore geometric symbols should always be explained in a legend [38]. (Alphanumerical point symbols also require a legend.)

5.2.2.2 Line symbols

Line symbols on topographic maps can represent features like roads, railways and contours. In thematic maps, lines can show the position of geological fault lines, ocean currents, and trade flows. However, the thin and elongated shape of line symbols makes them very difficult to handle as interactive Web objects, especially when they are highly curved [38].

5.2.2.3 Area symbols

Area symbols are used to represent area-based information. The graphic variables typically used in designing area symbols for maps are colour, value, texture, shape and orientation. These can create complex area patterns. The proper use of variables can promote the semantic meaning of an area symbol, e.g., green tree symbols filling an area to represent forest area. It can also decrease the possibility of confusion among adjacent area symbols, e.g., by assigning a different colour to each country on a map of the world and using a blue colour for seas and oceans [38]. In Web maps, area symbols can also function as clickable Web objects, e.g., the different countries in HealthCyberMap's map of the world (Figure 5.2).

5.2.3 Features Labelling and Typographic Variables

Text labels on Web maps cannot be omitted as text can express information like geographic names that are not possible to express using any other symbol. Map typography and map symbol design cannot be separated from each other. Inappropriate application of typographic variables may affect the legibility of the text on a map and/ or clash with other graphic variables.

Typographic variables include [38]:

- type size: expressed in points;
- shape: refers to variations in font or type face, e.g., serif types like Times or sans serif types like Helvetica, and sometimes lower and UPPER case of the same font are considered shape variations; and
- orientation: refers to upright or *italic* variations within one font. Value refers to light, medium and **bold** font variations, or some **grey** value.

Text placement is also important; it can be horizontal, inclined, or curved along a path. Users of Web maps cannot rotate it to read very inclined or upside down text as they do with paper maps. Other important issues include text-background relation, e.g., any outline around text to make it clearer, and any antialiasing (font smoothing) used and its amount. Anti-aliasing might not be a good choice for very small font sizes and together with many other options available for typing text on a map, like cast shadows, they tend to increase image file size. Also important is character kerning (space between letters).

Considering text on Web maps, one can distinguish two types [38]:

- text applied outside the map face, such as in the legend. This can be saved as graphic in raster format, or saved as real text on a Web page (to be displayed in a separate frame, with the usual precautions to avoid/ care for different fonts on users' machines, and control text flow, letter spacing and space between lines); and
- text within the map face, e.g., for geographical names. This is usually saved as part of the map in raster format to be sure no text changes will take place at the user's side. ToolTips used to label map features (on mouse over) also belong to this type.

The density of text (the amount of textual information on the map) is also very important and an overcrowded map can become illegible (Figure 4.2). Text in this case can be made available as selective ToolTips only appearing on mouse hovering or as message boxes available by clicking hotspots on the map [38]. This is the approach adopted by the author in HealthCyberMap.

5.2.4 Contrast and Visual Hierarchy

Contrast will increase the communicative role of the map since it will create a kind of visual hierarchy or figure-ground relation in map contents, since usually not all map information is of equal importance [37, 38]. The visual hierarchy of Web map information content deserves more attention compared to maps in general. Three distinct levels exist [38]:

- Primary content level. This is formed by the main theme of the map. Interactive Web objects such as hotspots, mouse-overs, etc. can also be considered to form part of this level. These objects trigger specific events resulting in the supply of main theme information, e.g., in HealthCyberMap, a list of Web resources (the list itself belongs to the secondary content level).

- Secondary content level. This refers to the (often topographic) base map information, but also to pop-up menus and windows, and any movies, tables, sounds, etc. supplying information on the main map theme.
- Supportive content level. This includes the legend, other marginal information like grid, and any information that is not directly related to the main theme of the map, e.g., map interface help. A legend is necessary to understand how the topic is represented, e.g., a key to the colour tints used in a choropleth map and the class ranges they represent. If the legend and the map are saved as one image, the legend will disappear when the viewer zooms in to part of the map (or scrolls away from the legend corner if the map does not fit the whole window). The legend in Web maps is thus best treated as a separate image so that it can be kept in view in a separate frame.

5.2.5 Scale, Generalisation and Isomorphism

The visualisation of information spaces on a two-dimensional screen can be severely impeded as the volume and complexity of the respective dataset grows. Skupin [37] cites methods such as windowing, fish-eye views, hierarchical display, and tree condensing as examples of methods used to reduce the complexity of information visualisations.

Cartographers are capable of creating maps where geographic meaning is preserved throughout the scales, despite the large number of features involved. The processes of abstraction that achieve such scale-dependent representation are collectively referred to as cartographic generalisation. Generalisation can be based on controlling the number of classes into which features are grouped, or on simplification of form (geometric generalisation). At the root of the complexity problem lies a conflict between the number of visualised objects, the size of symbols representing them (and their labels) and the size of the display surface. Cartographic generalisation is deeply tied to the notion of scale, which is defined as the ratio between the size of a feature on the map to its size in the real world. Digital cartography has further expanded the classical concept of scale. Now the accuracy with which a given map scale represents the location and details of features is known as resolution. Computers can enlarge the scale of any map (by zooming in), but no additional (true) detail will be gained and map accuracy will not change unless a new higher resolution map is loaded replacing the first [37].

Conventional maps are also isomorphic, i.e., identical to or similar to the domains they represent, though they usually present a selective view of reality, only showing a subset of the features in their domains. During the process of creating this scaled-down view of reality, the cartographer has to select those features that will become amplified, while discarding the rest of the information in the domain according to the purpose of the map [53].

However, sometimes isomorphism is abandoned all together for the sake of clarity. A good example of non-isomorphic maps is the very popular London Underground map, which Henry C. Beck first designed in 1931 (<http://www.thetube.com/content/tubemap/images/mapjp.pdf>). The map is not drawn to scale and therefore is not 100% isomorphic. If enlarged to the actual size of London it would diverge significantly from the actual geography of the city, a fact many tourists soon discover when they try to use the Underground diagram as a walking map. Isomorphism is sacrificed to give downtown areas better coverage. The density of information is roughly the same all over the diagram (distances between stations are nearly equal), while downtown areas in reality have a much higher density of Underground installations. The result is a product that is highly functional for its purpose and is a good model, which some cybermaps have followed [53, 54].

5.2.6 Download Time

If the information takes too long to download (large files), users will lose interest and abort the process [38, 55]. The ideal Web map should not be too large in both file size and image size (to cope with the limited size of display screens). This means graphic and information density in Web maps should be kept low [38]. The Web offers many techniques for adding interactivity to maps (see above), which can be utilised to make smaller, “smarter” maps that are faster to download and provide additional functionality.

5.2.7 Screen Resolution, Colour Palette and Web Graphic Formats

The screen resolution and colour palette settings of display adapters can limit the amount of map detail that can be displayed on a computer monitor (compared to a map of similar size printed on paper). Mobile Internet devices (e.g., WAP devices –

Wireless Application Protocol) carry with them even more challenges to the Web cartographer, because of their much smaller (and usually monochrome) display screens.

Moreover, although Web graphics, including maps, are stored in platform-independent formats like GIF (CompuServe Graphical Interchange Format), JPEG (Joint Photographic Experts Group) and PNG (Portable Network Graphic), the same graphic or map still might not appear exactly the same to every user. This is because of differences in users' devices, browsers and operating systems (which handle colours in different ways), and in the quality (e.g., resolution) of their graphic cards and display screens. Besides, users are able to manually adjust their displays for resolution, contrast, brightness and colour balance. Fortunately, there are established ways of dealing with most of these different output conditions.

Web map designers should adopt a cautious approach by assuming the minimum configuration and lowest settings on users' machines, e.g., 8-bit/ 256-colour depth (though this is becoming less common, with most new machines now having their display set to 24-bit/ True Colour). The famous 216-colour Web- (or browser-) safe palette fits well into this configuration and should be used for Web images saved in a palletised format like GIF. The 216 colours in this palette are guaranteed to be non-dithered (smooth, solid colours) on any configuration, unless combinations of them are used to represent other colours not originally in the palette. This does not mean however that these 216 colours will always appear exactly the same on any system, since much depends on the calibration of the monitor and other factors as mentioned above [38]. (For some applications, like HealthCyberMap, this is not a big problem.)

GIF is a palletised format, i.e., images saved in this format are limited to a palette with a maximum of 256 colours. GIF is also a lossless compression standard (cf. JPEG below). GIF is more suitable for line art images and images with solid colours. GIF images can have transparent backgrounds if a colour from the underlying palette has been defined as transparent. Animated GIF maps can be seen as the view only version of the dynamic maps.

JPEG does not work with a palette; it compresses images based on colour and intensity (usually lossy compression). Even if a map is designed to contain only Web-safe colours, some of them might shift slightly during JPEG compression, possibly leading to dithered result on a 256-colour display configuration. For maps making considerable use of colour blends, e.g., contour maps, JPEG is however the best compression algorithm (this is the format used in HealthCyberMap).

Web map designers should test both formats (GIF and JPEG) on the maps they intend to produce, varying the various parameters available, including JPEG compression/ quality settings and GIF palette sizes. (GIF palettes can be adaptive, i.e., include only the colours that need to appear in an image.) The aim is to find which setting gives good results (usually assuming a 256-colour configuration) while maintaining a reasonably small file size [38].

PNG (<http://www.w3.org/Graphics/PNG/>) is an emerging royalty free Web graphic format that promises:

- to put an end to differences in colour display on different platforms because of colour control through gamma settings (see: <http://www.cgsd.com/papers/gamma.web.html>);
- to deliver real transparency by specifying an alpha channel so that a transparent layer really blends into whatever colour is underneath;
- to achieve smaller file sizes compared to similar GIF images; and
- to be always editable with the possibility to resave without loss of information.

The latest Web browser versions are natively supporting the PNG format, but not all of its features [38].

The vector approach to Web maps [38] (see Scalable Vector Graphics format (SVG)—<http://www.w3.org/Graphics/SVG/>) and wavelets image compression are two relatively new techniques that also promise to achieve high quality images with smaller file sizes. However, special browser plug-ins are usually required to display images and maps saved in these formats.

5.3 Techniques for Implementing Web Map Interactivity

5.3.1 Basic Imagemaps

Images referenced in an HTML file can be made to have “sensitive areas” or hotspots, defined by their bounding co-ordinates. Hotspots can be specified as rectangles, circles, or polygons. Clicking within a hotspot will usually cause the browser to launch an Internet location, e.g., another image or document. Different hoptspots can be tied to different links. In this way, a single image can provide multiple hyperlink destinations, each associated with a defined region or regions of the image.

5.3.1.1 Server-side Imagemaps

The `ismap` attribute for the HTML `img` tag can be used to turn an image into a graphically active element, so that clicking different regions on the same image causes the server to take different actions. The bounding co-ordinates of hotspots are not stored in the HTML file that calls the map image, but in a separate map file (a text file) stored on the Web server. This map file also includes the corresponding actions to be taken when these hotspots are clicked.

When a user clicks on a server-side hypermap, the browser sends the mouse co-ordinates to Imagemap, a program on the server that looks in the corresponding map file for the action associated with those co-ordinates. Imagemap first determines which hotspot among the defined hotspots these co-ordinates fall within. It then reads the action associated with this hotspot, and returns a server redirect message back to the browser telling it which Internet location or document it should access.

Imagemap is a CGI program that runs on the Web server. CGI stands for Common Gateway Interface. It is an interface definition that allows an HTML application to invoke a program on the server, like Imagemap, and pass arguments to it. CGI programs can be programmed in many languages like C, Pascal and PERL (Practical Extraction and Report Language).

In the following example, a GIF image, `mymap.gif`, is declared active using the `ismap` attribute:

```
<a href="http://www.myserver.com/cgi-bin/imagemap.exe/maps/mymap.map">
<img src="mymap.gif" ismap>
</a>
```

Note the path to the map file (`mymap.map`) associated with `mymap.gif`. When a user clicks over the image, the browser sends the mouse click co-ordinates to the server, with respect to the image origin (0, 0—top left corner of the image); the co-ordinates are appended to the map file address after a question mark '?' symbol as follows:

```
http://www.myserver.com/cgi-bin/imagemap.exe/maps/mymap.map?239,80
```

It is noteworthy that Microsoft ASP pages (Active Server Pages) use a similar technique to pass arguments to the server as in the following example:

```
http://healthcybermap.semanticweb.org/icd.asp?SearchText=493
```

The main disadvantage of server-side imagemaps is that they create extra data traffic between the server and its clients. That is why client-side clickable maps have evolved as a replacement to server-side maps [38, 56, 57].

5.3.1.2 Client-side Imagemaps

Client-side imagemaps store hotspot co-ordinates and associated hyperlink information in the same HTML document in which the image is referenced, not in a separate map file on the server. When the user clicks a hotspot in the image, the associated hyperlink location is determined by the Web browser software (from the underlying HTML code) and the user is transferred directly to that location. This makes client-side image maps faster than server-side imagemaps and reduces server load [38, 58].

Many imagemap editors exist that allow users to define hotspots on an image as well as the actions associated with these hotspots, and then automatically generate the necessary client-side HTML code (or server-side map file if needed), e.g., Mapedit (<http://www.boutell.com/mapedit/>). The client-side HTML code below defines the co-ordinates of a rectangular hotspot on a JPEG image, `myimage.jpg`, near its lower right corner. The description "City University, London, UK" should appear when the user moves the mouse cursor over the defined hotspot. Clicking the hotspot will take the user to City University Web site (<http://www.city.ac.uk>). Nothing will happen if the image area outside the hotspot is clicked.

```
<img src="myimage.jpg" usemap="#mymap" width="445" height="523"
border="0">
<map name="mymap">
<area shape="rect" alt="City University, London, UK"
coords="237,311,443,467" href="http://www.city.ac.uk">
<area shape="default" nohref>
</map>
```

WebView (<http://www.zebris.com/english/main_webview.htm>), the Internet extension to ArcView GIS used in HealthCyberMap, also generates client-side imagemaps form ArcView themes. WebView adds extra interactive functionality to these maps using JavaScript. JavaScript is embedded directly in HTML pages (as readable text) and is interpreted by the browser completely at runtime. JavaScript statements can respond to user events such as mouse-over and mouse-clicks. Web map ToolTips can be also implemented in Javascript. (VBScript is another popular scripting language by Microsoft and an alternative to JavaScript.)

Client-side imagemaps provide better accessibility compared to server-side imagemaps, because authors are able to assign appropriate text to each imagmap hotspot by including the `alt` attribute and area description inside each `<area>` tag.

This feature means that someone using a screen reader can easily identify and activate regions of the map (see <http://www.access-board.gov/sec508/guide/1194.22.htm#(f)>—[59]).

Client-side interactive maps can be also created in Macromedia Flash (e.g., <http://kartoweb.itc.nl/webcartography/webmaps/static/si-example4.htm>), Macromedia Shockwave (e.g., <http://kartoweb.itc.nl/webcartography/webmaps/dynamic/explov.htm>) or as a Java applet (e.g., Descartes system <http://allanon.gmd.de/and/java/iris/>). These types of client-side interactive maps might be slow to download [38].

5.3.2 Zooming

Web map zooming options depend on users' systems, including installed plug-ins (WebView, the tool used in HealthCyberMap does not require any browser plug-ins), and also upon the presence of enough map detail to allow considerable enlargement. There are three distinct zooming strategies or options [38]:

- Static linear zooming. The relation between zoom factor and map content is static. When zooming into the map, the image is linearly enlarged but the content of the map does not actually change. It can be done on the client side using an appropriate browser plug-in or applet. In this case the map is stored simply as an image. Vector-based images such as those in SWF (Macromedia Shockwave File) and SVG format, will keep their sharp character when enlarged, while raster-based images such as those saved in GIF or JPEG formats will show jagged edges. There is, however, an ideal scale (or scale range) to display any particular map, depending on the density and accuracy of map detail. If a map is enlarged too much very few details may be visible in the image window and the positional accuracy of the symbols may be much less than what the users expect at such scale.
- Static stepped zooming. In this case a series of maps of the same area is available, each one designed for a different scale or scale range. When the user requests to zoom in or out, the software automatically selects the most suitable map for the desired scale. This system is widely used on route planning sites and by companies such as MapQuest (<http://mapquest.com/>). It offers better results compared to static linear zooming, and is the approach undertaken by WebView.
- Dynamic zooming (animated scaling). In this zooming strategy there is a continuous direct relation between scale and map content. The larger the scale the

more detail is shown in the image. A direct link between the image and some kind of database is necessary. Although not always required, the cartographic symbolisation may change with scale. For instance a town represented by a point symbol at a small scale may turn into an area symbol upon zooming into the map. This approach is more expensive, requires more data, processing and bandwidth, and is not always needed.

(See also “Semantic Zooming”—Chapter 9.)

5.3.3 Advanced Solutions for Publishing Desktop GIS Maps and Data on the Web

Two main options exist for sharing desktop GIS maps and projects on the Web as sensitive clickable maps:

- dynamic publishing to the Web using a dedicated Internet map server that maintains a live connection with the underlying GIS project/ database; or
- publishing a static snapshot of the project (representing the project’s maps and underlying data *at time of publishing*) as clickable client-side imagemaps using tools like ImageMapper from alta4, Germany (<http://www.alta4.com/eng/products_e/im/im30/index_e.htm>) and WebView from Zebris, Germany (HealthCyberMap method is a modified version of this option to partially compensate for its limitations—see Chapter 9).

5.3.3.1 Dedicated Internet Map Server Solutions for Serving Maps with Dynamic Database Drill-down Functionality

Advanced mapping applications running on the server side can be linked to the server software, e.g., using CGI. These applications can be used to provide live database access (browsing/ querying a map database on the server). Using a dedicated map server, users could get a map depicting the latest figures from a database, which can come from another remote server, visualised with the colours and classification the user has requested [38, 60].

ESRI Internet Map Server solutions allow for an existing ArcView project to be transparently ported to the Web with minimal effort (updates carried on the original project in ArcView will also show automatically in real time in the Web front-end). Almost all major GIS vendors have already done this and although their approaches differ in detail, most use a combination of server-side and client-side components.

In one typical ESRI set-up, the ArcView GIS program takes on a role “similar to conventional CGI applications” running on the server. An ArcView extension, called

Internet Map Server (IMS), is installed to receive commands from the Web browser via the Web server. A command can be for example a map query. It will be passed to and processed by ArcView GIS and the result (a map view) will be converted to a GIF or JPEG file and sent to the browser. A Java applet called MapCafe is used to implement in the Web browser an interface similar to the standard ArcView GIS interface (Figure 5.3). Users can for example click the zoom button and drag a rectangle on the map displayed by the applet. This would result in the applet building a command to implement the required zoom action (IMS will receive this command and hand it to ArcView). The last item in this set-up is a plug-in to the Web server software called `esrimap.dll` that enables the server to find the appropriate ArcView GIS application to handle the request. ArcView GIS can be run on another computer to decrease server load and the server plug-in can even distribute requests among a multitude of computers running the same ArcView application. The Java applet can be customised and the IMS can handle all functionality within ArcView, including its built-in scripting language. This makes the system very flexible but also expensive and more difficult to set-up and run [38, 61].

ESRI Internet Map Server allows users to easily look up places on the map, e.g., by typing the names of the places they want to locate on the map. However, typing errors and disagreement about correct spelling of map features can severely limit the usefulness of such feature [38].

5.3.3.1.1 Costs Associated with Dedicated Internet Map Servers

Unfortunately, all these excellent features of dedicated Internet map server solutions do come at a cost:

- price (several thousands of US dollars);
- expertise is required to install, customise and manage the Internet map server;
- full access to the hosting Web server is required to install and manage software components (not always possible with mainstream (cheap) shared virtual hosting packages offered by most Web Hosting Providers; for example, HealthCyberMap current Web hosting package does not allow full access to the hosting server to install extra software); and
- ESRI MapCafe Java applet might be slow to download (depending on speed of client’s Internet connection).

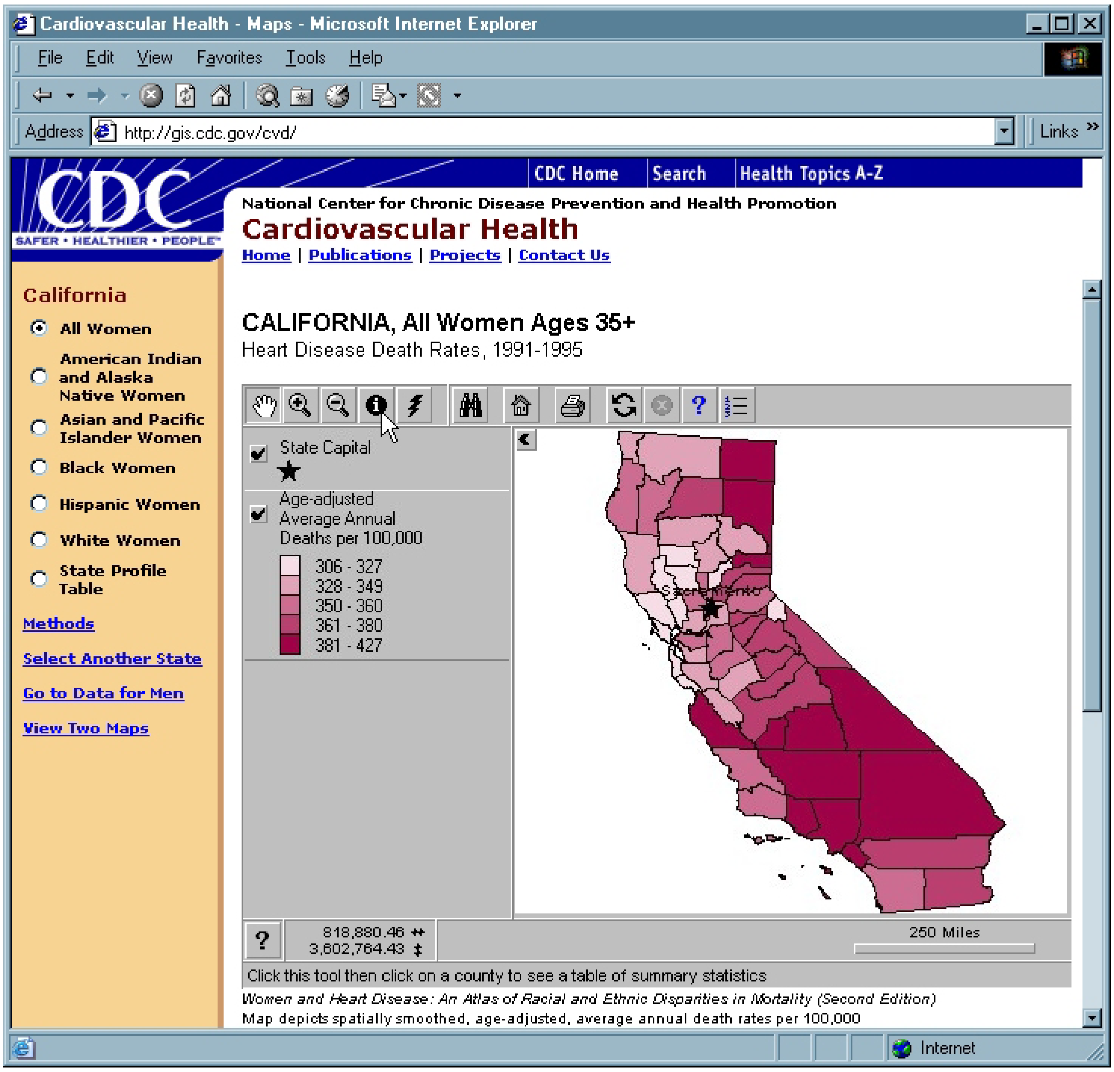


Figure 5.3. Screenshot from CDC Atlas of Heart Disease (<http://gis.cdc.gov/cvd/>) showing ESRI MapCafe Java applet in action. Users can select map layers to be displayed, pan, zoom, and view the attributes of features clicked using the identification tool.

In Part II, we describe a simple, low-cost way that the author developed for HealthCyberMap to serve hypermaps with dynamic database drill-down functionality (dynamic database links) without the need for a dedicated Internet Map Server.

5.4 Conclusion

Non-animated (static) interactive Web maps, also called clickable maps, imagemaps or hypermaps, can be served either as client-side imagemaps or as server-side imagemaps depending on where mouse-click co-ordinates are resolved.

Map comprehension can be enhanced with appropriate use of visual and typographic variables, and the application of sound cartographic design principles regarding cartographic symbols, map contrast and visual hierarchy, and map scale and abstraction. An important visual variable is value (differences in grey values or lighter/ darker shades of the same colour), which conveys differences in order or relative quantity and forms the basis of choropleth maps. Pictorial symbols are much easier to understand than geometric or abstract symbols, especially when a clear metaphorical relation exists between the symbol and the user's real world knowledge.

The ideal Web map should not be too large in both file size and image size to download quickly and cope with the limited size of display screens. The Web offers many techniques for adding interactivity to maps and responding to mouse events (e.g., zooming, panning and ToolTips on mouse-over for map feature labelling), which can be utilised to make smaller, "smarter" maps that are faster to download and provide additional functionality.

Image formats used in Web maps include GIF, which is more suitable for line art images and images with solid colours, and JPEG, the best compression algorithm for maps making considerable use of colour blends. Web map designers should test both formats on the maps they intend to produce, varying the various parameters available, including JPEG compression/ quality settings and GIF palette sizes to find which setting gives good results (usually assuming a 256-colour configuration) while maintaining a reasonably small file size.

There are three distinct zooming strategies: static linear zooming, static stepped zooming and dynamic zooming.

Currently two main options exist for sharing desktop GIS maps and projects on the Web as sensitive clickable maps. Dynamic publishing to the Web using a dedicated Internet map server that maintains a live connection with the underlying GIS project/ database is one option, but is an expensive and complex solution to acquire, run and maintain. The other cheaper and simpler option is to publish a static snapshot of the project as clickable client-side imagemaps representing the project's maps and underlying data at time of publishing (HealthCyberMap method is a *modified* version of this latter option to partially compensate for its limitations).

Part II: Core Service Methodology and Architecture

6 A Bird's-eye View of HealthCyberMap's Methodology and Architecture

"The ocean flows of online information are all streaming together, and the access tools are becoming absolutely critical. If you do not index it, it does not exist. It is out there but you cannot find it, so it might as well not be there."

—Barbara Quint, ASI San Diego Conference, 1994

HealthCyberMap comprises two main arms or layers. The top-level visualisation/ navigation arm (interface layer) is founded upon a robust semantic layer (Figure 6.1).

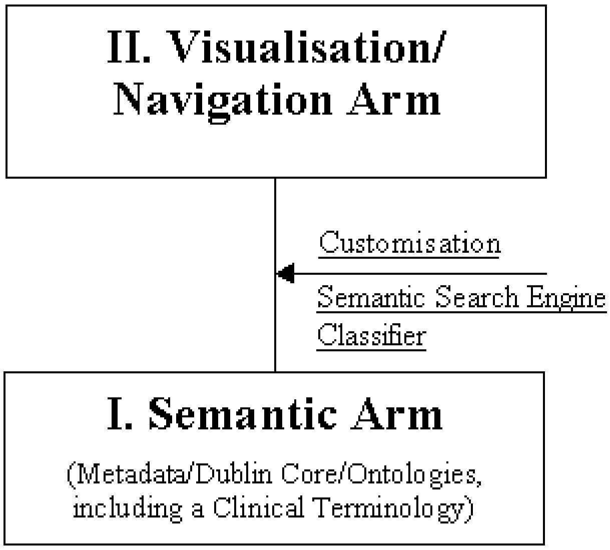


Figure 6.1. HealthCyberMap comprises two main arms or layers: a top-level visualisation/ navigation arm (interface layer) founded upon a robust semantic layer.

6.1 Semantic Arm

Pointers to good quality resources need to be described in a central metadata base and organised in such way to allow a browsing front-end (and in the future a sophisticated content management and customisation engine) to easily and suitably recall and re-use them in different browsing/ customisation scenarios.

Metadata are information about information resources and are totally transparent and invisible to the user. HealthCyberMap metadata base of resource pointers uses fields (elements) from the well-known Dublin Core (DC) metadata set scheme for resource description [22] with HealthCyberMap's own extensions to this scheme, including a field that has been introduced to store information about resource quality (Chapter 7).

A DC language field, for example, makes possible the selection of resources based on their language to match a user's preferred language, while a DC type element allows the classification and retrieval of resources by type, e.g., fact-sheets, guidelines, e-books, etc.

Multiple cardinality is allowed for the DC subject field, enabling richer descriptions of resource topics. Codes from a suitable clinical terminology (domain ontology) are used to populate this field in HealthCyberMap's resource metadata base. This conforms to DC recommended best practice of using a controlled vocabulary for this element, and greatly improves HealthCyberMap's precision and recall rates. It also forms the basis of HealthCyberMap Medical Semantic Subject Search Engine (Chapter 8) and plays a pivotal role in the provision of clinical problem to knowledge linking services (e.g., linking clinically-coded electronic patient record problems to appropriate Web resources answering them—Chapter 10).

6.1.1 Beyond Direct Metadata Queries: Mapping User Queries and Resource Metadata to a Brokering "Intelligent" Ontology

HealthCyberMap Medical Semantic Subject Search Engine must reliably "understand" the semantics of a user query, as well as the semantics of available resources that might contain an answer to this query in order to return semantically superior search results. It should be able to infer synonyms, implicit semantic relationships and contexts [62] not directly mentioned in user queries and/or resources/resource metadata (since it is not practical or computationally efficient to explicitly encode all synonyms, semantic relationships and other possibilities of related topics in resource metadata). This can only be done if queries and resources are mapped to the same domain ontology (an appropriate clinical coding scheme or vocabulary), so that these synonyms and semantic relationships can be made explicit and available for intelligent semantic information retrieval (Figure 6.2). A related approach has been described by Leroy et al [63].

A code (concept) location in a clinical vocabulary has its semantic topology preserved, i.e., it "knows" where it is and its semantic links to its surroundings (and of course has all possible descriptions or synonyms attached to it, possibly in more than one natural language). Some semantic relationships are explicitly mentioned, while others can be inferred by inheritance.

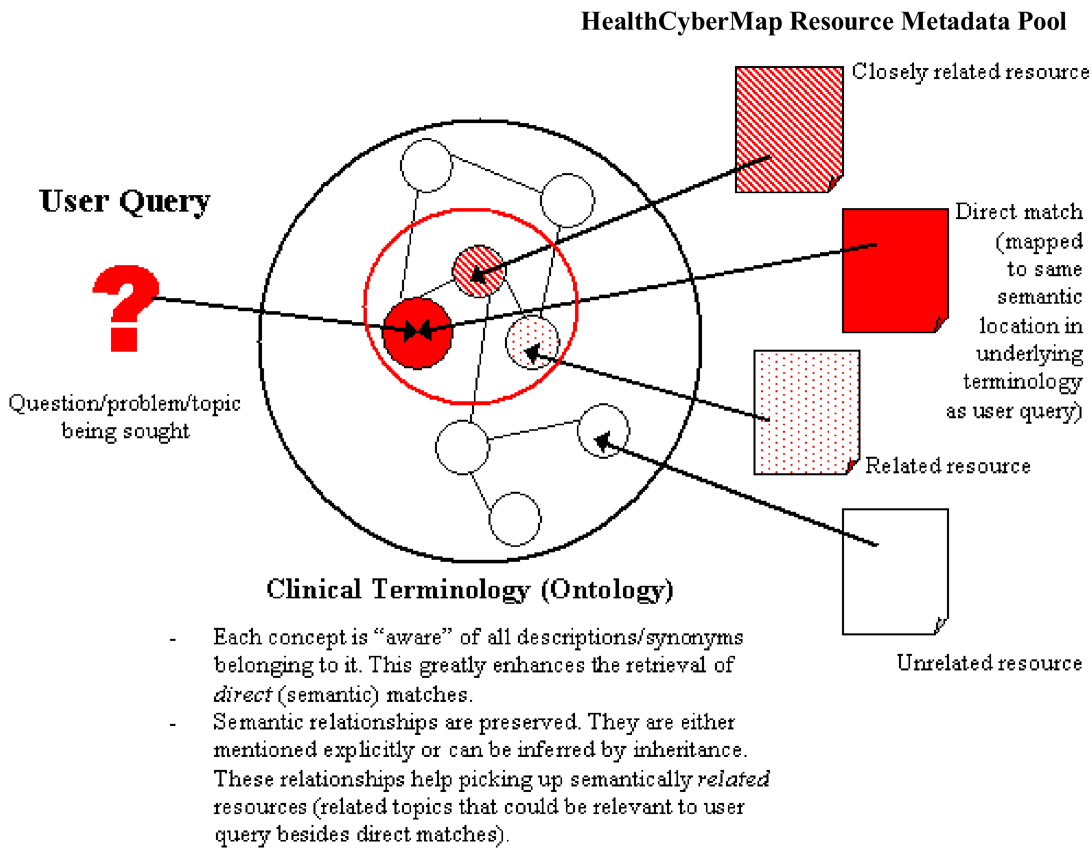


Figure 6.2. Explicit concepts in resource metadata map onto a domain ontology (a clinical terminology or classification) allowing a Semantic Web search engine to infer implicit meanings (synonyms and semantic relationships) not directly mentioned in either the resource or its metadata. Similarly, user queries would map to the same ontology allowing the search engine to infer the implicit semantics of user queries and use them to optimise retrieval.

6.1.2 Requirements for Resource Selection and Indexing

Resources selected for HealthCyberMap should be:

- pooled according to well-defined selection criteria/ editorial policy, using suitable quality benchmarking tool(s) or checklist(s) [64]. Manual selection ensures quality of selected resources;
- indexed using *only the most specific (narrow) clinical concept codes* corresponding to their topics, picked from a suitable clinical vocabulary (domain ontology) to ensure precision of topical indexing [65]. This approach (using suitable clinical codes for the topical indexing of resources) is far more superior than conventional indexing using textual keywords. A resource topic code can be always traced to the underlying clinical domain ontology to care for different topic synonyms and relationships to other topics of interest within the topic's semantic

neighbourhood (see Figure 6.2 above), and to allow automatic topic classification under all relevant broader categories (the human cataloguer is *not* involved in this topical categorisation process). Some tools exist that can partially automate this task and help human indexers select the appropriate clinical code(s) for a resource at hand, e.g., code browsers/ locators and more "intelligent" tools that scan a resource and suggest possible clinical concept codes that could describe its topic(s). Big sites should not be indexed as a single resource except when they cover a single narrow topic. Whenever possible, individual pages and subsections from these large sites covering individual (specific) topics should be indexed as distinct resources;

- structured by filling a complete DC metadata record for each selected resource, describing all aspects of that resource, e.g., content, currency information (dates), format, etc.;
- organised—this involves reasoning with the compiled resource metadata records to categorise/ classify HealthCyberMap's resource pool into groups (e.g., topics/ profiles of related topics) for themed navigation. The same resource can be included in more than one navigation path, map, or ontology. Themed navigation ontologies (maps of links) are dynamically generated and kept separate from resources. Compared to conventional hard-coded links, this approach ensures maximum flexibility and ease of maintenance (when resources are added, deleted, or their address changes) [66]; and
- maintained for currency (up-to-dateness) and persistence by regularly using a broken link checker.

6.2 Visualisation/ Navigation Arm – User Interface

HealthCyberMap's pool of resource metadata together with the semantics of the underlying clinical coding scheme form the basis of the service's navigational hypermaps (Chapter 9), which are used to semantically browse collections of health-related resources on the Web in intuitive graphical ways.

Designing and building an appropriate user interface, including proper navigational aids like visual maps that use suitable metaphors for browsing available resources, are very important ingredients for the success of the service. Principles of Web interface usability must be observed [67].

Suitable tools must be selected for building and maintaining such interface. Dynamic pages, e.g., Microsoft ASP—Active Server Pages, can be designed as templates to be filled in and customised on the fly with appropriate content from the underlying metadata base according to user's needs and preferences.

6.2.1 GIS as HealthCyberMap's Visualisation Engine

Geographic Information Systems (GIS) are robust and reliable tools, optimised for handling, cross-linking and visualising data with spatial and/ or spatialised components. In GIS, data semantics and visualisation are separated for maximum flexibility, but remain tightly coupled [47]. This corresponds to HealthCyberMap's core architecture model, making GIS a perfect choice as a visualisation engine.

6.3 Benefits Arising from the Separation of the Semantic and Navigation Layers

The separation of the navigation/ browsing interface from the underlying semantic infrastructure (separation of links from resources) is a very powerful feature of HealthCyberMap making it possible to introduce/ experiment with different navigation techniques, and even completely change the user interface and associated technology at any time (if needed) without touching the underlying resource metadata base.

Moreover, if a resource address changes, only its metadata record needs to be updated (no other documents need updating). The next time this resource is referenced (in any dynamically generated browsing front-end), the reference will point to the correct link. With this architecture, broken links that used to bug pre-compiled, hard-coded maps and lists of resource pointers become a thing of the past. Also, if a resource on a given topic is added to the metadata base (or deleted from it) and a query is run for this particular topic, this resource will be automatically listed in any generated browsing front-end (or disappear from it) [66].

HealthCyberMap's architecture also allows the service to be made available to other external clients besides HealthCyberMap itself as a Semantic Web Service (see <http://www.w3.org/2002/ws/>). Each external client can apply its own navigation/ interface layer (see Chapter 10).

6.4 Conclusion

HealthCyberMap comprises two main arms or layers: a top-level visualisation/ navigation arm (interface layer) founded upon a robust semantic layer. The semantic arm is essentially formed of metadata, and aims at making the context and meaning (semantics) of health and medical Web resources amenable to computer processing to improve HealthCyberMap's precision and recall rates.

The navigation arm features navigational aids (e.g., visual maps based on suitable metaphors) for browsing HealthCyberMap's pool of resource pointers. It relies on GIS as the visualisation engine. It also uses dynamic ASP pages to query and present results from the underlying metadata base (semantic arm).

The separation of the navigation/ browsing interface from the underlying semantic infrastructure is a very powerful feature of HealthCyberMap making it possible to experiment with different navigation techniques without touching the underlying resource metadata base.

Moreover, with this architecture, service maintainability is much improved, rendering broken links that used to bug pre-compiled, hard-coded maps and lists of resource pointers a thing of the past.

The same architecture also makes HealthCyberMap's pool of resource pointers and semantic search available to other external clients as a Semantic Web Service.

Part III: Service Implementation and Description

7 HealthCyberMap's Metadata Base of Medical/ Health Information Resources Based on the Dublin Core Metadata Set

This chapter describes the work undertaken to build HealthCyberMap's Semantic Arm (see Chapters 3 and 6). The author started by developing a model RDF (Resource Description Framework) metadata base based on the qualified Dublin Core (DC) metadata set. This approach based on RDF and the qualified DC metadata set offers many opportunities in the future, including the possibility of other services making use of HealthCyberMap's RDF metadata base to develop their own Web portals. However, for reasons explained below, the current HealthCyberMap pilot service uses a simpler metadata base developed in Microsoft® Access based on the non-qualified Dublin Core metadata set. HealthCyberMap's editorial policy for resource selection and quality benchmarking is also discussed.

7.1 HealthCyberMap's Tool for Building an RDF Metadata Base Based on the Qualified Dublin Core Metadata Set

7.1.1 Basic DCMI Metadata Packages

The author modelled the qualified DC metadata set based on three currently approved DCMI (Dublin Core Metadata Initiative) metadata packages with the following namespaces:

- <http://purl.org/dc/elements/1.1/>
- <http://dublincore.org/2000/03/13/dcq#>
- <http://dublincore.org/2000/03/13/dctype#>

7.1.2 DC Qualifiers

DCMI describes two broad classes of DC qualifiers which are also preserved in HealthCyberMap's model ontology and tool described below [23, 68]:

- Element refinement. These qualifiers make the meaning of an element narrower or more specific, e.g., the "table of contents" and "abstract" element refinements for the DC description element. A refined element shares the meaning of the unqualified element, but with a more restricted (specialised) scope.
- Encoding scheme. These qualifiers identify schemes that aid in the interpretation of an element value. These schemes include controlled vocabularies and formal notations or parsing rules. A value expressed using an encoding scheme will thus

be an instance selected from a controlled vocabulary, e.g., a term from a medical terminology or a classification, or a string formatted in accordance with a formal notation (e.g., "2002-03-01" using ISO 8601-based W3C date encoding rules [69]). This helps preventing any ambiguities and making the string machine-understandable (in the last example, the machine can be sure that we only mean 1st of March 2002).

DCMI is supporting the use of RDFS/RDF to express DC [68, 70, 71]. RDF metadata could be reasoned with in more intelligent semantic ways and are ideal for data exchange (being serialised as XML). For these reasons, the author decided to save HealthCyberMap's ontology in this format.

7.1.3 Extending DC

DC is not a complete metadata solution. For example, DC cannot be used to describe the quality [64] or location of a resource (location is different than DC coverage, although both are geographic elements). There are no DC elements covering these important aspects of a resource. HealthCyberMap attempts to fill these gaps using its own elements.

The author added HealthCyberMap own extensions, "quality" and "location", to the standard DC elements. The slot value for location stores the resource publisher or author(s)' main geographical location, whichever is more relevant, and takes the form of an ISO3166 country code, plus a city value from the Getty Thesaurus of Geographic Names (TGN – <http://www.getty.edu/research/tools/vocabulary/tgn/>).

This approach of patching DC is not completely new. For example the W3C RDFPic project for describing and retrieving digitised photos with RDF metadata extends the DC schema by adding its own technical schema to define important elements not covered by DC, e.g., camera, film, lens and film development date [72].

7.1.4 HealthCyberMap Three-Layer Model

HealthCyberMap has been designed to use clinical codes like ICD-9-CM (International Classification of Diseases, ninth revision, Clinical Modification) as basic (template) medical ontologies for mapping the health cyberspace (the codes are used to populate the DC subject field in HealthCyberMap metadata base – Figure 7.1).

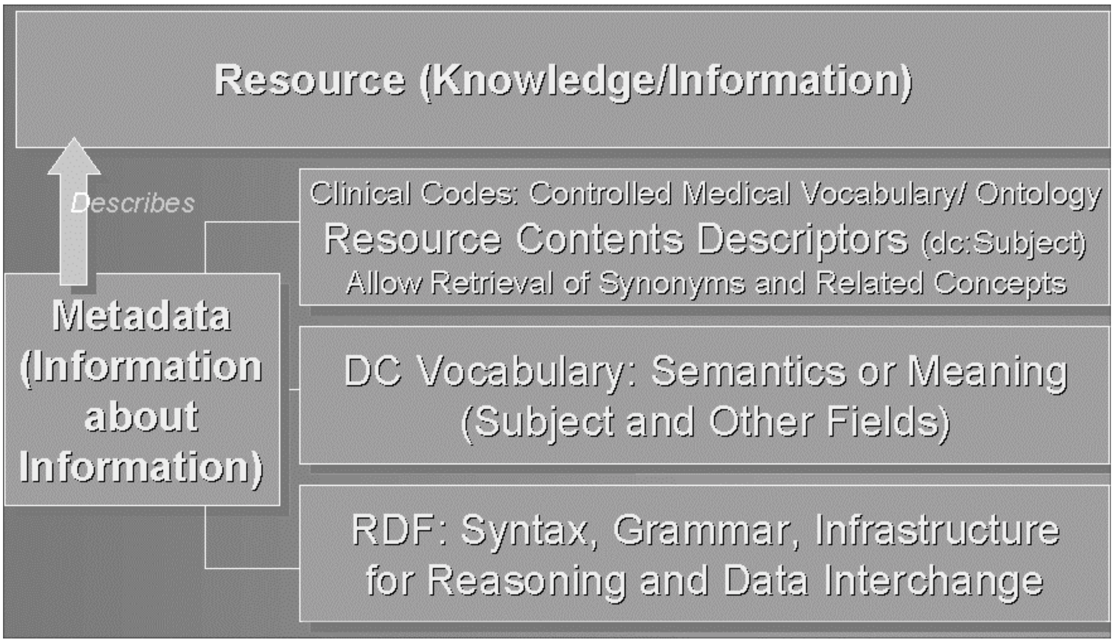


Figure 7.1. Clinical codes ontology/ DC vocabulary/ RDF interrelation. HealthCyberMap DC RDF instances use clinical codes to populate the DC subject field.

The author was able to adopt UMLS (Unified Medical Language System) for HealthCyberMap's DC model in Protégé-2000 (see below) alongside ICD-9-CM, thanks to DC Subject Scheme Qualifiers, which prevent any ambiguities that might arise when using more than one clinical coding scheme at the same time. UMLS provides a kind of mapping between many coding systems.

7.1.5 Tools Used: Protégé-2000 and the UMLS Tab

Ontology modelling languages and tools like Protégé-2000 (<http://protege.stanford.edu/>) supply the modelling primitives necessary to provide adequate power of expression and clarity and to support the function of inference and reasoning services. These primitives are based on the notions of classes (concepts or frames) having properties (roles or slots). Slots can have their own properties (constraints), e.g., domain and range, and can be arranged in a hierarchy (i.e., subslots).

Protégé-2000 allows users to construct a domain ontology, customise knowledge-acquisition forms for it and enter domain knowledge structure and instances. It can be extended with many useful plug-ins like the UMLS tab [73].

Protégé-2000 supports hybrid ontologies (see Chapter 3) and saves class definitions and instances in two separate files (under the same project). A Protégé JDBC (Java Database Connectivity) database back-end permits the storage of large Protégé-2000 knowledge bases in a relational database for speedier access. Users can also (at least theoretically) access the same database outside Protégé-2000 environment to benefit from the faster access and search facilities that a relational database offers. However, this latter situation is not encouraged since Protégé encodes its knowledge bases in a very specific way making access from other software components very difficult. An RDF Schema backend plug-in is also available that allows saving and opening projects in RDFS and OIL formats (RDF Schema and Ontology Inference Layer—see Chapter 3) [73, 74].

Noy and colleagues [75] propose using Protégé-2000 as an editor for Semantic Web languages like RDFS and OIL, as well as an interchange/ translation module between these languages. They also suggest that developers should create their own Protégé-2000 tab plug-ins to include custom Semantic Web applications that can benefit from the live connection to the knowledge base in Protégé-2000.

The UMLS tab is a handy Protégé-2000 plug-in that connects from within Protégé-2000 to the UMLS Knowledge Source Server (KSS) of the US National Library of Medicine (to protect the rights of the providers of some component vocabularies, a signed license agreement is required in order to use or access UMLS content; the user must then register with the UMLS KSS—<http://umlsks.nlm.nih.gov/>). The tab allows browsing and searching UMLS, and directly annotating an ontology in Protégé-2000 with elements imported from UMLS (Figure 7.2) [74].

HealthCyberMap DC implementation in Protégé-2000 uses UMLS terms as slot values for the DC subject field (Figures 7.3 and 7.4). Thanks to the UMLS tab, any imported element instance from the remote UMLS knowledge base becomes a permanent part of HealthCyberMap DC ontology, available for annotating (populating the DC subject field of) any number of current or future HealthCyberMap Resource instances, even when there is no more connection to the UMLS KSS. This virtually eliminates any need to manually copy, paste and/ or type terms from the remote knowledge source [74].

Other clinical coding systems, e.g., ICD-9-CM, are also allowed in our ontology. The *qualified* DC provides the necessary mechanism for using more than one coding scheme within the same system without causing any confusion. Each additional coding scheme is defined in its own subclass under DC SubjectScheme class.

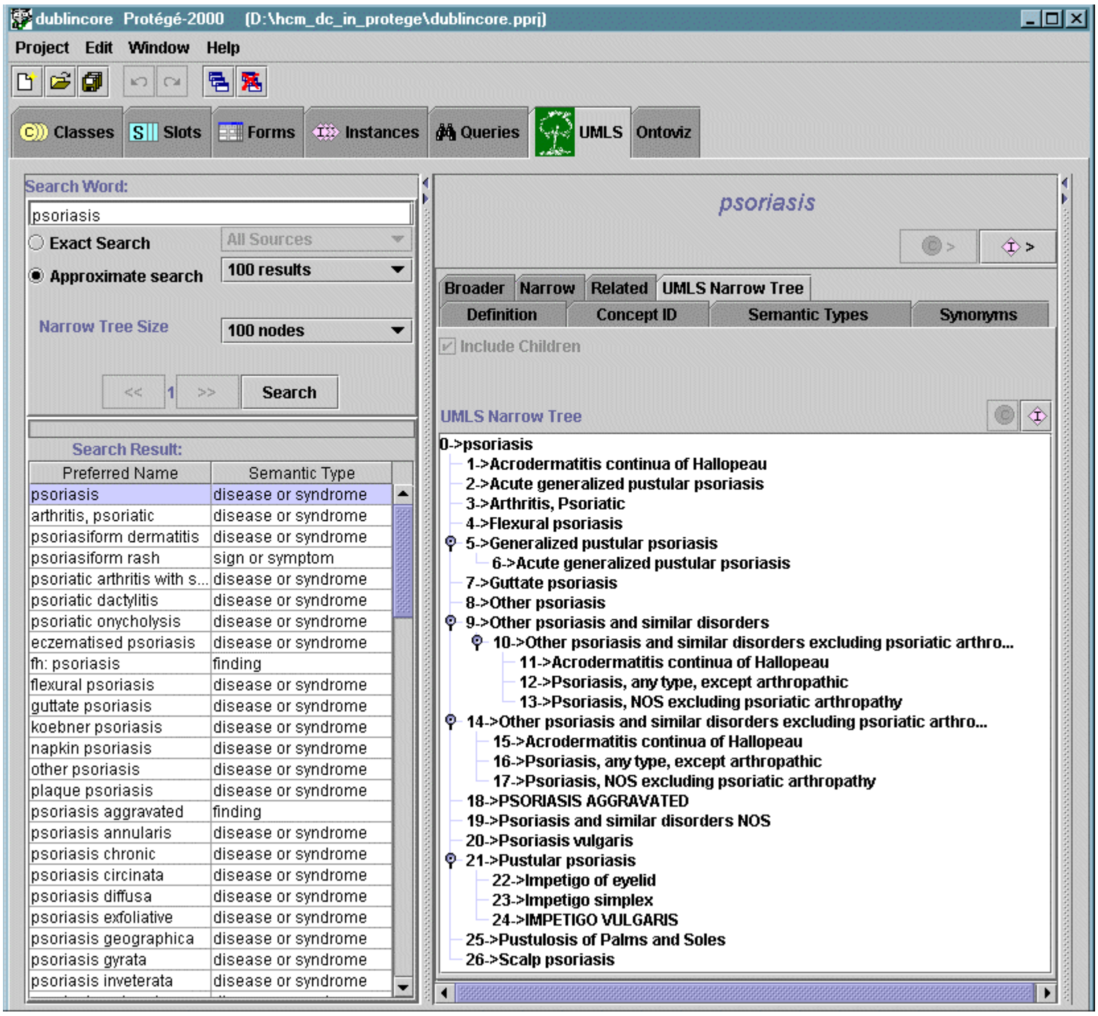


Figure 7.2. Screenshot of Protégé-2000 showing UMLS search results/ narrow tree for "psoriasis" in the UMLS tab.

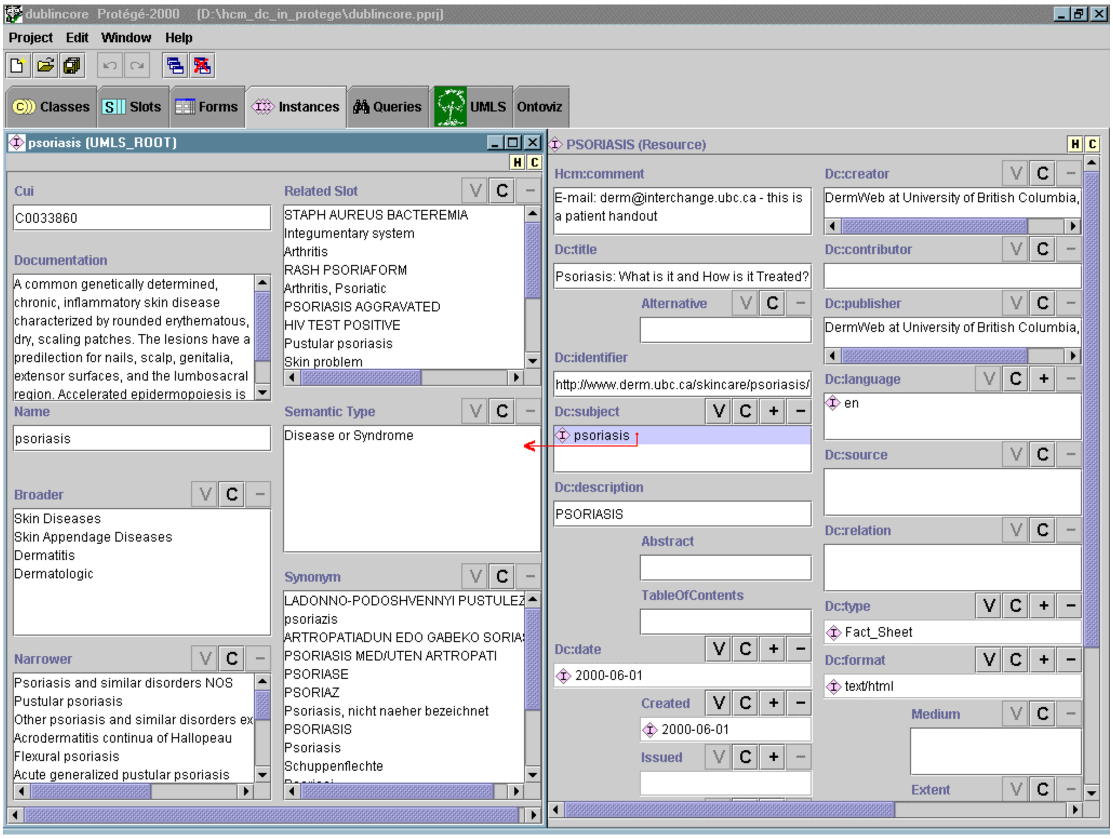


Figure 7.3. Screenshot of a Resource instance form from HealthCyberMap's DC project in Protégé-2000. The DC subject value is a UMLS term instance; double-clicking it displays all properties and relationships among other terms for this term ('psoriasis' in this example). To improve the quality of metadata, one should select the most specific term(s) that best describe a resource, avoiding whenever possible broader/ more general terms.

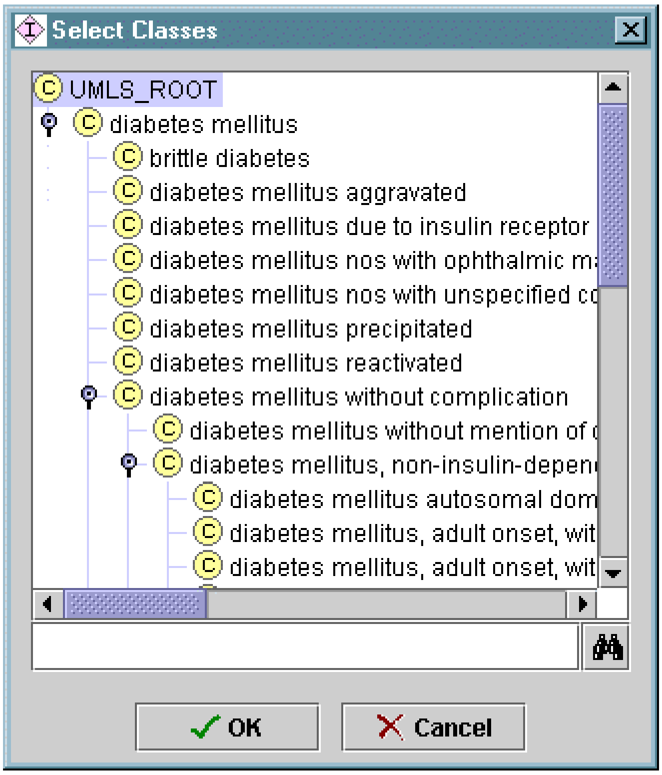


Figure 7.4. UMLS terms can be also imported as classes, preserving their tree (clinical) hierarchy and allowing easier navigation of the imported terms when selecting a value for the DC subject slot. Note that in Figure 7.3 above 'psoriasis' has been imported as a UMLS instance. (The DC subject slot can be set to take instances or subclasses of UMLS_ROOT depending on how we have opted to import UMLS terms.)

7.1.6 HealthCyberMap DC Implementation in Protégé-2000

The author defined a Resource class (template Bibliographic Card) in Protégé-2000 v1.6 (Figure 7.5). Instances of this class collectively form HealthCyberMap database of resource metadata. Each instance describes a single resource using the DC and HealthCyberMap elements that have been defined.

Custom data entry forms (knowledge acquisition forms) were also developed using the standard Protégé slot widgets to acquire instance data for the different classes in this project (Figure 7.3).

In Protégé, classes and instances of classes can act as slot values for instances of other classes. More than half of the slots in class Resource take instances of other classes as their values. For example, the author has defined language codes once as instances of ISO639-2 Language Scheme, and then used them to populate the `dc:language` slot of all instances of class Resource (Figure 7.6).

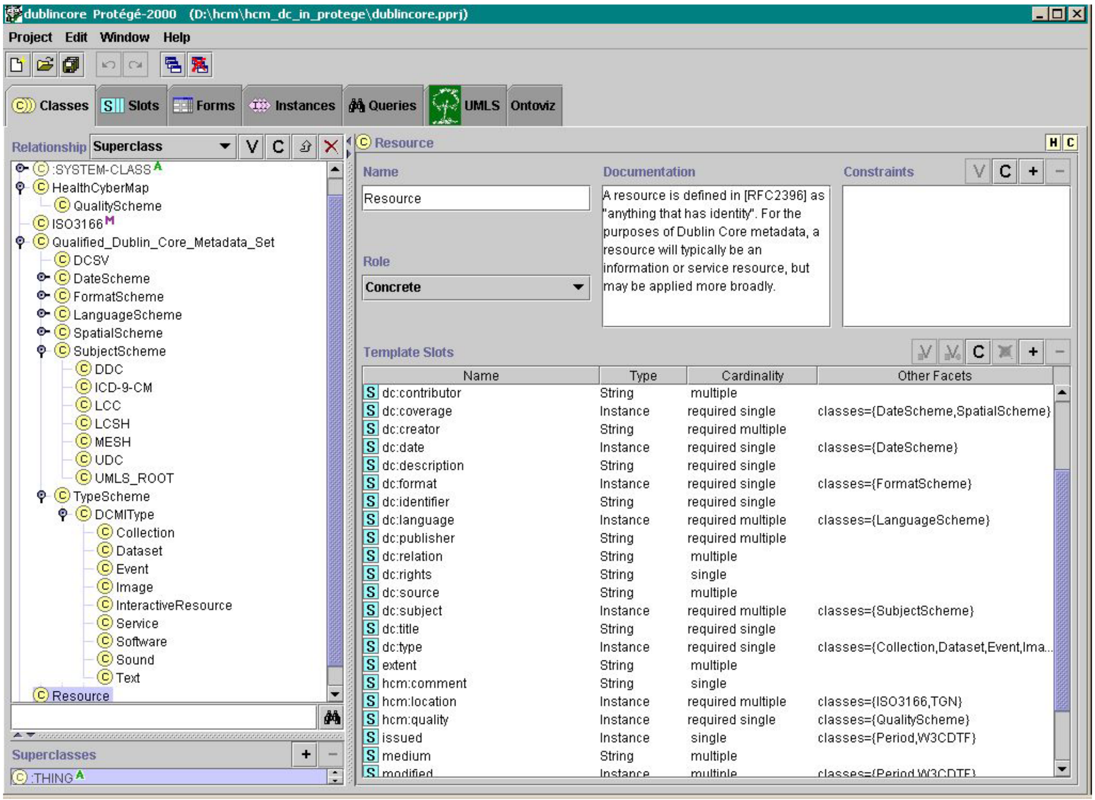


Figure 7.5. Screenshot of HealthCyberMap's DC project classes in Protégé-2000. The cardinality of some fields has been set to multiple, e.g., the DC subject field can take more than one value per resource but at least one value must be present (required). This results in richer resource metadata and is allowed in the original DC specification.

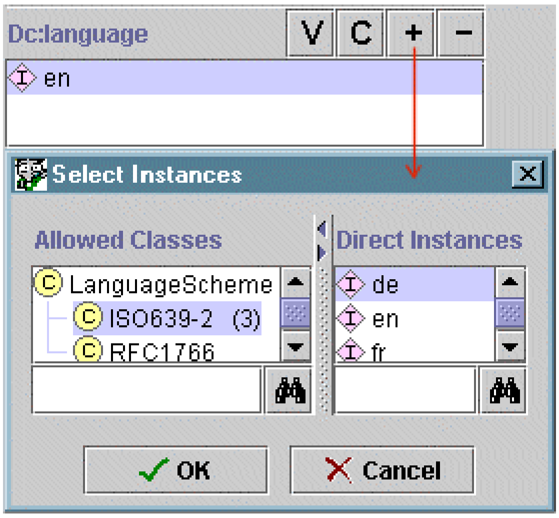


Figure 7.6. Language codes were defined once as instances of ISO639-2 Language Scheme, and then used to populate the `dc:language` slot of all instances of class Resource.

Protégé-2000 can also automatically generate HTML Documentation for a project's classes and subclasses, and their slots. The HealthCyberMap/ DC Protégé-2000 ontology described above and its Protégé-generated HTML documentation can be downloaded from: <http://protege.stanford.edu/ontologies/dublincore/hcm_dc_in_protege.zip> (requires Protégé-2000 v1.6 or later).

Protégé-2000 RDF output for HealthCyberMap DC project comprises an *.rdfs* file for the RDF schema (definitions of classes and slots) and an *.rdf* file for the RDF instances (where the actual resources are described; this is the equivalent of database records), plus Protégé's own *.pprj* project file, which contains information about the project's forms and the slot widgets they are using.

Protégé-2000 proved to be a potentially good tool for modelling and populating an RDF metadata base, though some users have argued its capacity to handle large RDF databases efficiently. The UMLS tab provides a friendly way for selecting DC subject field values; the connection to the UMLS KSS through this tab is very transparent and unnoticed by the user.

7.1.7 Practical Application of HealthCyberMap's Qualified DC Metadata Base in RDF

With a new Open Source version of Protégé's UMLS tab now available for downloading (<http://protege.stanford.edu/plugins/umlstab/umls_tab.html>), the full functionality of the described qualified Dublin Core ontology should become available to everyone as a free tool for building and maintaining an RDF medical/ health resource metadata base. In HealthCyberMap's model and tool, each instance holds the metadata record of a single Web resource following the blueprint of template classes. Users can keep adding resource metadata instances to build their metadata base. Thanks to the UMLS tab, users can search for concepts that best describe a resource, browse them in a tree view and directly import appropriate concepts to the DC subject fields. The resultant metadata base can then be shared and used with a search and inference engine, and have a textual and/ or visual navigation interface applied to it, to ultimately build a medical/ health Semantic Web portal. It is noteworthy that the Open Directory Project distributes its data of hand-categorised Web sites in RDF format (<http://dmoz.org/rdf/Changes.html>) [76]. Users then apply their own search, inference and textual/ visual navigation mechanisms to the distributed data to produce different added value interfaces to the same core data (e.g., <http://map.net/start>).

7.1.7.1 Exploiting Protégé-2000 RDF Output

Two main approaches have been proposed for querying RDF metadata [77]:

1. the SQL/XQL (Structured Query Language/ eXtensible Query Language) style approach, viewing RDF metadata as a relational or XML database (HealthCyberMap currently stores resource records in an Access database and queries them using SQL); and
2. viewing the Web described by RDF metadata as a knowledge base and, thus, applying knowledge representation and reasoning techniques on RDF metadata.

The first approach builds on the fact that the basic RDF model maps very directly to the relational database model (a record is an RDF node, the field (column) name is RDF propertyType, and the record field (table cell) is a value) [78].

The second approach is the one currently supported by the W3C RDF founders and working group, as well as other Semantic Web researchers. An RDF query language should provide facilities for simple property-value queries, path traversal based on the RDF graph model and more complex logical queries, deductions and inferences [77].

A new W3C querying language, Metalog, has been proposed that attempts to fulfil these criteria [79]. SiLRI (Simple Logic-based RDF Interpreter), one of the most cited RDF inference engines, represents another attempt. SiLRI implements a major part of Frame-Logic functionality [80]. There are also many other RDF query and inference languages like TRIPLE (the successor of SiLRI), RQL, SquishQL and RDFQL [81].

7.2 HealthCyberMap's Metadata Base in Microsoft® Access Based on the Non-qualified Dublin Core Metadata Set

ESRI ArcView GIS and its BodyViewer extension used to generate most of HealthCyberMap's navigational hypermaps (Chapter 9) cannot use an RDF data source and only work with ICD-9 codes not UMLS. Although possible, switching to Protégé's database backend to save the metadata base in some standard relational format compatible with ArcView GIS like Microsoft® Access is better avoided, since Protégé encodes its knowledge bases in very specific ways making access from other software components very troublesome. For these reasons and until standardised RDF databases and inference engines become widely available to make the potential benefits of RDF a robust reality, the author preferred to re-create HealthCyberMap's metadata base in Microsoft® Access. This "porting" to Microsoft® Access was straightforward. The qualified DC metadata set is needed to enable easy and reliable interpretation of element values in mainly two situations:

- when two or more encoding schemes co-exist in the same metadata base for a given metadata element; and/ or
- when the metadata base is shared with other services (outside HealthCyberMap).

As the above two scenarios do not apply to the current HealthCyberMap research pilot service, it was decided to base HealthCyberMap's metadata base in Microsoft® Access on the non-qualified DC metadata set for the sake of simplicity.

7.2.1 HealthCyberMap's Bibliographic Card

HealthCyberMap's metadata base was implemented in Microsoft® Access 97 based on the Bibliographic Card fields described below. This is the same database currently running on HealthCyberMap server, which users query by clicking the hypermaps (Chapter 9).

The current HealthCyberMap Bibliographic Card (resource metadata record) includes the following fields from the DC metadata set scheme [22] with HealthCyberMap's own extensions for resource quality and geographical provenance (dc=Dublin Core; hcm=HealthCyberMap-specific extensions):

```
dc:Creator="Author(s) name(s)"
dc:Title="Resource title"
dc:Subject (1, 2, 3)="ICD-9-CM Code, e.g., E948.2"
dc:Description="Textual equivalent (description) of ICD code, e.g.,
                CHOLERA VACCINE"
dc:Publisher="e.g., WHO"
dc:Date="Resource date of last update"
dc:Type="The category of the resource, e.g., electronic article, fact
         sheet, electronic journal paper, collection, e-book, digital
         atlas, audio-visual material, interactive resource, event,
         software, other online health service"
dc:Identifier="Resource URI"
dc:Language="e.g., en"
dc:Coverage="The spatial extent or scope of the content of the
             resource in the form of a spatial location (a place name
             or geographical co-ordinates from the Thesaurus of
             Geographic Names)"
hcm:Location (City, Country)="Publisher or author(s) geographical
                              location, whichever is more relevant"
hcm:Quality="Level of evidence (official guideline, systematic
             review, RCT, other peer-reviewed studies, official CAT,
             expert opinion), recognised code of ethics or quality
             seal, Trusted Publisher or Listed in Trusted Directory,
             e.g., OMNI, or unspecified"
hcm:Comment="Any additional information"
```

HealthCyberMap allows for three DC subject fields per resource record permitting up to three ICD-9-CM codes to be used to describe the topic(s) of each selected resource. The author did not distinguish between public, patient and professional audiences on purpose; the boundaries are now very hazy in this open world of free information and patient-centred care, self-help and self-care. The three audiences are partners in healthcare. The current trend in the NHS (UK National Health Service) is to consider health-related knowledge as a single blob (or pool of knowledge) that is relevant to both healthcare professionals and patients, and so should be made accessible to both of them equally without any distinction. Considerable numbers of patients are very well educated; these patients have more knowledge and understanding of their own conditions than do their treating general practitioners. Patients should be given more information and control of their conditions. We should not estimate patients' needs for information based on their looks, age, social class or the way they are dressed [82]. However, most respondents to HealthCyberMap's formative evaluation questionnaire thought it would be useful in a future implementation of HealthCyberMap to also organise information resources by intended primary audience as patients, health professionals, or basic researchers (see Chapter 12). To implement this request, another additional field will be required in the future to extend the DC set and store information about the intended primary audience of a resource.

7.2.2 Selecting Resources and Populating the Metadata Base

HealthCyberMap research pilot service only maps a limited number of health and medical information providers world-wide and just part of their resources. A resource is defined in RFC2396 (<http://www.ietf.org/rfc/rfc2396.txt>) as "anything that has identity". For the purpose of HealthCyberMap, a resource will typically be an information resource. [N.B.: Requests for Comments (RFCs) are a series of notes that started in 1969 about the Internet (originally ARPANET—US Advanced Research Projects Agency Network).]

Aided by conventional Web directories and search/ meta-search engines, candidate Internet resources were hand-selected by the author. Their metadata attributes, including their URI, ICD-9-CM code(s) representing their subject(s) and any recognised quality/ code of ethics rating they bear (e.g., a Health On the Net Foundation HONcode seal—<http://www.hon.ch>), were manually compiled in

HealthCyberMap's metadata base using a resource metadata entry form that the author created in Microsoft® Access (Figure 7.7).

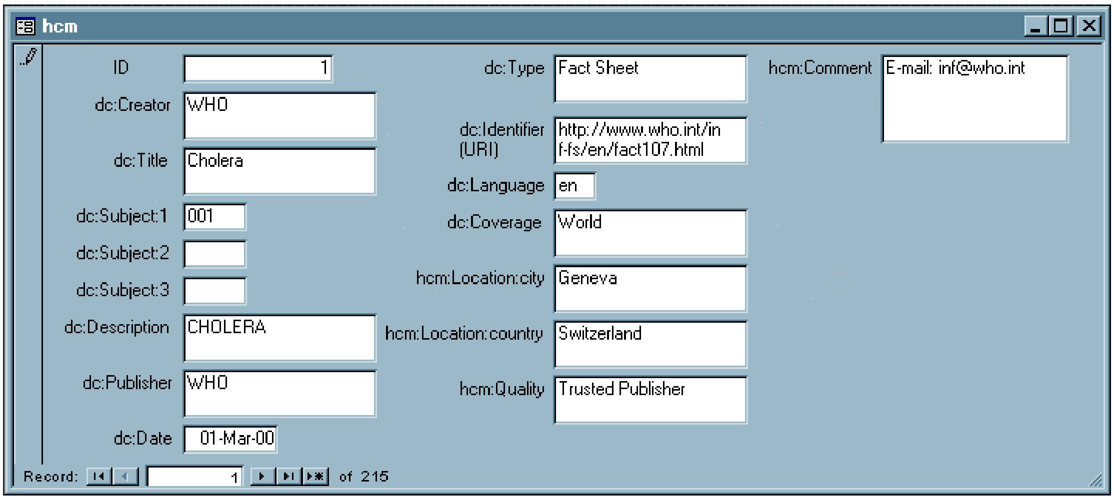


Figure 7.7. HealthCyberMap resource metadata entry form in Microsoft® Access 97.

There are 1640 resource records in HealthCyberMap's pilot metadata base (at the time of writing) covering a wide and assorted range of medical/ health topics across the full ICD-9-CM spectrum. The author did not treat big sites as a single resource (e.g., <http://www.orpha.net>—rare, congenital and hereditary diseases), but rather indexed individual pages/ collections from these sites that cover individual (specific) topics as distinct resources (e.g., <http://orphanet.infobiogen.fr/Site/Exp.stm?Lng=FR&Expert=281>—Cri-du-chat syndrome) to increase the relevance of HealthCyberMap's answers to user information needs.

7.2.2.1 Locating Suitable ICD-9-CM Codes

The author used two advanced online ICD-9-CM code locators that support rich code descriptions and synonyms: Yaki Technologies' ICD-9-CM search engine (<http://www.eicd.com/>) and e-MDs ICD-9 Search (<http://www.e-mds.com/icd9/>—Figure 7.8) to locate codes that best describe topics covered by a given resource (procedural topics, e.g., colonoscopy, were excluded in this research pilot of HealthCyberMap).

N.B.: ICD-9-CM is based on the World Health Organisation's (WHO) Ninth Revision of the International Classification of Diseases (ICD-9), a statistical hierarchical classification. ICD-9-CM is the current official system used to code and classify morbidity data from hospital inpatient and outpatient records and physician offices in the United States. The National Centre for Health Statistics (NCHS), American Public Health Service and the Health Care Financing Administration are the US governmental agencies responsible for overseeing all changes and modifications to the ICD-9-CM [83].

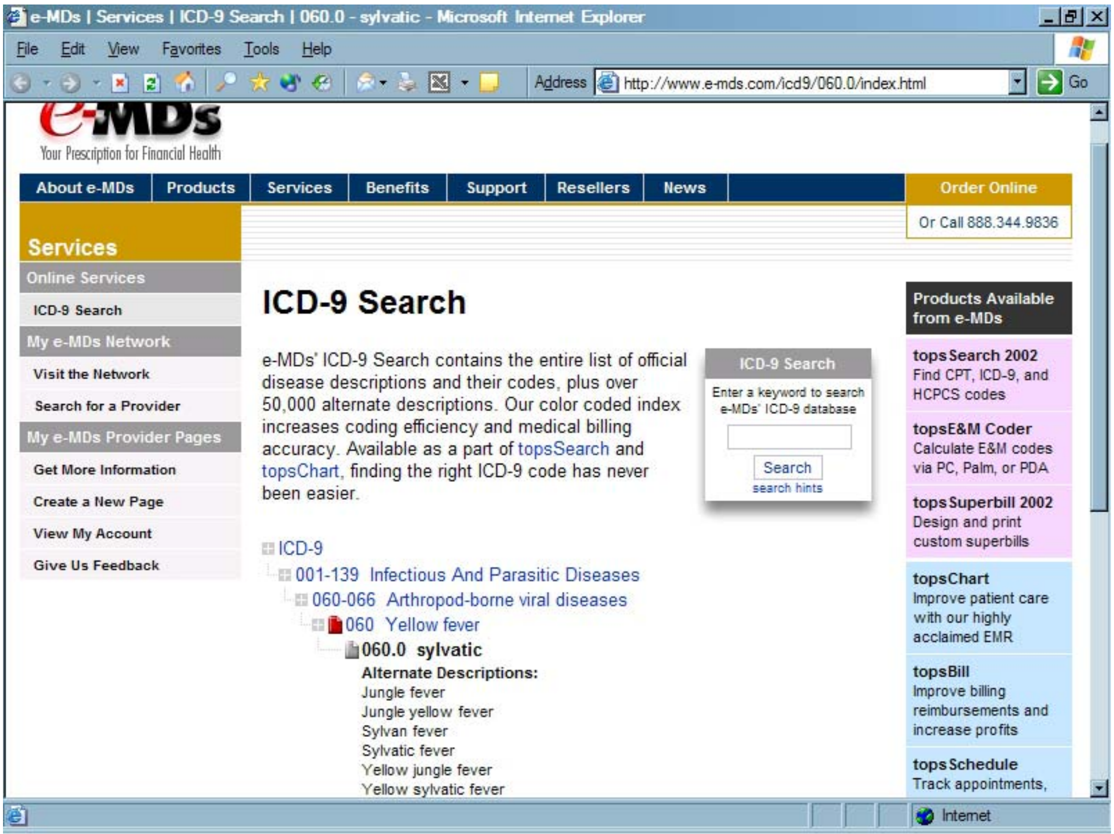


Figure 7.8. e-MDs ICD-9 Search (<http://www.e-mds.com/icd9/>).

7.3 HealthCyberMap's Editorial Policy for Resource Selection/ Quality Benchmarking

Health and medical Web resources are not all written by qualified, unbiased professionals, hence the need for rigorous quality benchmarking [64]. Quality also remains an important issue for other non-health-related and non-medical Web resources. In Tim Berners-Lee's vision of Semantic Web architecture, a "Web of Trust" forms the top layer (<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>), enabling content publishers and quality reviewers to take responsibility for (or be accountable for) what they publish, review or annotate on the Web and its quality.

The quality element in HealthCyberMap's extended DC set has been introduced to store a resource's level of evidence (whether it is an official guideline, systematic review, randomised controlled trial—RCT, other peer-reviewed study, official critically appraised topic—CAT, or expert opinion), and any other relevant information regarding its compliance to a recognised code of ethics (e.g., Health On the Net—<http://www.hon.ch/>) or quality seal, and whether it has been published by a trusted publisher or listed in trusted directory (e.g., OMNI—Organising Medical Networked Information—<http://omni.ac.uk>).

Back in the year 2000, the author reviewed the different quality benchmarking tools and checklists for health-related Web resources as part of his second MSc thesis [5] and in a peer-reviewed paper published in 2001 [64]. In selecting resources for the current HealthCyberMap pilot service, the author tried to observe the spirit of all reviewed tools and checklists, rather than sticking to any specific individual quality benchmarking checklist as no single checklist is complete. The fact that the author is medically qualified and worked as a clinician (dermatologist) for many years also helped a lot in discerning content quality and value of selected resources.

For each selected resource, the author tried to find an answer to the following questions:

- Who has written it and why? Can we confirm authors' identity and credibility/ contact them?
- When was the information first published/ last updated?
- Where is the information from? (Which country/ type of site: personal, institutional, academic, non-academic, commercial, etc.)
- Is the content suitable for the purpose it was written for? Are there any identifiable errors in it?

Sometimes it was necessary to spend extra time reading any About/ Aims/ History/ Disclaimer/ Copyright pages to find an answer to some of the above questions, and before deciding whether to add a resource to HealthCyberMap or not.

7.3.1 HealthCyberMap Subscribes to HONcode Principles

HealthCyberMap itself as a service has applied for and received the Health On the Net active seal (HONcode—Figures 7.9 and 7.10), meaning that its editorial policy abides by the eight principles of the HON Code of Conduct (authority, complementarity, confidentiality, attribution, justifiability, transparency of authorship, transparency of sponsorship, and honesty in advertising and editorial policy—[84]).

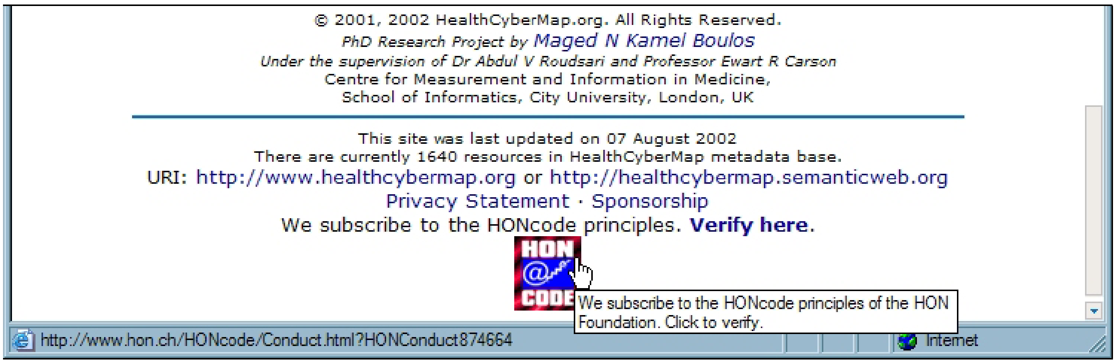


Figure 7.9. HealthCyberMap subscribes to the HONcode principles. This screenshot shows the active HONcode seal at the bottom of HealthCyberMap front page. Clicking this seal will retrieve and display HealthCyberMap's specific HONcode certificate from Health On the Net server (<http://www.hon.ch/HONcode/Conduct.html?HONConduct874664>—Figure 7.10 below).

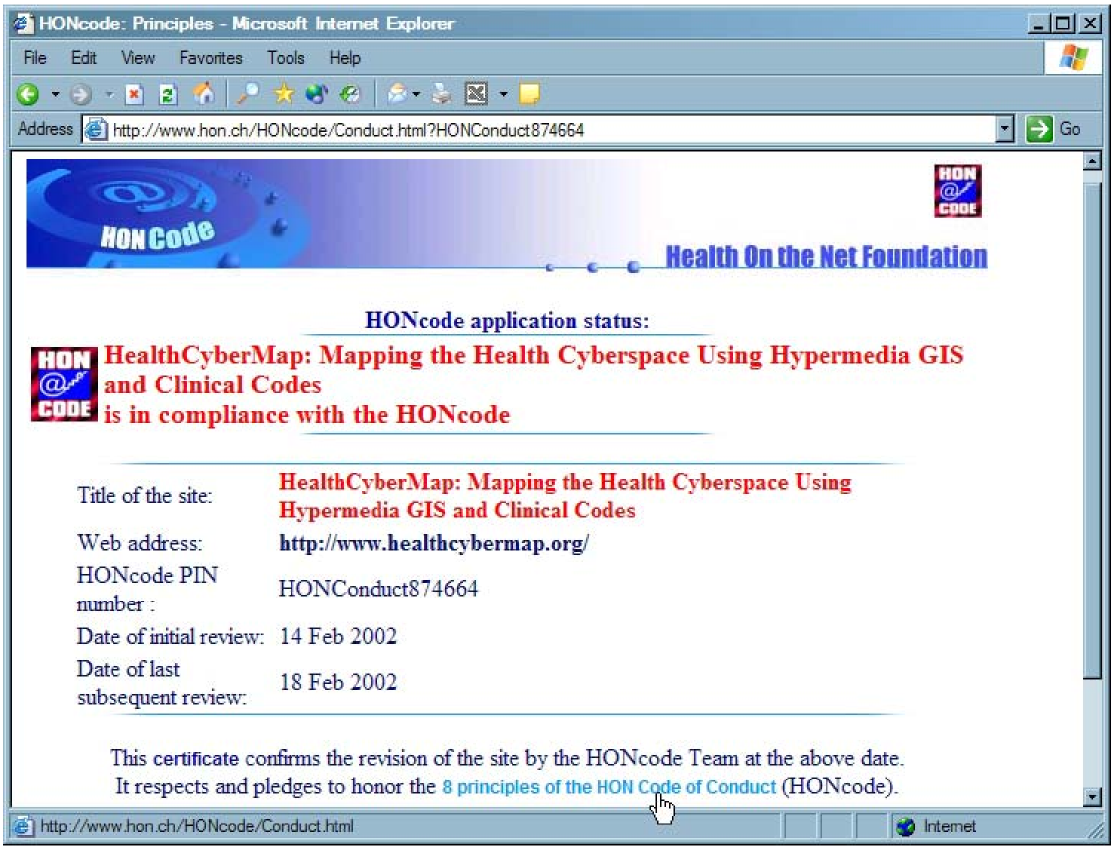


Figure 7.10. HealthCyberMap's specific HONcode certificate from Health On the Net server (<http://www.hon.ch/HONcode/Conduct.html?HONConduct874664>).

7.4 Conclusion

In this chapter, we have described the work undertaken to build HealthCyberMap's Semantic Arm. The author started by developing a model RDF metadata base based on the qualified DC metadata set with HealthCyberMap's own extensions for resource quality and geographical provenance. The DC subject fields in HealthCyberMap's metadata base are populated with UMLS terms directly imported from the UMLS Knowledge Source Server using the UMLS tab, a Protégé-2000 plug-in. This approach based on RDF and the qualified DC metadata set offers many opportunities in the future, including the possibility of other services making use of HealthCyberMap's RDF metadata base to develop their own Web portals. The different ways of exploiting Protégé-2000 RDF output were also discussed.

However, because ESRI ArcView GIS and its BodyViewer extension used to generate most of HealthCyberMap's navigational hypermaps cannot use an RDF data source and only work with ICD-9 codes not UMLS, the author developed another simpler metadata base in Microsoft® Access based on the non-qualified DC metadata set for use in the current HealthCyberMap pilot service. Candidate Internet resources were hand-selected by the author for inclusion in this metadata base. Their metadata attributes, including ICD-9-CM code(s) representing their subject(s) and any recognised quality/ code of ethics rating they bear, were manually compiled in HealthCyberMap's metadata base in Microsoft® Access. Big sites were not treated as a single resource, but rather subdivided into and indexed as distinct resources covering specific topics to increase the relevance of HealthCyberMap's answers to user information needs.

Health and medical Web resources are not all written by qualified, unbiased professionals, hence the need for rigorous quality benchmarking when selecting resources for HealthCyberMap. HealthCyberMap also subscribes to the HONcode principles and has received HONcode active seal and certificate. Manual indexing by trained humans (assisted by suitable tools like code locators) ensures the quality of selected resources and the precision of their topic indexing. Automatic free-text resource indexing by conventional Web spiders, although possibly providing much wider coverage in less time, cannot ensure the quality or topic indexing precision of spidered resources, and cannot index non-textual, multimedia Web resources.

Clinical codes can crisply and unambiguously describe the subject of medical Web resources, and automatically establish the semantic relationships (as defined by the coding scheme in use) between related resources for exploitation by suitable tools in different user interfaces. They also help automating the topical categorisation of these resources (e.g., automatically classify a resource on "diabetes mellitus" under "endocrine disorders").

8 HealthCyberMap Medical Semantic Subject Search Engine

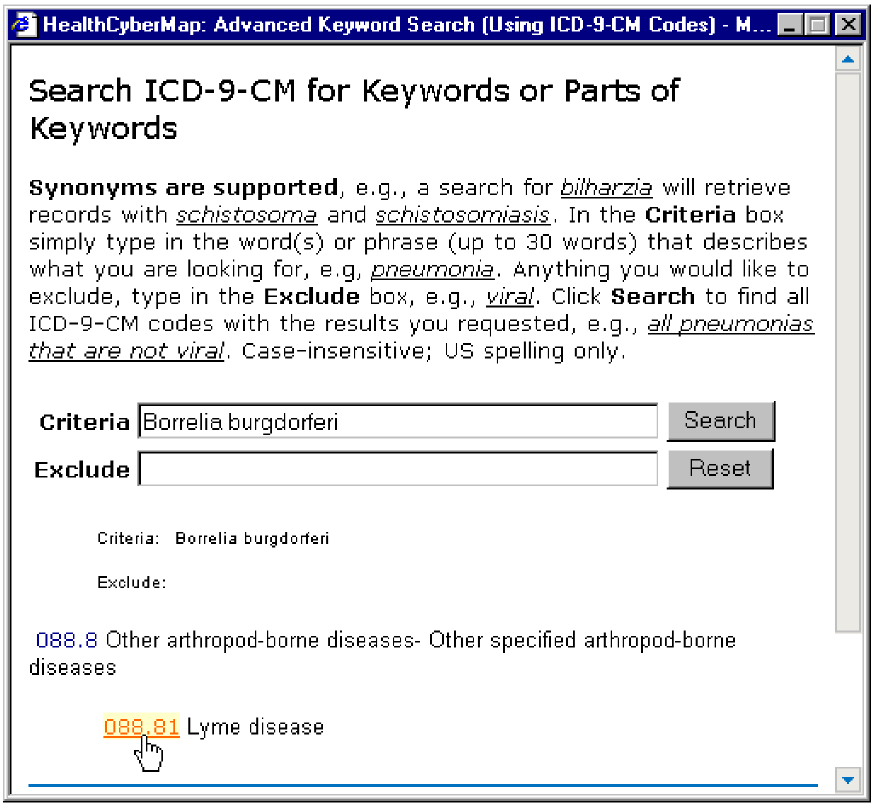
8.1 Background

HealthCyberMap uses clinical codes from a clinical coding scheme (currently ICD-9-CM) for topical resource indexing to populate the Dublin Core (DC) subject element. This conforms to DC recommended best practice of using a controlled vocabulary for this element to preserve semantics. HealthCyberMap currently allows three DC subject fields per resource record for richer, more complete descriptions. However, it is not practical or computationally efficient to encode all synonyms, semantic relationships and other possibilities of related topics in a resource or a metadata record of it (especially given the fact that resource indexing is still a largely manual process—see Chapter 7). The ideal system should be able, given the concept code(s) of a resource topic, to infer all allowed textual synonyms/ descriptions (even in multiple languages) for this topic, as well as the codes of any other relevant topics related to this resource via semantic relationships (Figure 6.2).

8.2 Current Pilot Implementation

The current HealthCyberMap Semantic Subject Search Engine (available online at <http://healthcybermap.semanticweb.org/icd.htm>) is an attempt to realise the design concepts described in Chapter 6, though with a much less sophisticated clinical coding scheme (ICD-9-CM). The tool uses a brokering ICD-9-CM ontology implemented using proprietary technology from Yaki Technologies (<http://www.eicd.com/>) in the form of a database where all ICD-9-CM codes and associated textual descriptions, synonyms and relations are represented. The ontology acts as "the intelligent middleman" between user queries and HealthCyberMap ICD-9-CM-coded resource metadata pool.

The current HealthCyberMap Semantic Subject Search Engine supports synonyms, disease variants and subtypes. For example, a user query for "bilharzia" will retrieve records with "schistosoma" and "schistosomiasis" and a query for "crohn's disease" is able to retrieve "regional enteritis". A Criteria box allows users to type what they are looking for (keywords or parts of keywords), e.g., "pneumonia". Another Exclude box allows them to type anything they would like to exclude, e.g., "viral" (to retrieve "all pneumonias that are not viral" in this example). Some semantic relationships are also supported.



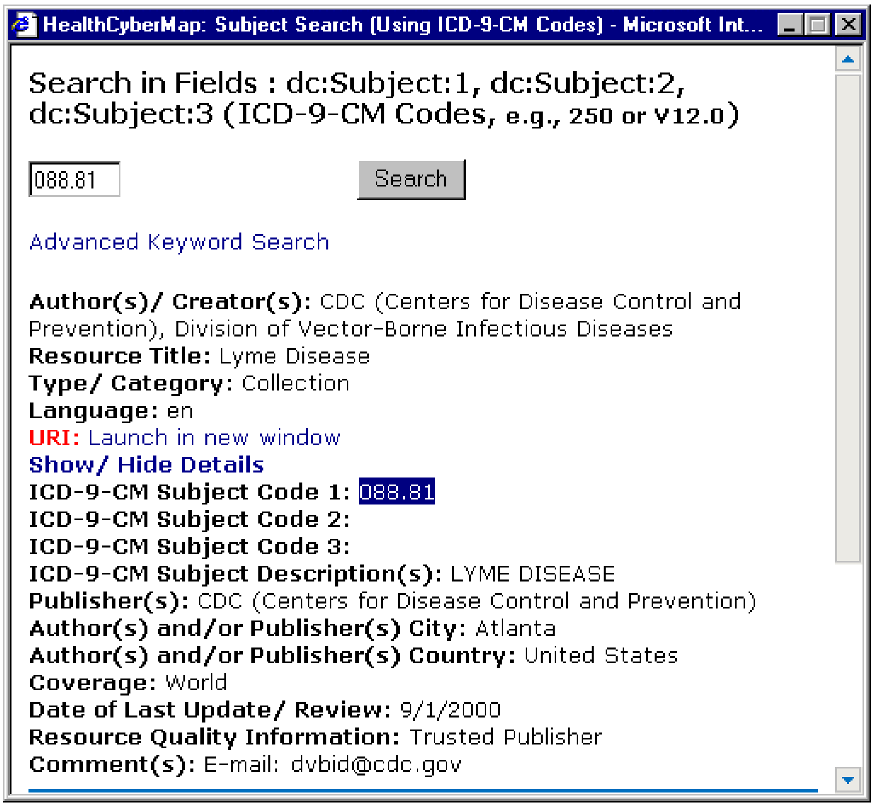


Figure 8.1. A search for "Borrelia burgdorferi" retrieves resources on "Lyme disease" (ICD-9-CM code: 088.81). HealthCyberMap ICD-9-CM brokering ontology "knows" (or has the relation properly defined or represented in it which says) that "Lyme disease" *is-caused-by->* "Borrelia burgdorferi" (an organism). Note that in the second screenshot above, the words "Borrelia burgdorferi" are not present in the retrieved resource metadata record.

For example, a search for “cataract” will retrieve “phacolytic glaucoma” (ICD-9-CM code: 365.51) among search results. The word “cataract” does not occur as such in the metadata term “phacolytic glaucoma”, but a relationship can be inferred from the meanings of the latter (“phacolytic glaucoma” is the sudden onset of open-angle glaucoma caused by a leaking cataract—see also Figure 8.1 for another example). A conventional free-text search engine cannot deliver all these advanced search features. Ideally, search results should be classified according to how semantically close a match is to user query into direct matches (including synonyms), e.g., resources dealing primarily with the organism, “Borrelia burgdorferi”, in the example shown in Figure 8.1, and related matches (based on semantic relationships), e.g., resources covering topics related to “Borrelia burgdorferi” like “Lyme disease”, to use the same example. However, we could not implement this classification of returned matches in this pilot search engine due to limitations in both ICD-9-CM (its coarse semantic granularity) and the way relationships are represented in the current tool.

8.3 Conclusion

In this chapter, we have described the early pilot version of HealthCyberMap Semantic Subject Search Engine that attempts to overcome the limitations of conventional free text search engines.

Explicit concepts in resource metadata map onto a brokering domain ontology (a clinical terminology or classification) allowing a Semantic Web search engine to infer implicit meanings (synonyms and semantic relationships) not directly mentioned in either the resource or its metadata. Similarly, user queries would map to the same ontology allowing the search engine to infer the implicit semantics of user queries and use them to optimise retrieval. The early pilot HealthCyberMap Semantic Subject Search Engine currently available online supports synonyms, disease variants and subtypes (<http://healthcybermap.semanticweb.org/icd.htm>).

9 HealthCyberMap's Visualisation/ Navigation Arm — Spatialised Browsing of Medical/ Health Internet Resources

9.1 Introduction

HealthCyberMap (<http://healthcybermap.semanticweb.org>) is a Web-based service that aims at mapping selected parts of medical/ health information resources in cyberspace in novel semantic ways to improve their retrieval and navigation. This is achieved through "intelligent" categorisation and interactive hypermedia visualisation of the medical/ health information cyberspace using metadata (information about information resources), clinical codes (to describe resource topics—see Chapters 6 and 7) and GIS (Geographic Information Systems) technologies. HealthCyberMap pilot service currently provides six different interfaces to its metadata base, which has over 1600 resource records in it. Some of these interfaces are visual (maps, e.g., Figure 9.1), while others are textual (e.g., HealthCyberMap Semantic Subject Search Engine—Chapter 8, and directory of topical categories).

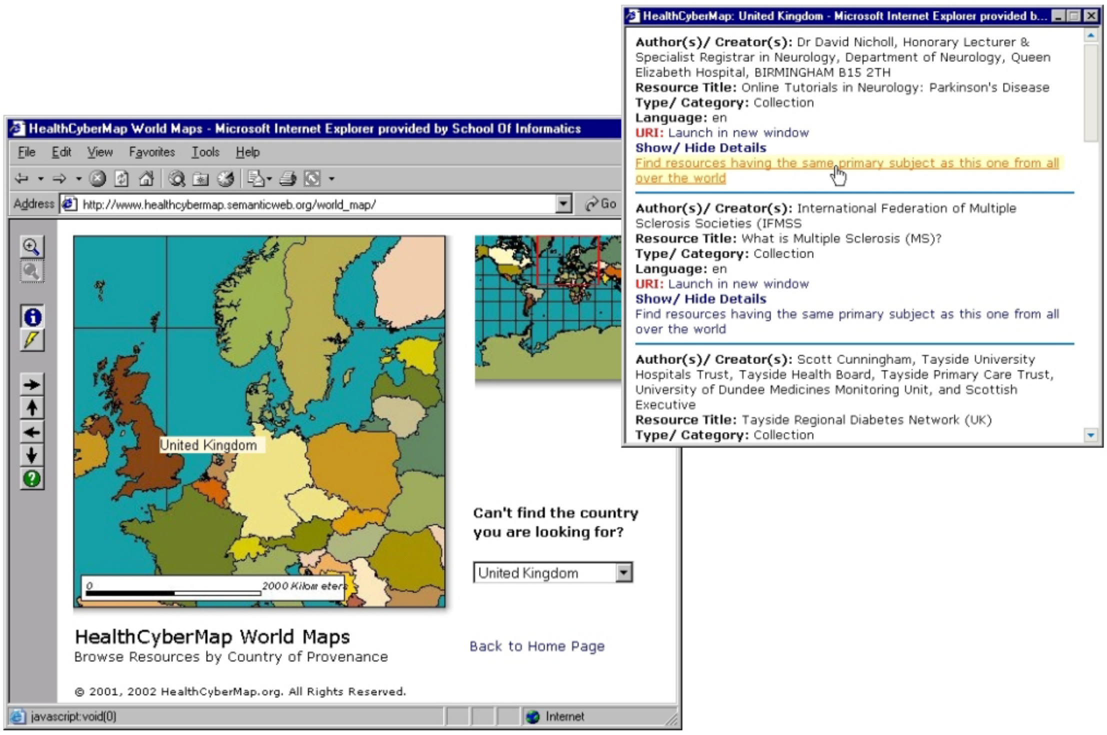


Figure 9.1. Screenshot of HealthCyberMap World Map Web interface (<http://healthcybermap.semanticweb.org/world_map/>). Note the country name ToolTip ("United Kingdom") and the different map interface buttons on the left. Also note the clickable overview map with a red positional square on the right. The world map is rendered as a chorochromatic map. Medical/ health information resources are mapped to countries (of authors/ publishers) rather than cities and listed in a separate pop-up text window (query result page) to avoid map clutter. The latter would have been unavoidable had the author opted to represent each resource using a distinct point symbol on the map. Note the *"Find resources having the same primary subject as this one from all over the world"* link at the end of each resource bibliographic card in the resource list pop-up window to the right.

HealthCyberMap features a novel and unconventional use of GIS to map conceptual spaces occupied by collections of medical/ health information resources based on the metadata attributes of these resources (stored in HealthCyberMap's resource metadata base) and using suitable metaphors like human body organs/ systems. The resultant maps are conceptual information space maps used as a visual navigational aid for browsing mapped resources.

9.2 HealthCyberMap Semantic Distance Metric

In cyberspace, conventional map metrics like distance, map projection, scale and grid assume new meanings, and new metrics also arise. In HealthCyberMap, the author used a "distance" metric based on the "semantic locations" of resource topics within a clinical coding scheme projected on a human body organs/ systems map. The clinical coding scheme acts as a semantic conceptual space with resources occupying different locations in this space based on their meaning (semantics or subject topics). The "semantic distance" between two resources will then depend on how close (or related) the two resources are from a semantic perspective (based on their subjects and their semantic locations as determined by the clinical coding ontology they are mapped to). For example, a resource on "myocardial infarction" should be much closer to a resource on "angina pectoris" than to another resource on "psoriasis".

The author used ICD-9-CM (International Classification of Diseases, ninth revision, US Clinical Modification—[83]) as the clinical coding scheme/ ontology in HealthCyberMap pilot service. Table 9.1 shows the top-level grouping or classification of resources in HealthCyberMap based on ICD-9-CM code ranges.

Ideally, resources on multi-organ/ -system diseases should be spatialised to all relevant human body locations, not just a single location, i.e., they should be listed under all pertinent categories. This will ensure that users will always find the information they are looking for in the places where they expect it to be present.

Code assigned to resource (semantic location)	**Corresponding body organ/ system location (projected spatial location on human body maps)**
(001-139)	Infectious and Parasitic Diseases*
(140-239)	Neoplasms*
(240-259)	Endocrine Diseases
(260-279)	Nutritional and Metabolic Diseases and Immunity Disorders
(280-289)	Diseases of the Blood and Blood-forming Organs
(290-319)	Mental Disorders
(320-389)	Diseases of the Nervous System and Sense Organs
(390-459)	Diseases of the Circulatory System
(460-519)	Diseases of the Respiratory System
(520-579)	Diseases of the Digestive System
(580-599)	Diseases of the Urinary System
(600-629)	Diseases of the Reproductive System
(630-676)	Complications of Pregnancy, Child Birth, and the Puerperium
(680-709)	Diseases of the Skin and Subcutaneous Tissue
(710-739)	Diseases of the Musculoskeletal System and Connective Tissue
(740-759)	Congenital Abnormalities*
(760-779)	Certain Conditions Originating in the Perinatal Period
(780-799)	Symptoms, Signs and Ill-defined Conditions*
(800-999)	Injury and Poisoning
(V01-V82.9)	V Codes (Factors Influencing Health Status and Contact with Health Services, e.g., Vaccinations)
(E800-E999)	E Codes (External Causes of Injury and Poisoning)
	*Sub-ranges should be spatialised to appropriate locations on human body maps

Table 9.1. Top-level grouping or classification of resources in HealthCyberMap based on ICD-9-CM code ranges.

9.3 Spatialisation in HealthCyberMap: The Different Resource Navigation Maps

The current HealthCyberMap pilot service uses “conventional” geographical hypermaps to map health resources on the Web to the country of their corresponding providers (Figure 9.1). Another set of hierarchical human body topical maps allows navigating resources by body location/ system according to ICD-9-CM, which acts as HealthCyberMap’s medical ontology (Figure 9.2). A third type of hypermaps categorises resources by type, e.g., electronic journal articles, digital atlases, etc. All maps are used to locate (by querying an underlying database), launch health resources on the Web, and display their bibliographic metadata records.

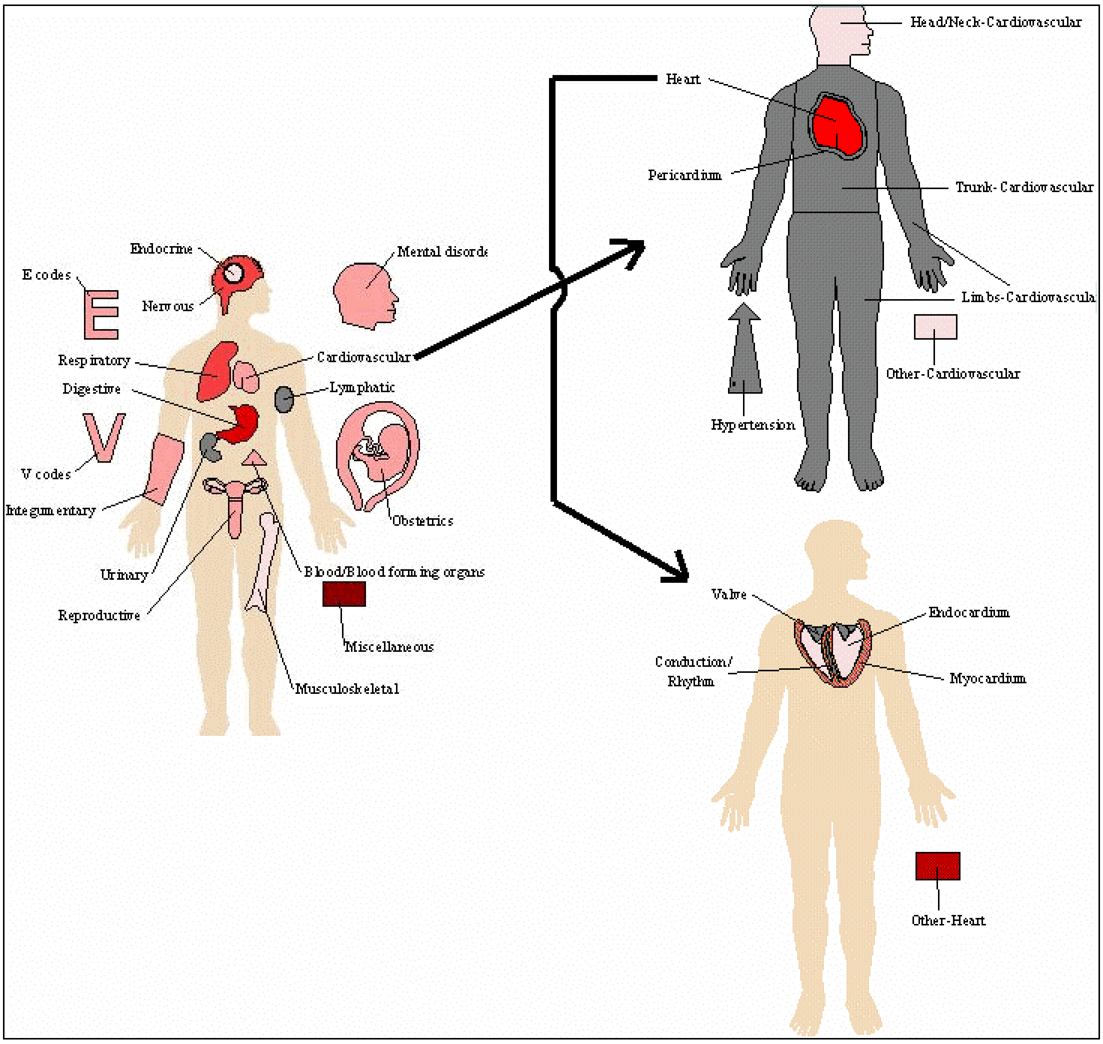


Figure 9.2. Three BodyViewer human body hypermaps from HealthCyberMap (<http://healthcybermap.semanticweb.org/bodyviewer/>). Clicking the Heart (Cardiovascular) on the main Human Body Map brings another detailed map of the Cardiovascular System. Clicking the Heart again on the latter map, calls a third map with a detailed view of the Heart.

When a spatialised Web space is rendered in the form of an interactive (clickable) map, the spatial representation (map) becomes the territory and also the means in which to navigate this territory. Links on such map are not just a representation of corresponding Web resources, but can also directly launch (take users to) these resources. This is not true for conventional real world maps, which were never meant to be the territory; they just allow users to visualise land and routes, then users will need to drive along real-world streets (not the maps obviously) to physically reach their destination [42].

9.4 HealthCyberMap Implementation in ArcView GIS

HealthCyberMap has been developed as an ArcView GIS project and features GIS-driven spatialisation based on an underlying resource metadata base where ICD-9-CM codes describing the topics of mapped resources are stored alongside other useful information about theses resources, including their geographic provenance and Web addresses (Chapter 7). The author used ESRI ArcView GIS Version 3.1 for Windows (<http://www.esri.com>). WebView 1.1, the Internet extension to ArcView GIS, was then used to translate HealthCyberMap Views (maps) from ArcView to the Web in the form of client-side imagemaps in JPEG format. The author also used another ArcView extension in HealthCyberMap's project, namely BodyViewer v2.1 for ICD-9 codes from GeoHealth, Inc. to generate HealthCyberMap's human body maps (see below).

9.4.1 WebView Features

WebView was developed by Thomas Zerweck at ZEBRIS in Munich, Germany (<http://www.zebris.com/>—[85]). It was programmed in Avenue (an ArcView scripting language), HTML (for the Web templates) and JavaScript (to add additional interactive functionality to its client-side Web maps). It is much cheaper compared to a dedicated ESRI Internet Map Server solution, though not as powerful as the latter (see Chapter 5).

A WebView wizard leads users through the necessary steps in ArcView GIS and creates the project's Web pages (based on the active view in ArcView at the time the wizard is launched—Figures 9.3, 9.4 and 9.5). These pages can then be edited manually if necessary in any HTML editor. The created map pages and interface can provide the following functionality [85]:

- Detail and overview maps (Figure 9.1 above). The detail map displays all visible themes of the active view in the chosen scales; it only displays part (one tile) of the whole view area at a time. The overview map displays the overview themes of the whole area at once in miniature form. A red positional rectangle moves over the overview map to show the location of the area currently displayed in the detail map. Users can also navigate to a different area in the detail map by clicking in the overview map.
- Panning in the detail map is also possible using four arrow buttons for the four directions (up, down, right, left).

- Legend for map contents.
- Scale bar.
- WebView offers three-way hotspots with two-way clicks depending on which toolbar button (Identify button with a blue 'i' icon  or HotLink button with a yellow spark icon ) is selected when the user clicks a map object:
  - attribute information can be displayed on mouse over (map feature ToolTip, e.g., country name or body organ/ system name in HealthCyberMap's maps);
  - other attribute information can be displayed on mouse click while the Identify button is selected, e.g., to display more country information in a message box based on one or more attribute fields (see Figure 9.21 below); and
  - mouse clicks while the HotLink button is selected can be associated with an image, video, sound file, Web page or email address. (In HealthCyberMap, the author associated them with database query pages to be executed on HealthCyberMap Web server.)
- Zoom in and zoom out in the detail map (up to three zoom levels in WebView 1.1).
- Up to five themes can be selected as interactive layers for each of the three zoom levels; attributes of these themes can be associated with the different mouse events outlined above (Figure 9.4). Different themes (layers) can be associated with the different zoom levels. This allows for different map contents and detail at different zoom levels (scales). This zooming strategy is called static stepped zooming.

Version 2.1 of WebView provides additional features (not available in WebView v1.1). These include five zoom levels, transparent layers (users can switch layers visible/ invisible) and object layers (Figure 9.6). The latter provides a client-side map feature lookup functionality "similar" to, though not as powerful or sophisticated as that provided by dedicated map servers. Developers select up to two themes as object layers. An object layer is a theme, in which the user may search for certain attribute value. The attribute values are listed in a combo box, e.g., a list of country names as in the screenshot below (Figure 9.6). Selecting a country from the list, will cause the detail map (on the left) to zoom to that country and mark its exact position on the map (Figure 9.6) [85].

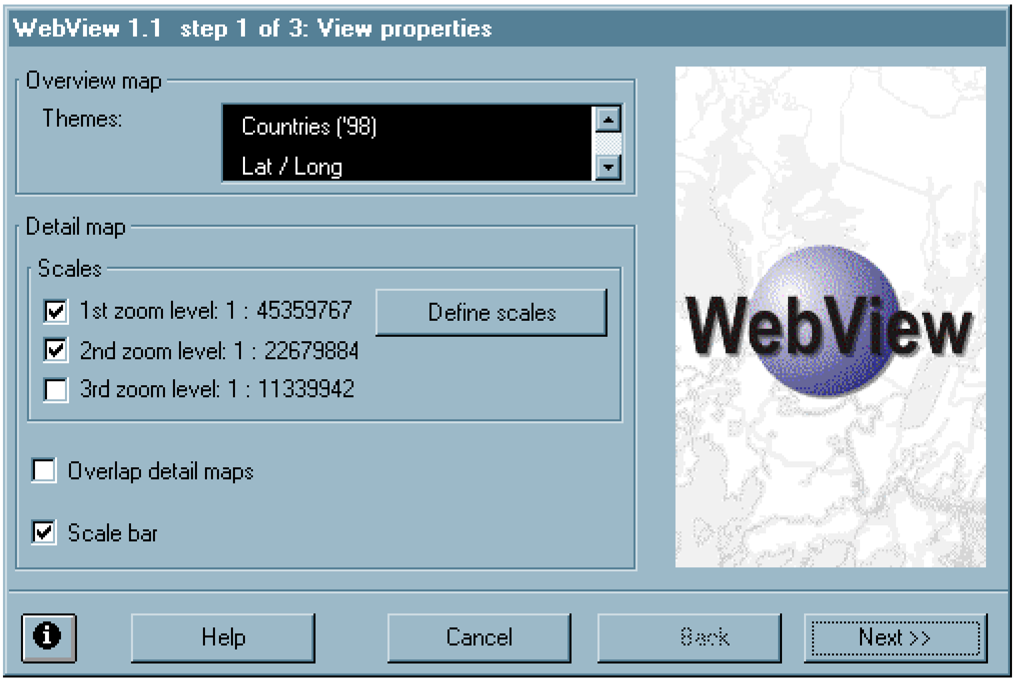


Figure 9.3. Step 1 of WebView 1.1 wizard in ArcView. Users can select themes for the overview map and determine how many zoom levels they want for the detail map (maximum 3) and the map scale for each level.

Figure 9.4. Step 2 of WebView 1.1 wizard in ArcView. Up to five themes can be selected as interactive layers for each zoom level; attributes of these themes can be associated with the different mouse events (on mouse move, on click/ Identify, and on click/ HotLink). Different themes can be associated with the different zoom levels. The (red) arrow points to the HotLink attribute, a field the author added to the table of "Countries ('98)" theme in ArcView (see later). It stores Web addresses of database query pages on HealthCyberMap server.

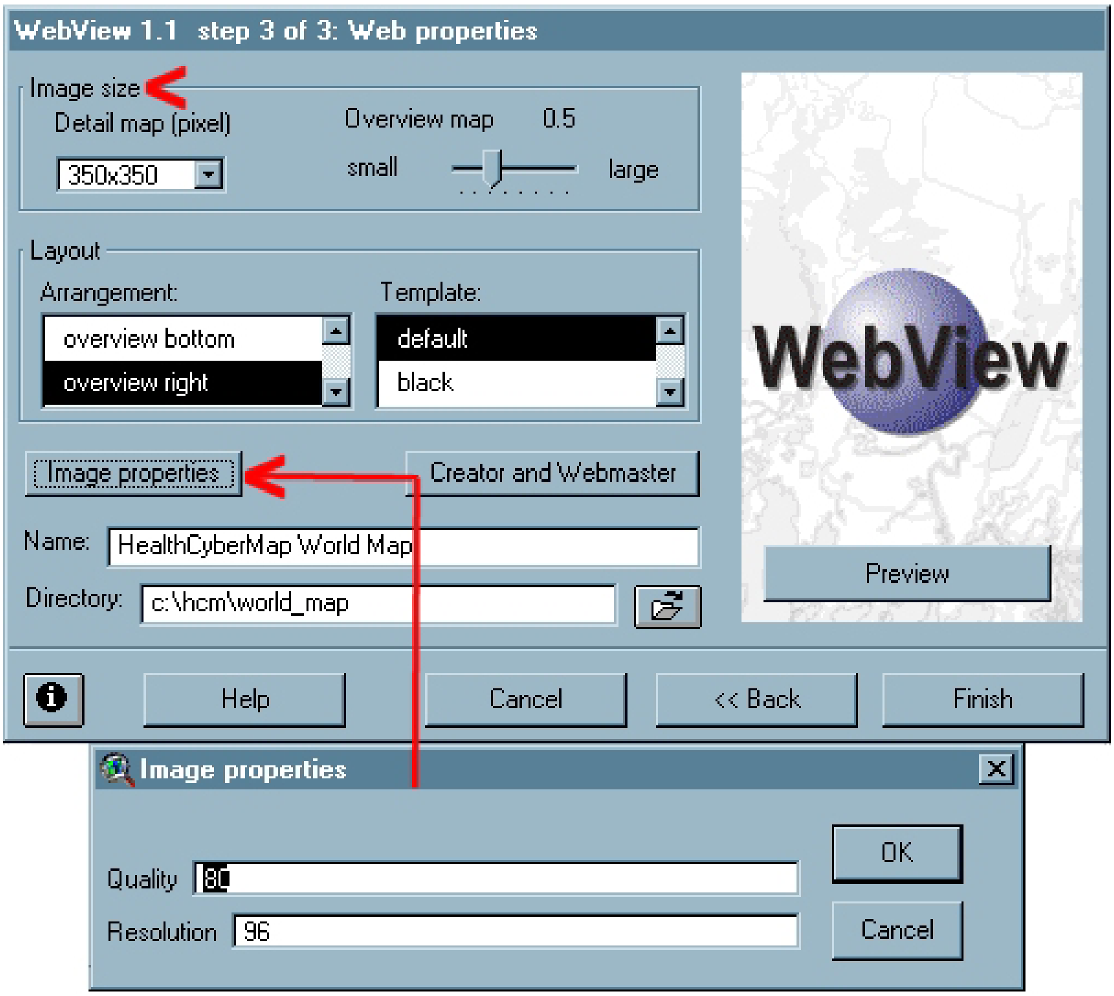


Figure 9.5. Step 3 of WebView 1.1 wizard, showing the Web Properties and Image Properties dialogue boxes. The user can control the size of the saved image files by adjusting the size of the output images in pixels (individual tiles or detail maps and overview map) and their JPEG quality (better quality is achieved on the expense of lower compression; a quality setting of 80 is WebView default).

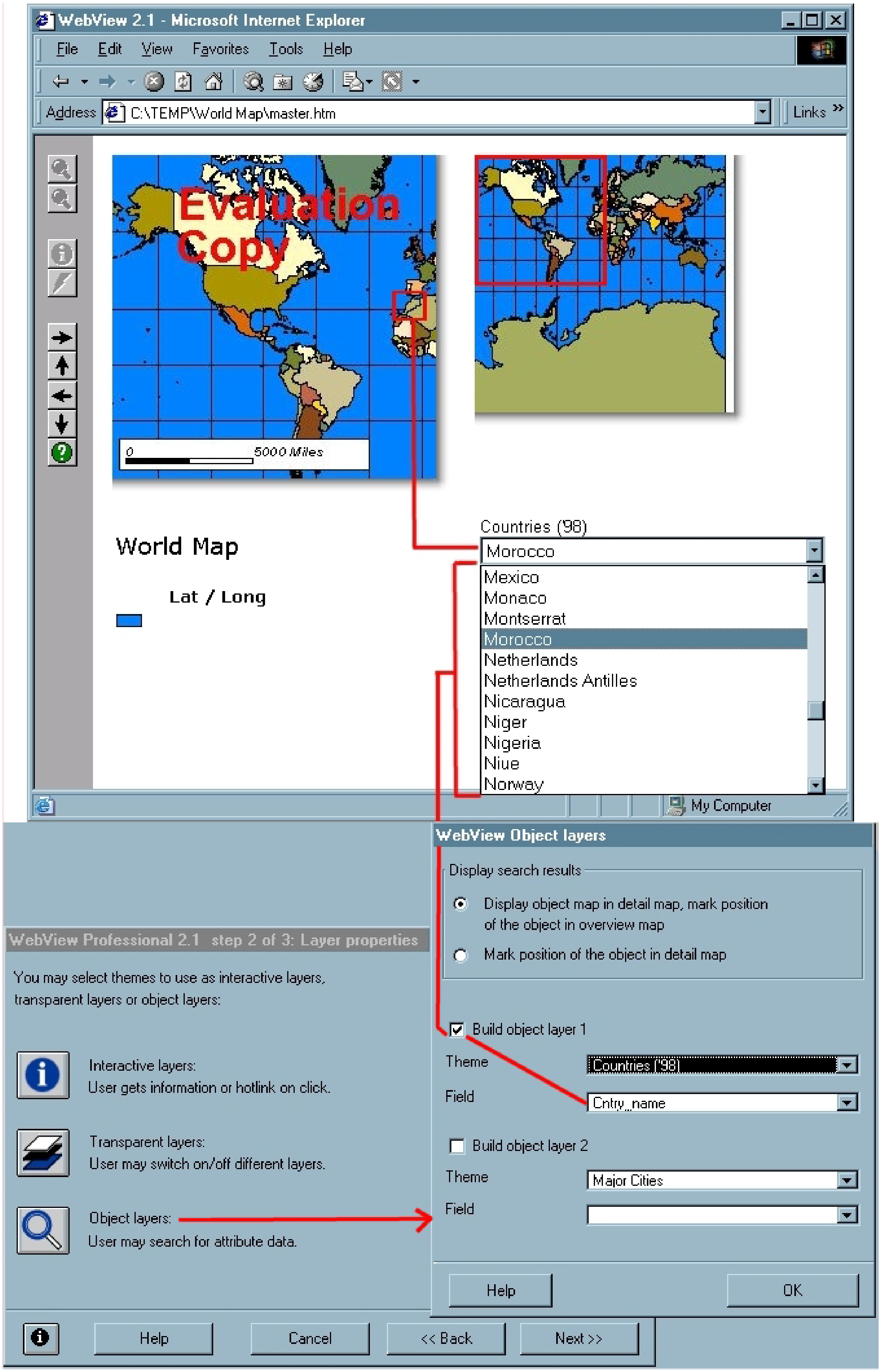


Figure 9.6. Screenshot of WebView v2.1.

9.4.2 WebView Limitations

Although it saves users the trouble of setting-up and running more complex Internet Map Server software while offering similar user interface features, the basic WebView set-up does not support any real GIS database drill-down functionality (the generated maps cannot communicate with the corresponding underlying databases). Moreover, projects published by WebView on the Web are uncoupled or disconnected from the original corresponding projects in ArcView.

9.4.3 "Patching" the Basic WebView Set-up

In HealthCyberMap, the author developed his own (partial) workarounds for these limitations of WebView (Figure 9.7). This solution makes use of WebView HotLink functionality to implement a dynamic database drill-down that will always reflect the latest updates to this database. By clicking different hotspots on the client-side imagemaps in HealthCyberMap, users are actually triggering server-side pre-formulated SQL (Structured Query Language) queries against an underlying database of resource metadata. The database is registered on HealthCyberMap server (a Windows 2000/NT 5 IIS Server—Microsoft Internet Information Server/ Services) as an ODBC (Open DataBase Connectivity) Data Source (Figure 9.8) and is the same database we are connecting to in ArcView.

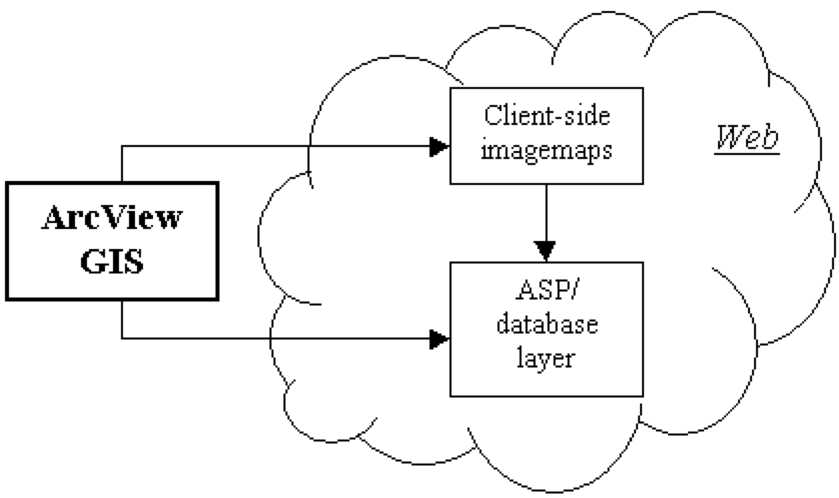


Figure 9.7. HealthCyberMap (partial) workarounds for WebView limitations. WebView converts HealthCyberMap's Views in ArcView to client-side imagemaps for the Web. Clicking the different hotspots on these client-side imagemaps will trigger server-side pre-formulated SQL queries against an underlying database of resource metadata on HealthCyberMap server. The database is the same database we are connecting to in ArcView. The author coded the SQL queries in ASP pages for execution on the server.

The author coded the SQL queries in ASP (Active Server Pages) pages for execution on the server. The ASP pages returned to users in their browsers only contain query results in the form of formatted HTML. The actual SQL and ASP code as found in the ASP pages stored on the server is never sent to the end user.

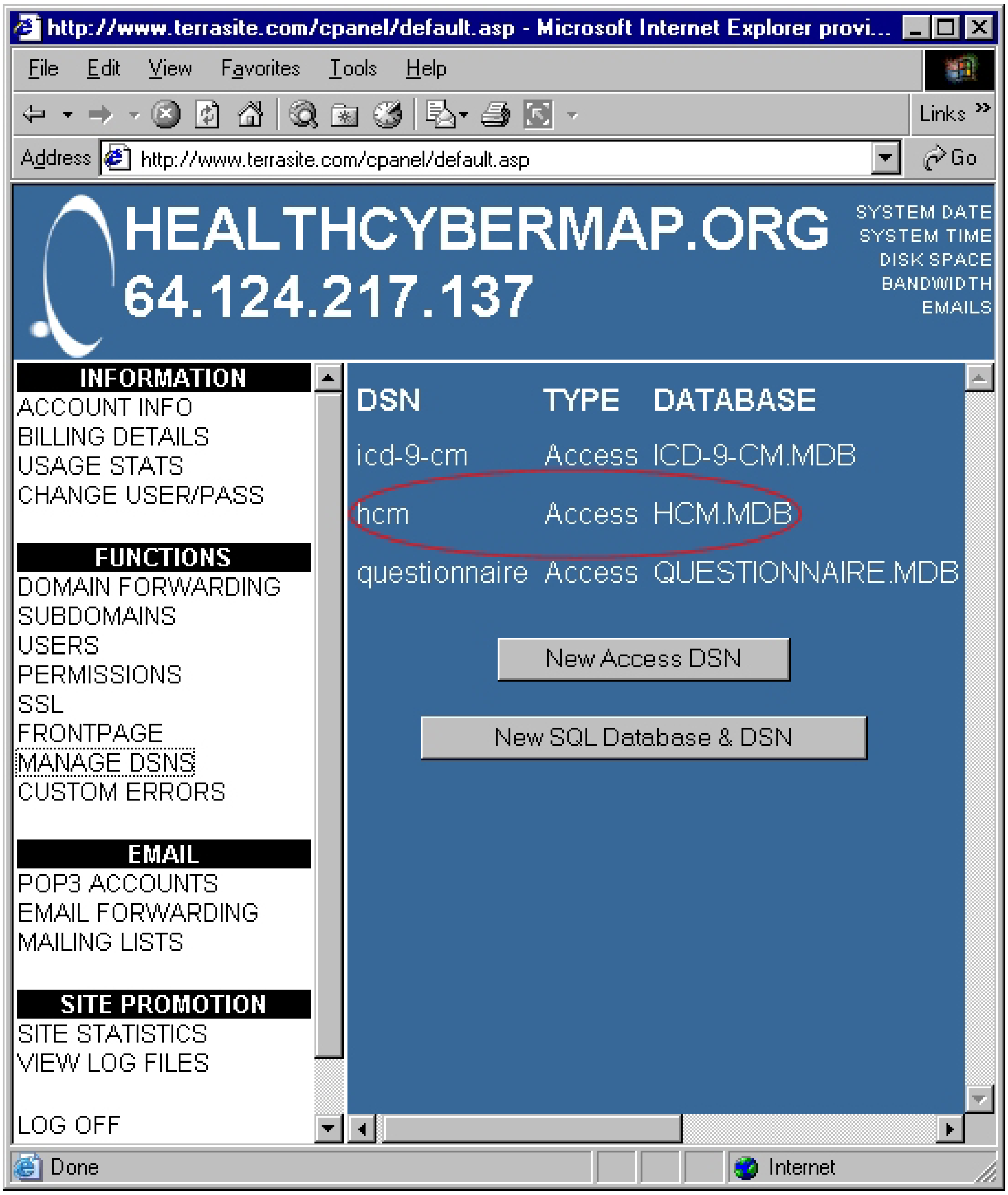


Figure 9.8. Screenshot of HealthCyberMap Web Server Control Panel (only accessible by server administrator). The same Microsoft® Access database we are connecting to in ArcView (HCM.MDB) is shown registered on HealthCyberMap server (a Windows 2000/NT 5 IIS Server) as an ODBC Data Source.

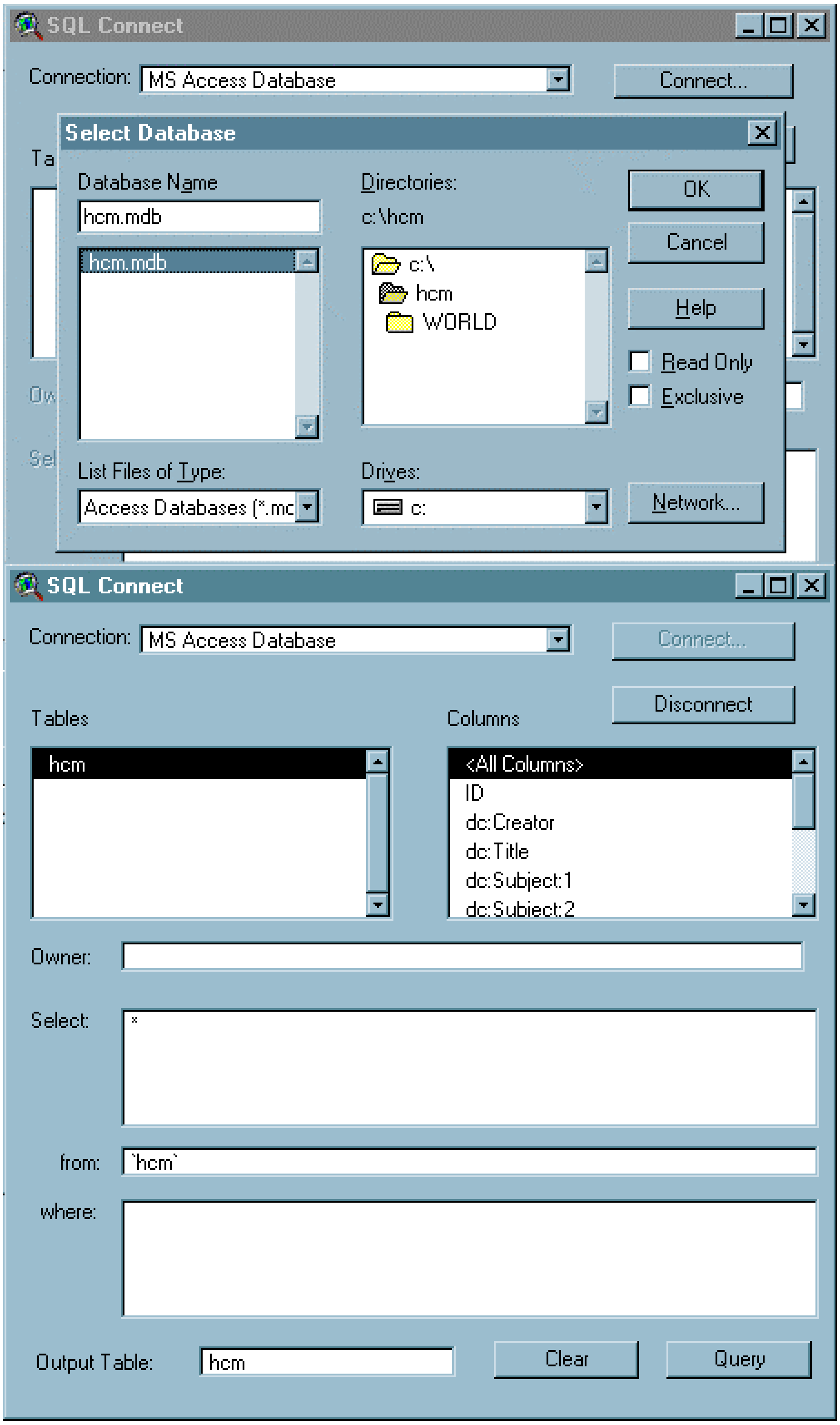


Figure 9.9. Screenshot of ArcView "SQL Connect" dialogue box. The author established a connection to HealthCyberMap metadata base (Microsoft® Access) and created an ArcView table ("hcm") containing all fields (and records) from the input Microsoft® Access table.

9.4.4 HealthCyberMap Metadata Base in ArcView

The resource metadata base was implemented in Microsoft® Access based on Dublin Core (DC) metadata set ([22]—Chapter 7). The author used ArcView "SQL Connect" feature to connect to HealthCyberMap metadata base and import all fields and records from it into an ArcView table that will refresh each time the project is opened in ArcView (Figures 9.9 and 9.10). This is the same database currently registered on HealthCyberMap server as an ODBC Data Source (Figure 9.8 above), which users query by clicking the hypermaps.

hcm

ID	Dc:Creat	Dc:Title	Dc:Subject:1	Dc:Subject:2	Dc:Subject:3	Dc:Description	Dc:Publisher	Dc:Date	Dc:Type	Dc:Identifier
33	WHO	Silicosis	500	502		COAL WORKERS' PNEUMOC	WHO	2000-05-01	Fact Sheet	http://www.who.int/
34	WHO	Solar Radiation and Human He	692.7	172	370.24	SOLAR RADIATION DERMAT	WHO	1999-08-01	Fact Sheet	http://www.who.int/
35	WHO	Solar Eclipses: Danger to Eyes	363.31			SOLAR RETINOPATHY	WHO	1999-08-01	Fact Sheet	http://www.who.int/
36	WHO	Tobacco Dependence	305.1			TOBACCO USE DISORDER	WHO	1999-04-01	Fact Sheet	http://www.who.int/
37	WHO	Tobacco - Health Facts	305.1	989.84		TOBACCO USE DISORDER, T	WHO	1999-04-01	Fact Sheet	http://www.who.int/
38	WHO	Tuberculosis	010	011		PRIMARY TUBERCULOUS IN	WHO	2000-04-01	Fact Sheet	http://www.who.int/
39	WHO	Typhoid Fever	002			TYPHOID AND PARATYPHOI	WHO	1997-03-01	Fact Sheet	http://www.who.int/
40	WHO	Typhus	080			LOUSE-BORNE TYPHUS	WHO	1997-05-01	Fact Sheet	http://www.who.int/
41	WHO	Yellow Fever	060	V04.4		YELLOW FEVER, VACCINE FI	WHO	1999-08-01	Fact Sheet	http://www.who.int/
42	CDC (Ce	Herpes Zoster (Shingles); Varic	053	052		HERPES ZOSTER, CHICKENI	CDC (Centers f	1999-11-03	Fact Sheet	http://www.cdc.gov
43	CDC (Ce	West Nile Virus Home Page	066.3			OTHER MOSQUITO-BORNE I	CDC (Centers f	2001-04-04	Collection	http://www.cdc.gov
44	CDC (Ce	Lyme Disease	088.81			LYME DISEASE	CDC (Centers f	2000-09-01	Collection	http://www.cdc.gov
45	CDC (Ce	Dengue Fever & Dengue Hem	061			DENGUE	CDC (Centers f	1997-06-01	Collection	http://www.cdc.gov
46	CDC (Ce	CDC Plague Home Page	020			PLAGUE	CDC (Centers f	2000-10-04	Collection	http://www.cdc.gov
47	CDC (Ce	Yellow Fever Disease and Vac	060	V04.4		YELLOW FEVER, VACCINE FI	CDC (Centers f	2000-10-28	Fact Sheet	http://www.cdc.gov
48	CDC (Ce	Japanese Encephalitis Home P	062.0			JAPANESE ENCEPHALITIS	CDC (Centers f	1999-04-01	Collection	http://www.cdc.gov
49	CDC (Ce	Amebiasis	006			AMEBIASIS (AMOEBIASIS)	CDC (Centers f	2001-02-13	Collection	http://www.cdc.gov
50	CDC (Ce	Alveolar Hydatid Disease	122			ECHINOCOCCOSIS	CDC (Centers f	2001-02-13	Collection	http://www.cdc.gov
51	CDC (Ce	Chagas Disease	086.0	086.1	086.2	CHAGAS DISEASE OF HEAR	CDC (Centers f	2001-02-13	Collection	http://www.cdc.gov
52	CDC (Ce	Hookworm Infection	126	126.9		ANCYLOSTOMIASIS (ANCYL(	CDC (Centers f	2001-02-13	Collection	http://www.cdc.gov

Figure 9.10. HealthCyberMap metadata base as a table in ArcView.

9.4.5 Using ArcView's BodyViewer Extension to Generate Multi-level Human Body Maps of ICD-9-coded Resources

BodyViewer is an ArcView GIS extension from GeoHealth, Inc. (<http://www.geohealth.com/>) that combines the power of GIS with computerised body organ system diagrams. It lets users see where their ICD-coded healthcare data (medical/ health Internet resources in our case) map onto the human body based on the body region(s) they cover. Multiple levels of analysis with multi-level human body maps are provided. Each organ system is broken down into its major components for a finer level of detail, e.g., the Digestive System is broken into Mouth, Oesophagus, Stomach, Liver, Gallbladder, Pancreas, Small Intestine, Large Intestine, Rectum, Other-Digestive, and Metabolic Disorder. Public Issues are a special category. Instead of body diagrams, communicable diseases are categorised according to their mode of transmission (Food/ Water, Sexual, Animal to Person, Air, Person to Person, and Other) using meaningful symbols [86].

The author used this extension to generate the human body topical maps in HealthCyberMap. These maps allow the navigation of resources by body location/ system according to ICD-9-CM.

In BodyViewer human body maps, map symbols are miniature simplified drawings or icons of the different body organs and systems that observe all applicable cartographic rules for good map symbol design [38]. They act as easy-to-understand *visual labels* (familiar metaphors) to the different resource categories that have been classified and mapped according to their DC subject fields (ICD-9 codes). These icons (on the corresponding Web hypermaps) are linked to respective ASP query pages that are executed on HealthCyberMap Web server to retrieve the appropriate resources based on the ICD-9 codes represented by the clicked icon. For example, if the cardiovascular icon is clicked, a query will be launched to retrieve resources with cardiovascular ICD-9 codes. HealthCyberMap's bibliographic/ cybergraphic use of this extension to map ICD-coded medical/ health Internet resources is the first of its kind and was never suggested in BodyViewer documentation by GeoHealth, Inc. (the manufacturer of BodyViewer).

The author used BodyViewer "Setup Wizard" to map the Internet resources listed in HealthCyberMap onto the human body (Figures 9.11 and 9.12). These resources have been indexed in the metadata base that we imported into ArcView in the previous step (see Figure 9.10 above). The mapping is based on the three ICD-9 fields in the imported table. BodyViewer can aggregate more than one ICD-9 code field at a time, and so was able to use all three DC subject fields in HealthCyberMap's metadata table combined to compute resource counts by clinical subject category/ body region.

9.4.5.1 A Choropleth Rendition for Spotting Topical Coverage Gaps

BodyViewer classifies resource counts per body region into ranges and associates each range with a colour shade or tint, i.e., a choropleth rendition (organs with darker red tints have more resources associated with them than organs with lighter red shades; a grey colour denotes no resources—see Figure 9.16 below). This allows us to visually spot *infogaps* and *infoclusters*, a useful form of "cyberspatial analysis". Infogaps represent body areas (topics) where resources are deficient and should be addressed by information providers (topical coverage gaps). They can be also due to insufficient indexing by HealthCyberMap.

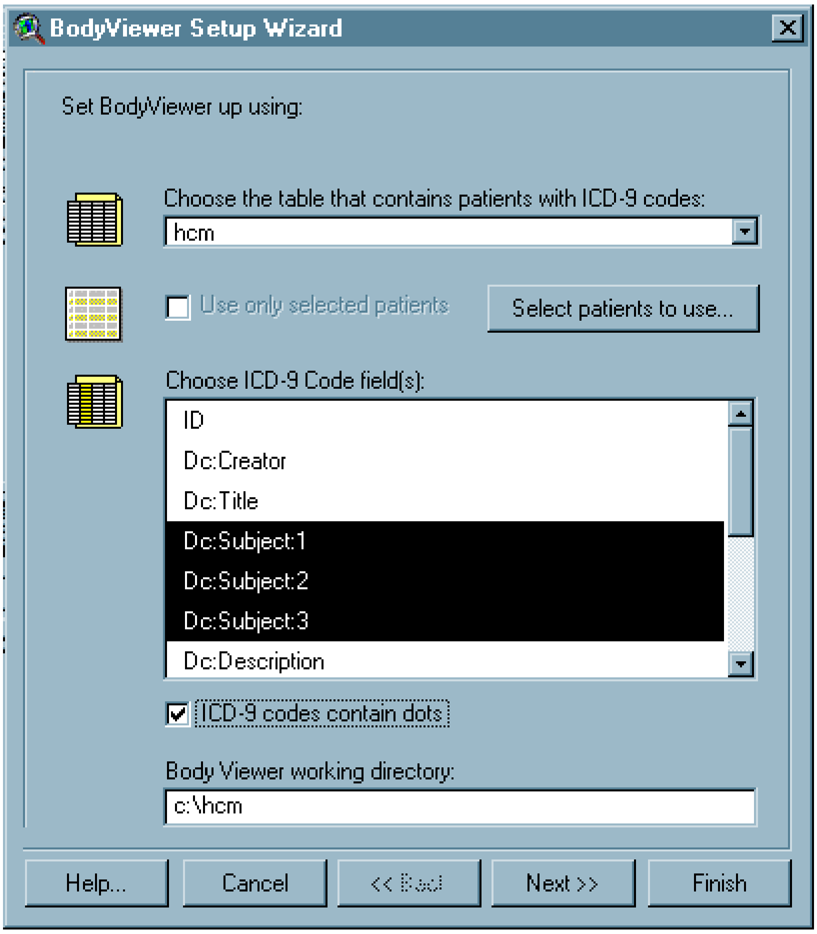


Figure 9.11. Screenshot of BodyViewer "Setup Wizard". BodyViewer can aggregate more than one ICD-9 code field at a time, and so was able to use all three DC subject fields in "hcm" table at the same time.

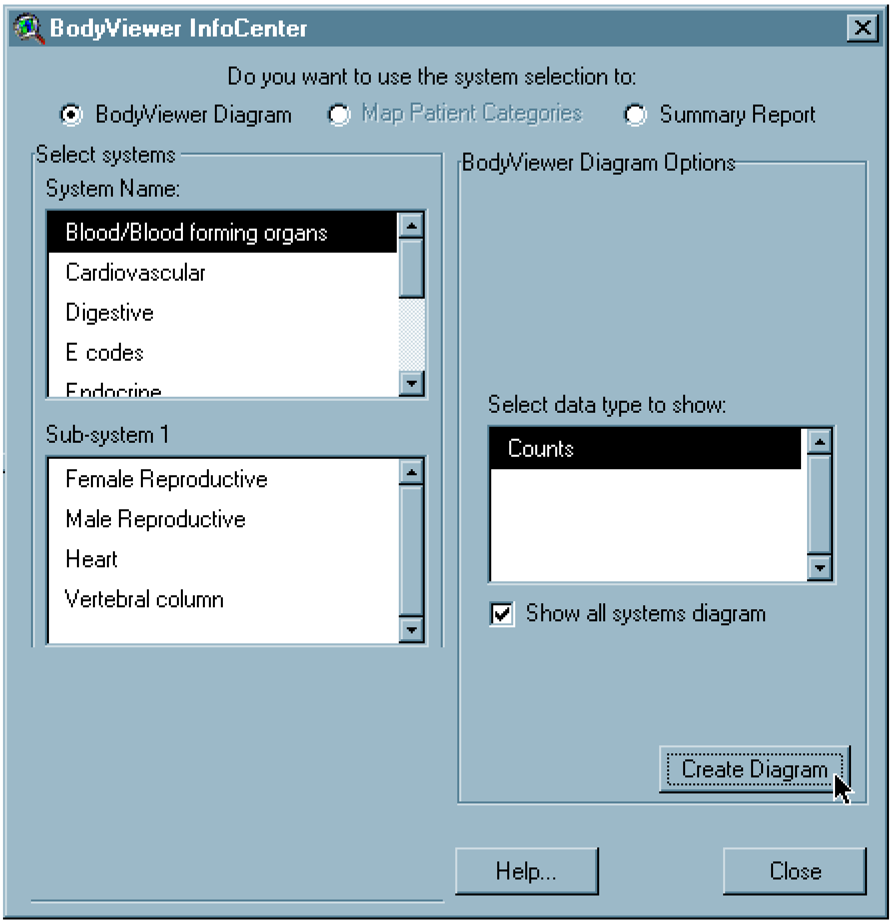


Figure 9.12. Screenshot of the final step in BodyViewer "Setup Wizard". Clicking the "Create Diagram" button will generate an "all systems" human body choropleth map showing resource counts in different organ systems. If "Show all systems diagram" is unchecked, a more detailed human body map will be created for (only) the system or sub-system whose name is selected in one of the two list boxes on the left.

In ArcView, BodyViewer views are not hot-linked to the underlying resource metadata table (*within ArcView*) until we perform what is called "Link Patients" in BodyViewer, though in this case we will be linking resources not patients. Although BodyViewer was able to aggregate three ICD fields in the previous step, the linking can only be done using one DC subject field at a time (Figure 9.13). The result of the linking in ArcView is shown in Figure 9.14. In this regard, the corresponding HealthCyberMap human body maps *on the Web* are superior since the linking query (running on HealthCyberMap Web server) looks in all three DC subject fields in the underlying metadata base.

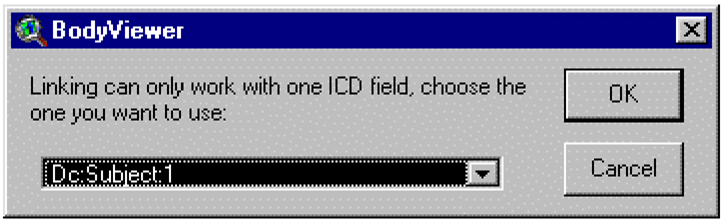


Figure 9.13. BodyViewer linking of its views to the underlying resource metadata table within ArcView GIS can only be done using one DC subject field at a time. In this regard, the corresponding HealthCyberMap human body maps on the Web are superior since the linking query looks in all three DC subject fields in the underlying metadata base (see below).

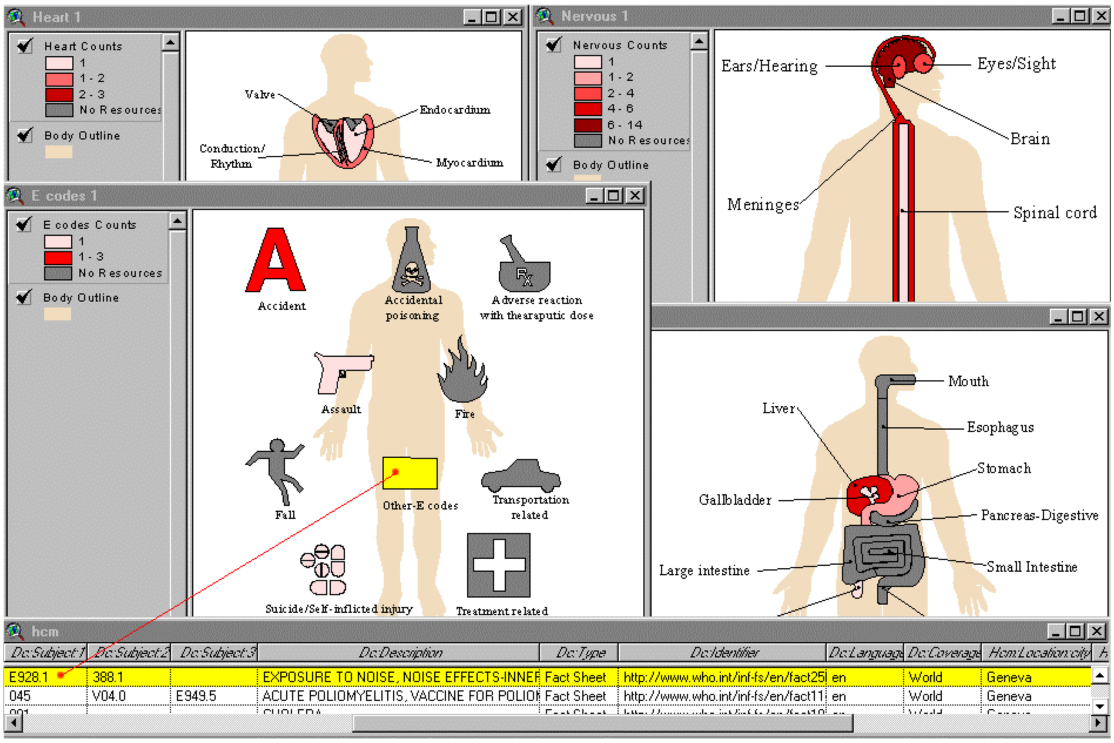


Figure 9.14. Screenshot of four BodyViewer Systems and Sub-systems close-up maps (views) and the table underlying one of them in ArcView GIS showing the "Link Patients (Resources)" feature of BodyViewer in action. Notice the database drill-down (red line); clicking "Other-E codes" on BodyViewer E codes Map selects the corresponding field(s) (in yellow) from HealthCyberMap's metadata table (in this case Exposure to Noise – ICD: E928.1). There are many other BodyViewer close-up diagrams besides those shown here offering much more detail than the general "All Systems" body map.

To export BodyViewer maps (views) to the Web using WebView, we have to first select a Map Unit ("meters") for the view or else WebView will not be able to determine a scale for rendering the detail map and will abort the process (Figure 9.15).

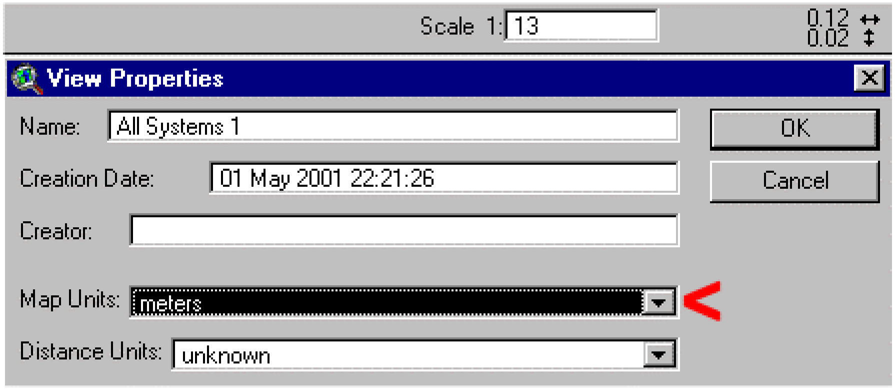
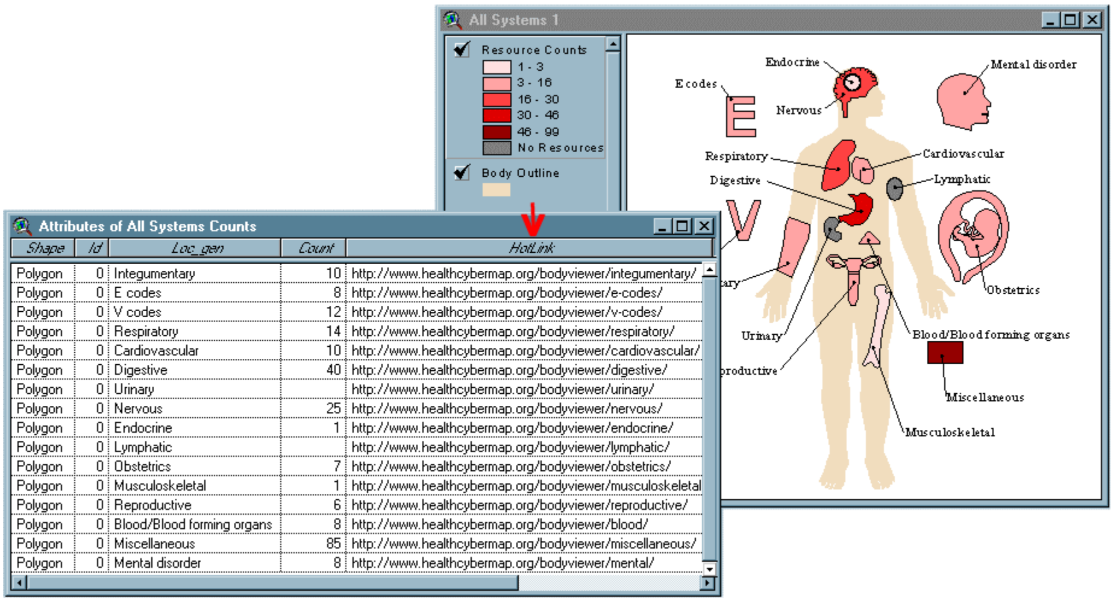


Figure 9.15. Selecting a Map Unit ("meters" in this screenshot) for a BodyViewer view in ArcView. This is necessary for the successful execution of WebView.

The author inserted a HotLink field in BodyViewer map tables to store the Web addresses of corresponding ASP query pages that will run on HealthCyberMap Web server; this field is associated with the HotLink mouse event feature of WebView (Figure 9.16).



Shape	Id	Loc_gen	Count	HotLink
Polygon	0	Integumentary	10	http://www.healthcybermap.org/bodyviewer/integumentary/
Polygon	0	E codes	8	http://www.healthcybermap.org/bodyviewer/e-codes/
Polygon	0	V codes	12	http://www.healthcybermap.org/bodyviewer/v-codes/
Polygon	0	Respiratory	14	http://www.healthcybermap.org/bodyviewer/respiratory/
Polygon	0	Cardiovascular	10	http://www.healthcybermap.org/bodyviewer/cardiovascular/
Polygon	0	Digestive	40	http://www.healthcybermap.org/bodyviewer/digestive/
Polygon	0	Urinary		http://www.healthcybermap.org/bodyviewer/urinary/
Polygon	0	Nervous	25	http://www.healthcybermap.org/bodyviewer/nervous/
Polygon	0	Endocrine	1	http://www.healthcybermap.org/bodyviewer/endocrine/
Polygon	0	Lymphatic		http://www.healthcybermap.org/bodyviewer/lymphatic/
Polygon	0	Obstetrics	7	http://www.healthcybermap.org/bodyviewer/obstetrics/
Polygon	0	Musculoskeletal	1	http://www.healthcybermap.org/bodyviewer/musculoskeletal
Polygon	0	Reproductive	6	http://www.healthcybermap.org/bodyviewer/reproductive/
Polygon	0	Blood/Blood forming organs	8	http://www.healthcybermap.org/bodyviewer/blood/
Polygon	0	Miscellaneous	85	http://www.healthcybermap.org/bodyviewer/miscellaneous/
Polygon	0	Mental disorder	8	http://www.healthcybermap.org/bodyviewer/mental/

Figure 9.16. Screenshot showing the HotLink field that has been added to the underlying table of a BodyViewer view in ArcView.

9.4.6 HealthCyberMap's World Map in ArcView

The author used the basic World map dataset that ships with ArcView GIS 3.1 as a project example (world.apr), and added a HotLink field to the "Countries ('98)" table (Figures 9.17 and 9.18). This field was associated with the HotLink mouse event in WebView wizard when the author generated HealthCyberMap's World Map for the Web. Clicking a country on HealthCyberMap's World Map (on the Web) will launch the appropriate country query to retrieve only those resources authored in and/ or published in the clicked country. Query results will *always* reflect the latest updates carried on the metadata base without the need to change any code.

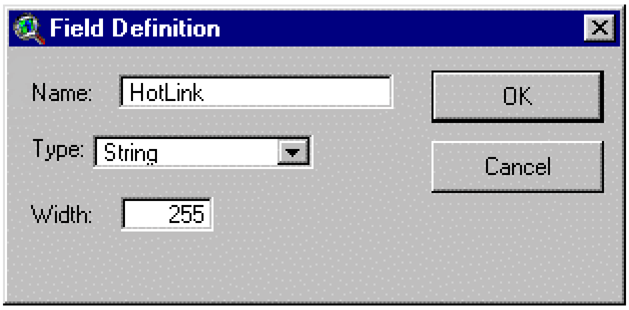


Figure 9.17. Screenshot of ArcView "Field Definition" dialogue box which is used to add a new field to an existing table and define its properties. In this screenshot, the properties of the HotLink field are shown. The field is defined as "String". It can hold up to 255 character (the maximum allowed for this type in ArcView) and is used to store Internet addresses (URIs).

Attributes of Countries ('98)

Shape	Fips_cntry	Gmi_cntry	Cntry_name	HotLink	Pop_cntry	Sqkm_cntry	Curr_type
Polygon	AS	AUS	Australia	http://www.healthcybermap.org/australia.asp	17827520	7706142.000	Australia Dollar
Polygon	BE	BEL	Belgium	http://www.healthcybermap.org/belgium.asp	10032460	30479.609	Franc
Polygon	CA	CAN	Canada	http://www.healthcybermap.org/canada.asp	28402320	9904700.000	Dollar
Polygon	FR	FRA	France	http://www.healthcybermap.org/france.asp	57757060	546728.875	Franc
Polygon	GM	DEU	Germany	http://www.healthcybermap.org/germany.asp	81436300	356108.812	Mark
Polygon	IT	ITA	Italy	http://www.healthcybermap.org/italy.asp	57908880	300979.500	Lira
Polygon	SZ	CHE	Switzerland	http://www.healthcybermap.org/switzerland.asp	6713839	41178.398	Franc
Polygon	UK	GBR	United Kingdom	http://www.healthcybermap.org/united_kingdom.asp	56420180	243137.203	Pound Sterling
Polygon	US	USA	United States	http://www.healthcybermap.org/united_states.asp	258833000	9450720.000	Dollar

Figure 9.18. Screenshot of the "Countries ('98)" table in ArcView showing the inserted HotLink field that stores the addresses of the "resources by country" ASP query pages on HealthCyberMap server.

It is also possible to *join* the resource metadata and World Countries tables (within ArcView) based on the values of a common field (country names – Figure 9.19). This kind of functionality is not possible with WebView.

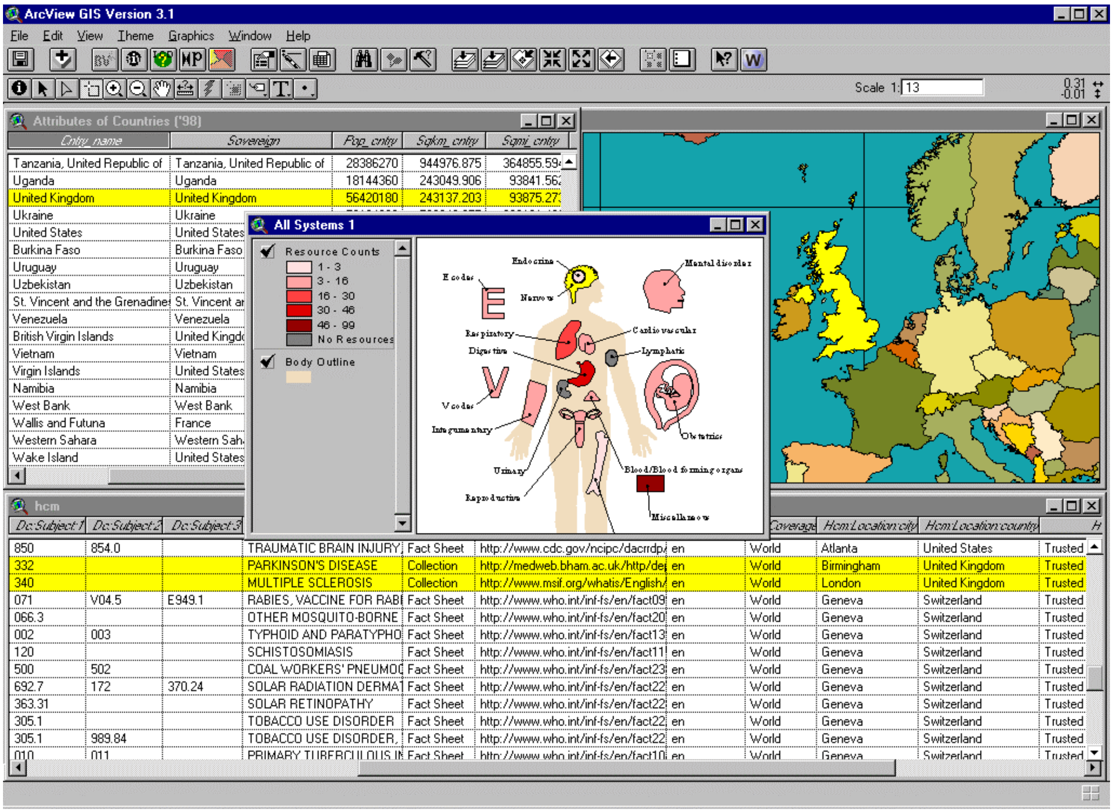


Figure 9.19. Two Web resources from the United Kingdom (Birmingham and London) related to the Nervous System (Parkinson's Disease - ICD: 332 and Multiple Sclerosis - ICD: 340) are selected (in yellow) at the same time in all open maps (World Map and BodyViewer All Systems body map) and tables ("Countries ('98)" and "hcm"). To achieve this functionality, the author first joined the two tables based on the values of a common field (country names). This kind of functionality is not possible with WebView.

9.5 HealthCyberMap Web Interface

9.5.1 HealthCyberMap's BodyViewer Maps on the Web

HealthCyberMap's BodyViewer maps are available on the Web at the following address: <http://healthcybermap.semanticweb.org/bodyviewer/> (Figure 9.20). These human body topical maps can be used to visually browse selected medical/ health resources on the Web by clinical subject according to ICD-9-CM classification.

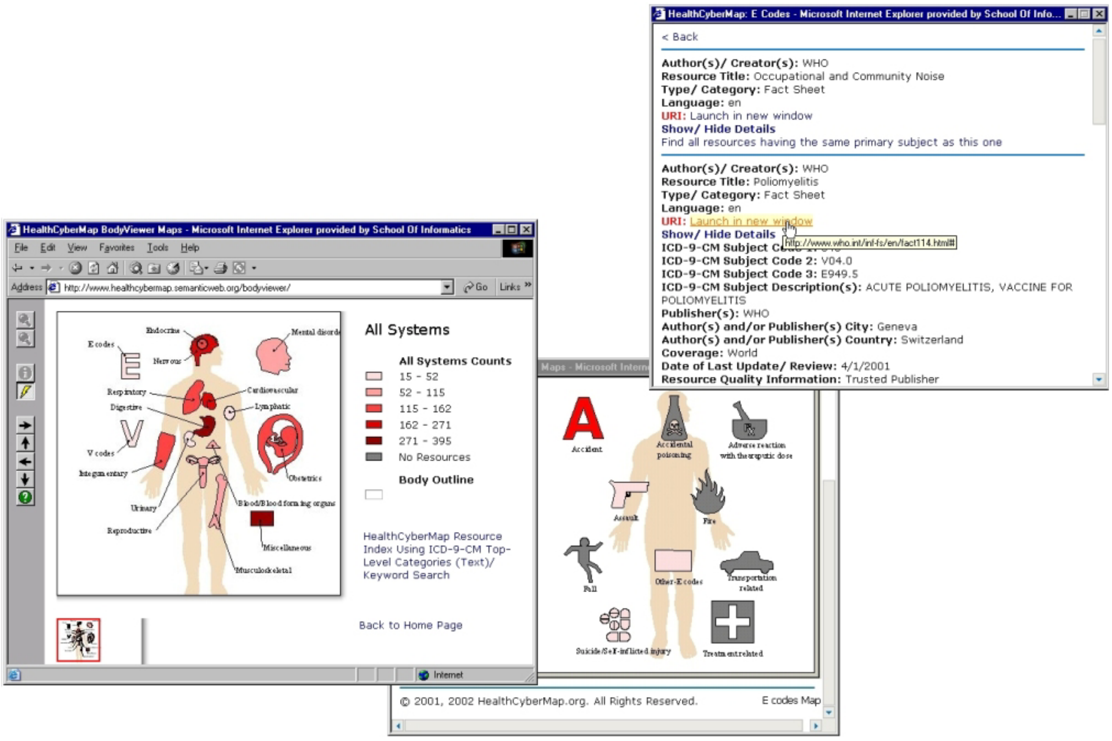


Figure 9.20. Screenshot of HealthCyberMap's BodyViewer Maps on the Web (<http://healthcybermap.semanticweb.org/bodyviewer/>). These hierarchical human body topical maps with semantic zoom can be used to visually browse selected medical/ health resources on the Web by clinical subject. The resources have been categorised and spatialised to the different human body organs on the map according to a clinical coding scheme.

Clicking a human body icon on these maps triggers a server-side dynamic query. This is for example the pre-formulated SQL query that currently runs on HealthCyberMap server in real-time to retrieve resources having E codes (codes for External Causes of Injury and Poisoning) in any of their DC subject fields (<http://healthcybermap.semanticweb.org/bodyviewer/e-codes.asp>—for full code of this page, see Appendix 1: Example ASP page from HealthCyberMap):

```
sql = "SELECT hcm.[dc:Creator], hcm.[dc:Title],
hcm.[dc:Subject:1], hcm.[dc:Subject:2], hcm.[dc:Subject:3],
hcm.[dc:Description], hcm.[dc:Publisher], hcm.[dc:Date],
hcm.[dc:Type], hcm.[dc:Identifier], hcm.[dc:Language],
hcm.[dc:Coverage], hcm.[hcm:Location:city],
hcm.[hcm:Location:country], hcm.[hcm:Quality], hcm.[hcm:Comment] FROM
hcm WHERE (((hcm.[dc:Subject:1]) Like 'E%')) OR (((hcm.[dc:Subject:2])
Like 'E%')) OR (((hcm.[dc:Subject:3]) Like 'E%'))"
```

In the above code snippet, `hcm` is the name of the database table. All three DC subject fields in each resource record are searched for matching ICD-9-CM codes.

9.5.1.1 Semantic Zooming

HealthCyberMap human body maps adopt a semantic zooming approach. With a conventional geometric zoom (as in HealthCyberMap's World Map) all objects change only their size; with semantic zoom they can additionally change shape, details (not merely size of existing details) or, indeed, their very presence in the display, with objects appearing/ disappearing according to the context of the map at hand [87]. A good example of semantic zoom in HealthCyberMap is illustrated in Figure 9.2 above; the lung, digestive system, and other body organs disappear (when the heart icon is clicked on the first map), and new (not just enlarged) details of the cardiovascular system appear in consecutive maps.

9.5.2 HealthCyberMap's World Map on the Web

HealthCyberMap's World Map Web interface is available on the Web at the following address: <http://healthcybermap.semanticweb.org/world_map/> (Figures 9.1 and 9.21). The maps can be used to browse Web resources by country of provenance.

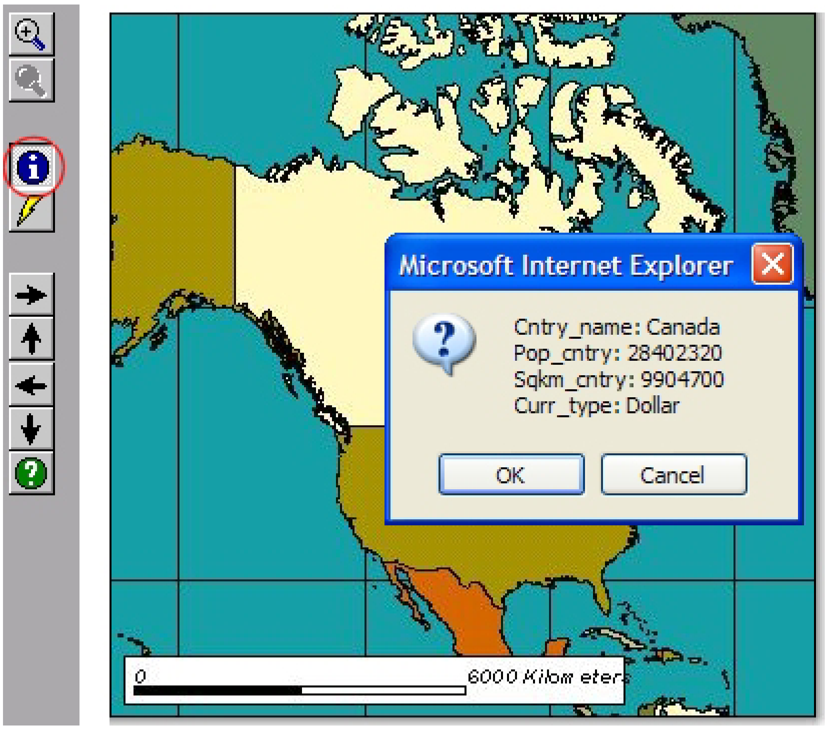


Figure 9.21. Screenshot of HealthCyberMap's World Map on the Web (<http://healthcybermap.semanticweb.org/world_map/>) showing the Identify button function.

In Figure 9.21 above, note the depressed (selected) Identify button  in the toolbar on the left and the pop-up message box displaying additional information on Canada after the latter was clicked. Any other information, e.g., health related, could have been displayed instead or additionally depending on what is available in the underlying table in ArcView from which these attributes are "pulled" and associated with the hypermaps when they are generated by WebView. Other buttons on the toolbar allow zooming in and out and panning the map, and activating the HotLink function  (instead of Identify), so that clicking a country retrieves the bibliographic cards of resources associated with it in a separate pop-up window (Figure 9.1). Unlike the dynamic, always up-to-date information (database queries) associated with the HotLink button, information associated with the Identify button is static (exported with the maps from ArcView and detached from the latter) and can be only updated (if needed) by regenerating the maps using WebView.

9.5.3 Online Map Interface Help

Online Help is available for both BodyViewer and World Map Web interfaces. Clicking the green question mark '?' button in Figures 9.20 and 9.21 above will display Help instructions in a separate pop-up window (Figure 9.22).

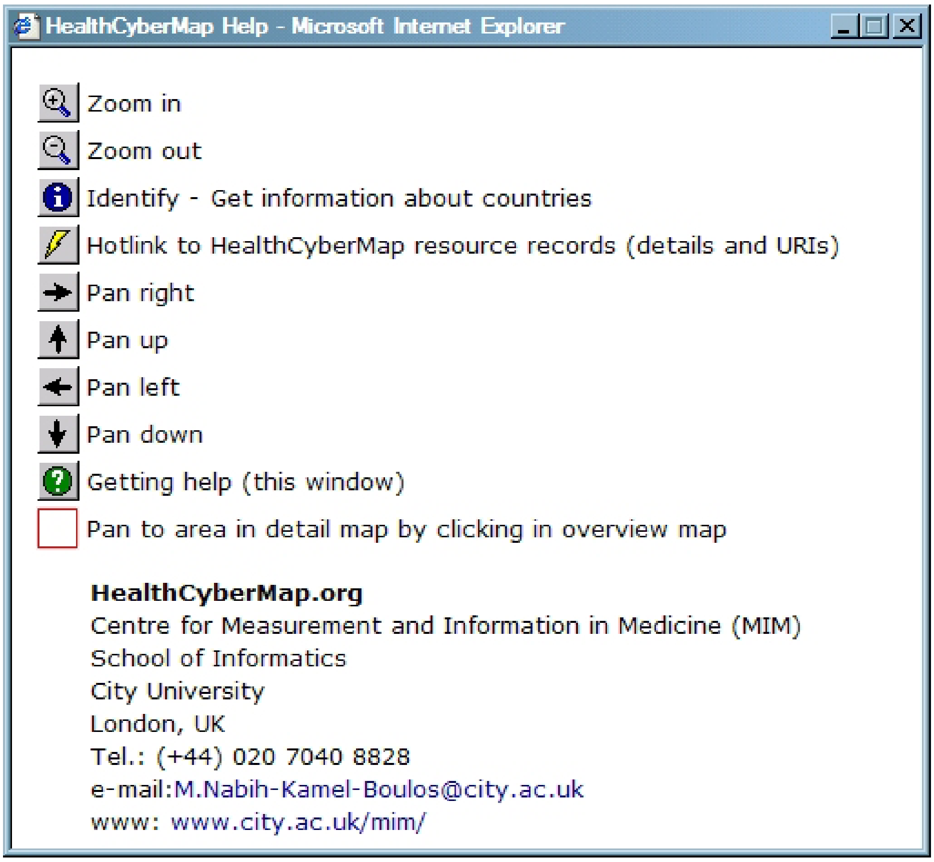


Figure 9.22. HealthCyberMap Help Window explaining the functions of the different map interface buttons.

9.5.4 HealthCyberMap's Browse by Resource Type Map

This is the only map in the pilot HealthCyberMap service that was created outside ArcView GIS using Jasc Paint Shop Pro 7 (<http://www.jasc.com>) to edit the image. The author then used Mapedit v2.63 for Windows 9x/NT, an imagemap editor from Boutell.Com, Inc. (<http://www.boutell.com/mapedit/>), to define (draw) the required hotspots on the image, as well as the actions (ASP query links) associated with these hotspots, and Mapedit generated the necessary (client-side) HTML code (Figure 9.23). Part of the HTML code for the map shown in Figure 9.23 follows:

```
<img src="type.jpg" usemap="#type" width=445 height=523 border="0">
<map name="type">
(...)
<area shape="rect" alt="Digital Atlases, e.g., of Dermatology"
coords="237,311,443,467"
href="javascript:Start('http://www.healthcybermap.org/digiatlas.asp')"
; title="Digital Atlases, e.g., of Dermatology">
(...)
<area shape="default" nohref>
</map>
```

Clicking on the rectangular hotspot defined in the above code snippet would launch a query on the server (<http://www.healthcybermap.org/digiatlas.asp>) and open a separate window to display the query results in it.

The resultant map (Figure 9.24) structures and organises health information resources by type based on the DC type field into the following categories:

- Electronic Journal Articles/ Papers/ Abstracts
- Guidelines
- e-books
- Fact Sheets
- Collections
- Multimedia/ Audio-Visual Material
- Digital Atlases, e.g., of Dermatology
- Interactive Resources
- Health Services
- Software
- Events

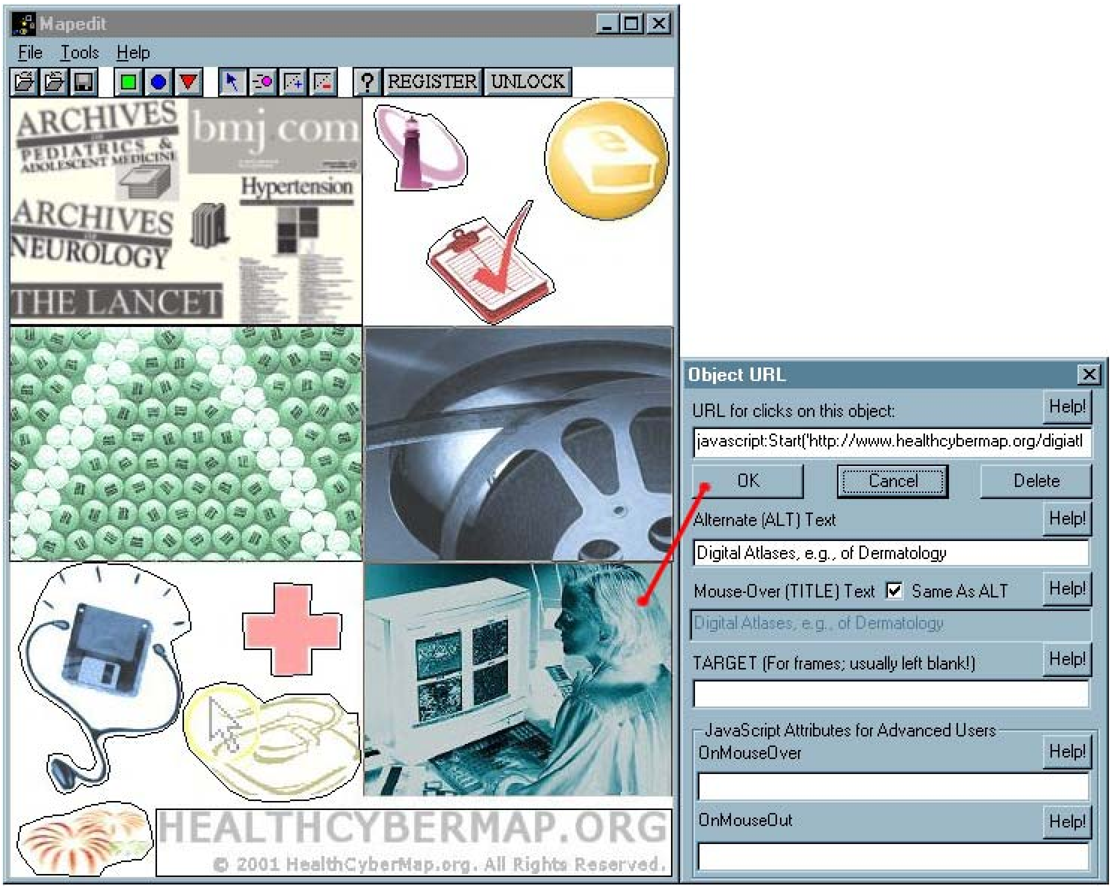


Figure 9.23. Screenshot of Mapedit v2.63 showing the map the author has created in it to browse HealthCyberMap resources by resource type based on the DC type field in the underlying database (this map is available at: <http://healthcybermap.semanticweb.org/type.htm>). A ToolTip text can be also defined for each hotspot.

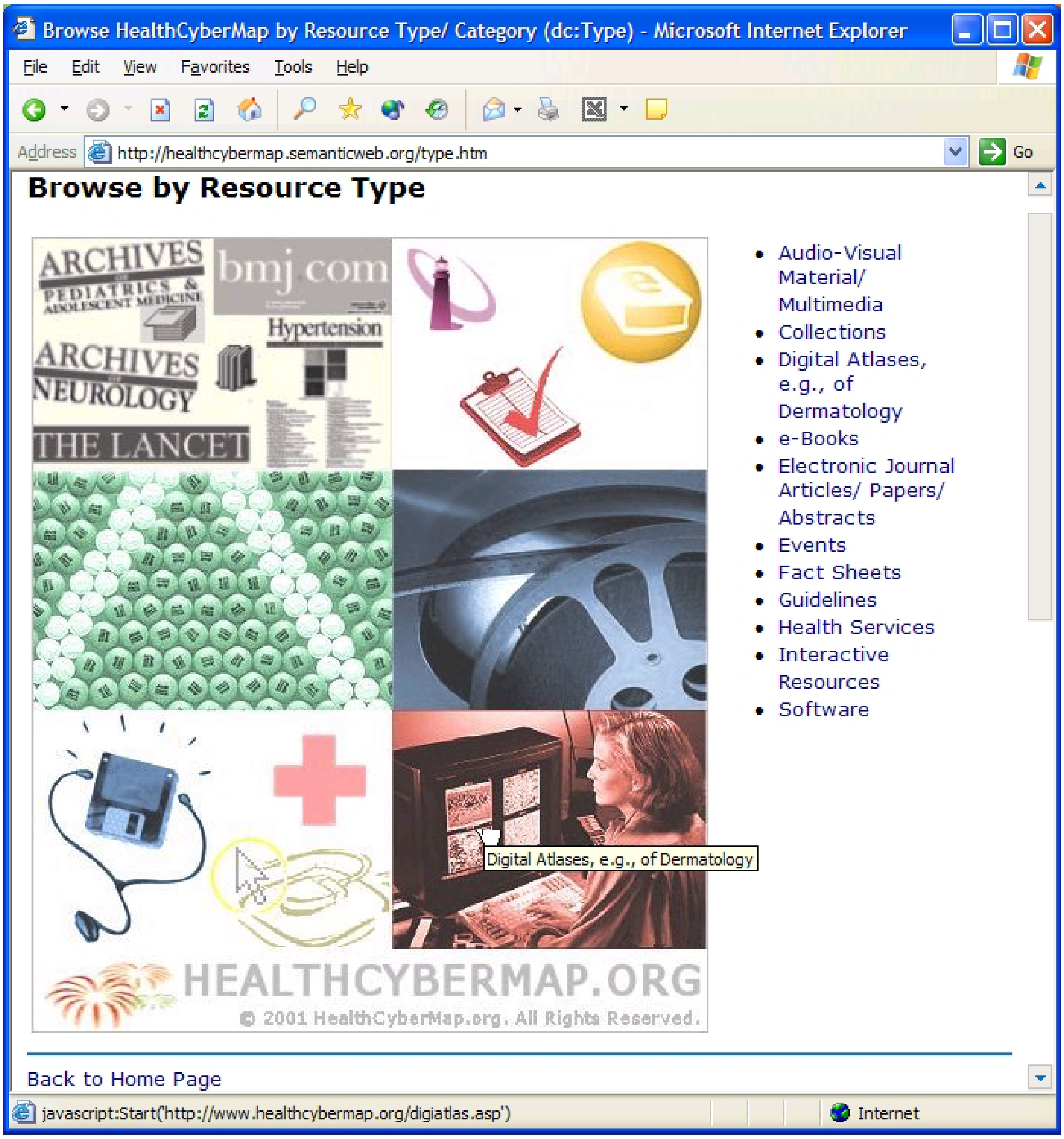


Figure 9.24. Screenshot of HealthCyberMap's "Browse by Resource Type" hypermap (<http://healthcybermap.semanticweb.org/type.htm>).

9.5.5 Complementary Textual Interfaces

There are also alternative ways to browse HealthCyberMap's resource metadata base in case users find it difficult to visually locate what they want on the maps. These alternative interfaces include multilingual interfaces for browsing resources by resource language based on DC language field (Figure 9.25), a textual resource index (Figure 9.26), and HealthCyberMap Semantic Subject Search Engine (Chapter 8).

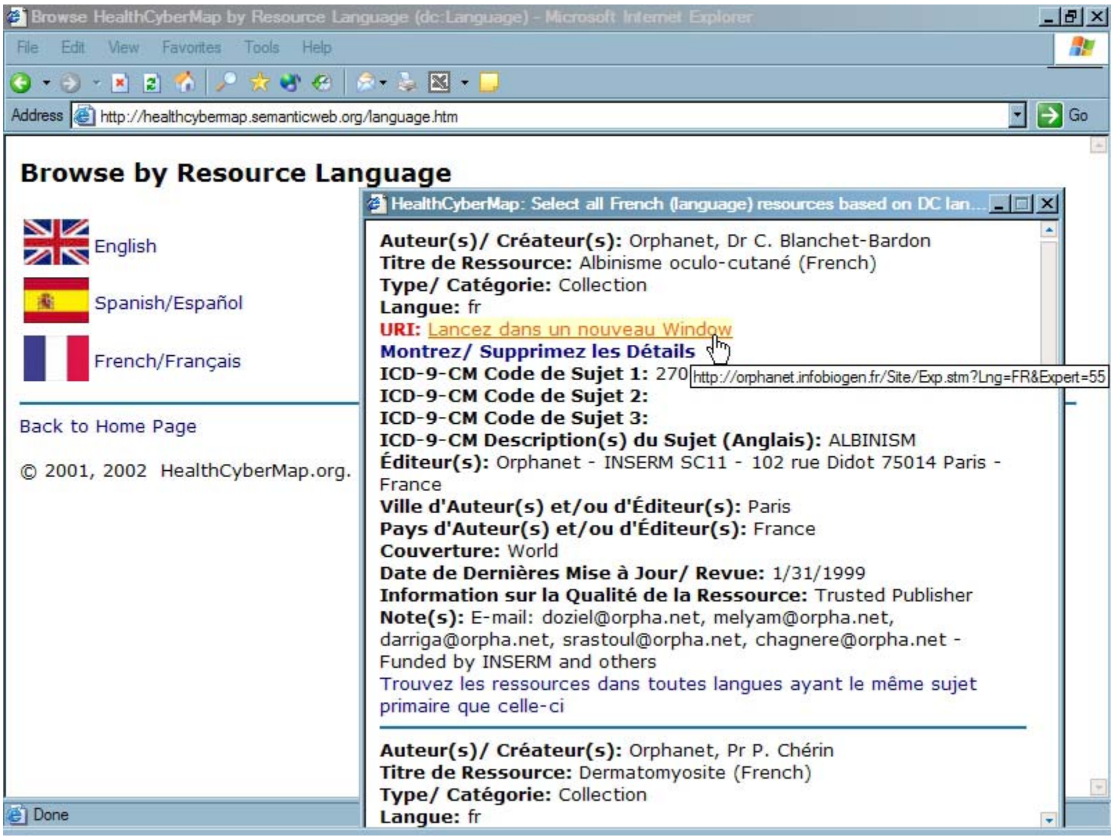


Figure 9.25. HealthCyberMap pilot French language interface for browsing French language resources (<http://healthcybermap.semanticweb.org/language.htm>).

Ideally, the classification of resources by language should be transparently integrated into other interfaces (as part of the proposed HealthCyberMap future work on customisation—Chapter 14) rather than offered as standalone interfaces.

HealthCyberMap's textual resource index is meant to be a directory of topical resource categories and could be considered the textual equivalent of BodyViewer's human body maps. The textual resource index in the current pilot service is only based on ICD-9-CM top-level (broadest) categories; clicking one of those categories triggers a query that returns a large number of resources covering different narrower topics (codes), all belonging to the main category that has been clicked (e.g., all "Diseases of the Blood and Blood-forming Organs"—Figure 9.26). A proposed future improvement is to enable users to browse and select hierarchical sub-levels (subcategories) of the main (top-level) ICD-9-CM categories to focus queries on smaller groups of resources/ narrower topics (Figure 15.4).[††]

[††]This is possible because resources in the current HealthCyberMap pilot metadata base are indexed using the most specific (narrow) ICD-9-CM codes representing their topics.

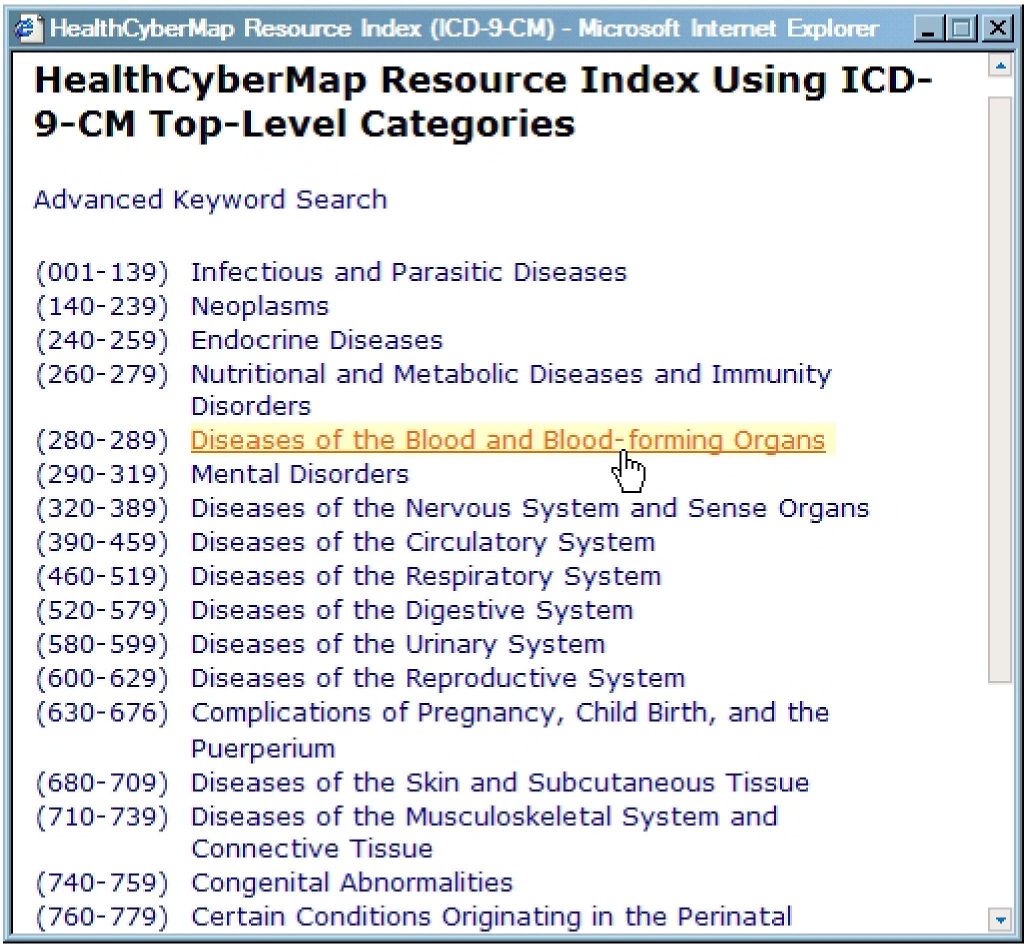


Figure 9.26. HealthCyberMap's textual resource index, a directory of topical resource categories.

9.5.6 Pagination of HealthCyberMap Metadata Base Query Results

Pagination becomes an important issue when we have a large number of resource records to display. HealthCyberMap ASP query pages are programmed to display only ten resource records per page, with "<< First | Next > | < Previous | Last >>" navigation links at the bottom of each page to move back and forth between returned resource records (Figure 9.27—for code, see Appendix 1: Example ASP page from HealthCyberMap).

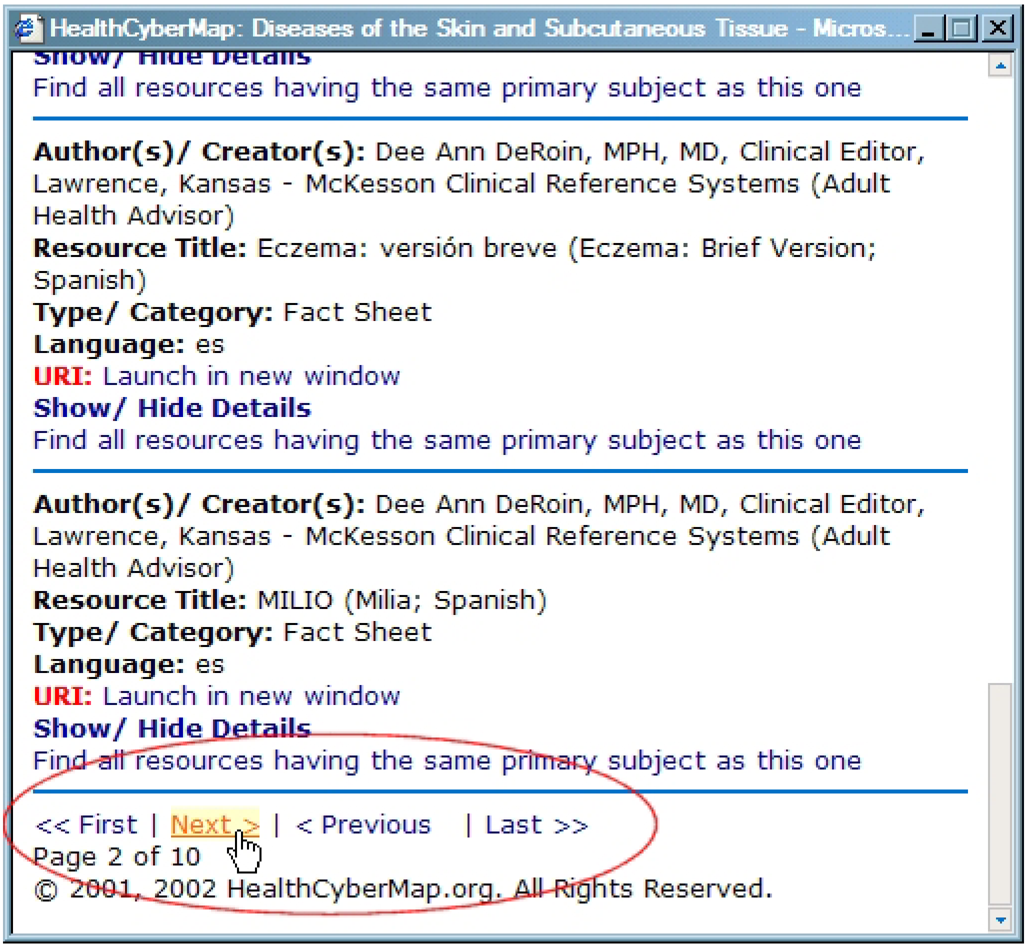


Figure 9.27. Pagination of HealthCyberMap metadata base query results. HealthCyberMap ASP query pages display only ten resource records per page, with "<< First | Next > | < Previous | Last >>" navigation links at the bottom of each page to move back and forth between returned resource records.

9.6 Maintenance of HealthCyberMap Web Maps and Managing Broken Resource Links

Since WebView does not allow the *dynamic* generation of Web maps from ArcView, some of HealthCyberMap Web maps will ultimately need to be manually regenerated using WebView when the underlying data change if this change has implications on the maps' appearance. In cases when only map attribute data change without a corresponding effect on map appearance, e.g., updating the address of a Web resource in the metadata base, nothing needs to be done; the same ASP query pages (unmodified) will retrieve the latest updates. HealthCyberMap GIS-generated Web maps can be grouped into:

- Choropleth maps that need to be regenerated when the underlying data are updated, e.g., the BodyViewer maps as the colour shades of the various body organ systems in these maps reflect the number of resources associated with them, and so will change whenever the database is updated (resources added and/ or deleted). In this case, the Web maps must be recreated in ArcView using WebView then uploaded to the Web server to replace older ones. Associated query pages need not be changed.

- Chorochromatic maps that do not usually need to be updated, e.g., the world maps. Whenever new resources are added (or existing ones updated/ deleted), they will automatically appear (or disappear) in the query results when the corresponding countries on the map are clicked. There is no need to change the map as long as the addresses of the dynamic ASP query pages on HealthCyberMap server do not change (e.g., <http://healthcybermap.semanticweb.org/canada.asp>). The only reason to regenerate these maps would be if some of their underlying attribute values that are used with WebView Identify function change, or if the political boundaries between some countries change (which is not very frequent).

9.6.1 Managing Broken Resource Links (The '404' Problem)

The success of a distributed information system such as the Web for research depends on the long-term consistency of the inter-links between online resources and the persistence of the links that are provided in the catalogues, indexes and listings of resource discovery services. The National Library of Australia has recently published a document titled "Managing Web Resources for Persistent Access" on how to overcome the deadly broken link message "HTTP 404 Not Found" [88]. It appears that there is *no* fully automated solution for this problem. Moreover, any potential solution does not lie solely in the hands of online catalogues and services like HealthCyberMap.

What HealthCyberMap can do is to regularly run a link checker on its database of resource addresses to identify links that are no more functional (and possibly exclude/ delete them)—Figure 9.28. Replacing these links with more current, functional versions still requires manual intervention. HealthCyberMap should also perform regular quality checks on listed resources (in case they have changed or the information they contain is no more valid). This is another manual task for humans to complete, as it depends on their discernment capabilities.

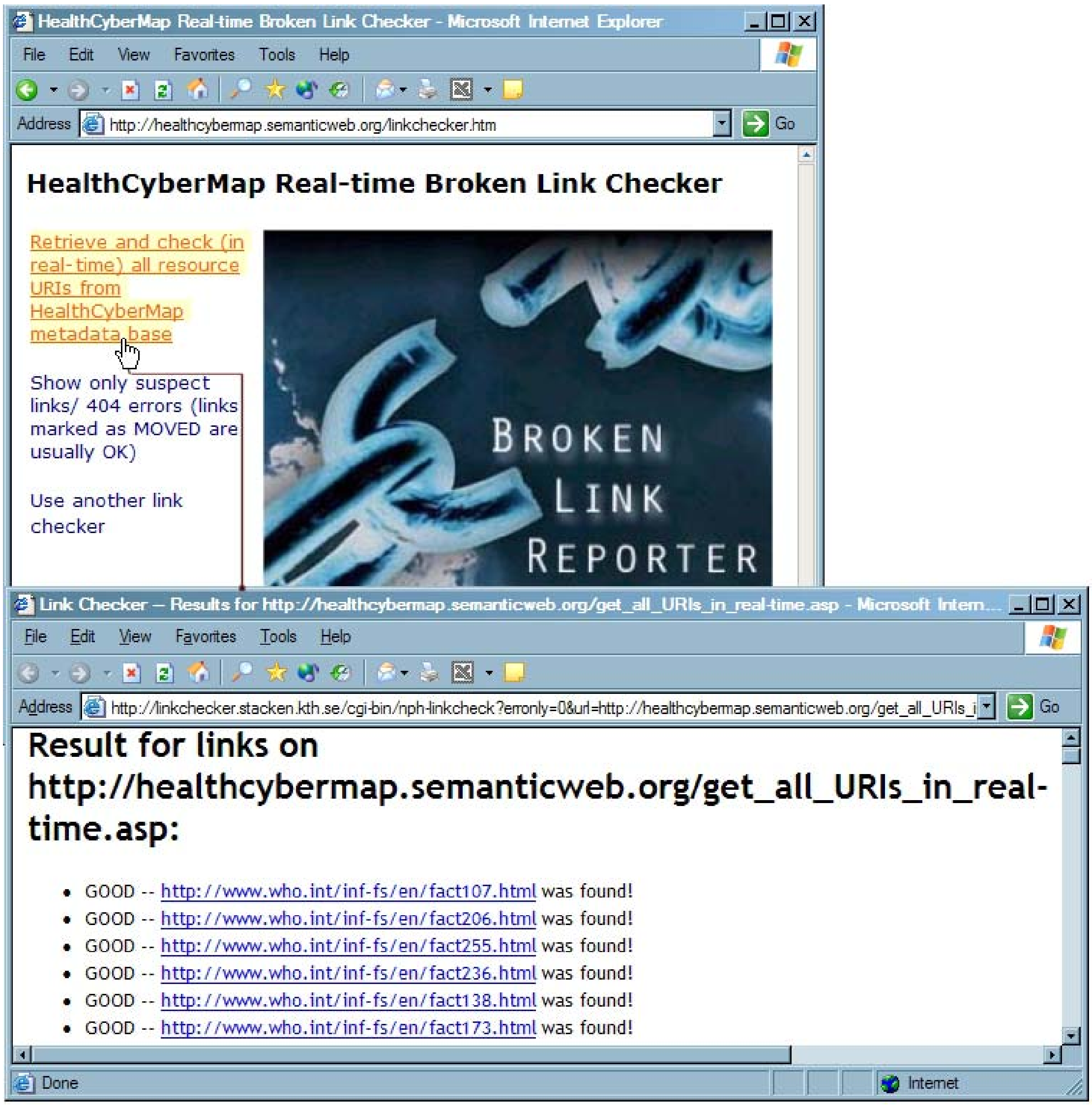


Figure 9.28. HealthCyberMap real-time broken link checker (<http://healthcybermap.semanticweb.org/linkchecker.htm>). A dynamic ASP query page extracts all resource URIs (Uniform Resource Identifiers—Internet addresses) from HealthCyberMap's metadata base and passes them to an external link checker.

The National Library of Australia guidelines are intended to assist those responsible for the management of published online materials (creators, publishers, libraries and other information repositories) to make sure links made to those resources continue to work. Even the most sophisticated solution presented in these guidelines, which is using a persistent identifier system, e.g., DOI (Digital Object Identifier—<http://www.doi.org/>) and PURL (Persistent Uniform Resource Locator—<http://purl.oclc.org/>), requires good management or else it will fail. When resources are moved, the current location must be associated with the persistent identifier in whatever system is chosen. This is usually achieved by a *resolver* database. There is *no* system that does not require such management [88]. (N.B.: A persistent identifier is a name for a resource which will remain the same regardless of where the resource is located; thus links to the resource will continue to work even if it is moved.)

Some of these guidelines also apply to HealthCyberMap *as a service Web site*, so that links to it from other sites around the Web remain valid. For example, it is advised to organise a Web site in such a way as to reduce the need to move material around, and if the need arises to revamp the site for a "new look" to do so by redesigning the graphics, presentation and navigation *without* moving any of the underlying files on which it is built. Also, it is recommended to keep older material accessible (without changing its address), as it may be valuable to others for continuing reference, research or maybe for historical purposes [88].

9.7 Conclusion

HealthCyberMap (<http://healthcybermap.semanticweb.org>) is a Web-based service for healthcare professionals and librarians, patients and the public in general that aims at mapping parts of medical/ health information resources in cyberspace in novel ways to improve their retrieval and navigation. HealthCyberMap pilot service currently provides six different interfaces to its metadata base, which has over 1600 resource records in it. Some of these interfaces are visual (maps for browsing resources by clinical/ health topic, by provenance and by type), while others are textual (multilingual interfaces for browsing resources by language, and a directory of topical resource categories, besides HealthCyberMap Semantic Subject Search Engine). The visual maps are used to locate, launch medical/ health resources on the Web, and display their bibliographic metadata records.

HealthCyberMap is an ArcView 3.1 project that uses GIS spatialisation methods to generate interactive navigational cybermaps from an underlying resource metadata base. (The visual map used to browse resources by type is the only map created outside ArcView in Boutell's Mapedit v2.63.) WebView, the Internet extension to ArcView, publishes HealthCyberMap ArcView Views as Web client-side imagemaps. The basic WebView set-up does not support any GIS database connection, and published Web maps become disconnected from the original project. A dedicated Internet map server would be the best way to serve HealthCyberMap database-driven interactive Web maps, but is an expensive and complex solution to acquire, run and maintain. In this chapter, we have described HealthCyberMap's simple, low-cost method for "patching" WebView to serve hypermaps with dynamic database drill-down functionality on the Web (using dynamic ASP pages to query the same metadata base used in ArcView and registered on HealthCyberMap server as an ODBC data source). The proposed solution is currently used for publishing HealthCyberMap GIS-generated navigational information maps on the Web while maintaining their links with the underlying resource metadata base.

HealthCyberMap human body maps with their "semantic zooming" feature allow the navigation of medical/ health Internet resources by body location/ system according to ICD-9-CM codes, which act as HealthCyberMap medical ontology and are used to describe resource topics (subjects) in the metadata base of the current pilot service. The "semantic distance" between two resources on these maps depends on how close (or related) the two resources are from a semantic perspective based on the "semantic locations" of their topics within ICD-9-CM.

The author believes that the visual categorisation of medical/ health Internet resources using familiar spatial metaphors for image-word association could give users a broad overview and understanding of what is available in this complex conceptual space and help them navigate it more efficiently and effectively. HealthCyberMap also introduces a useful form of cyberspatial analysis for the detection of topical coverage gaps in the resource metadata base using its human body topical choropleth (shaded) maps of resource counts. Detected topical coverage gaps should be then addressed by information providers.

The author also believes his map serving approach as adopted in the current HealthCyberMap pilot service has been very successful, especially in cases when only map attribute data change without a corresponding effect on map appearance. It should be also possible to use the same solution to publish other interactive GIS-driven maps on the Web, e.g., maps of real world health problems. The management of broken resource links in HealthCyberMap's metadata base was also discussed.

10 HealthCyberMap's Dynamic Problem to Knowledge Linking Service Demonstrator

10.1 Introduction

Many information needs arise during everyday clinical practice, including questions related to diagnosis, determining risks/ prognosis, planning an optimum investigative or therapeutic strategy, disease prevention and patient education material, etc. The quality and outcomes of patient care will suffer if these information needs are unmet or are answered with inaccurate, non-current or misleading (low quality) information [89].

Contextual relevance [5] is an important aspect of medical information quality that is frequently overlooked. It is essentially an information retrieval issue. An information resource that is excellent on its own merits can end up becoming mere distracting noise if it does not properly address the user's specific knowledge needs, or does not fit the clinical context in which it was recalled.

Problem to knowledge linking (PKL) aims at providing contextually appropriate medical knowledge in the right place and at the right time. It should be a major goal when designing the Electronic Patient Record (EPR), since mapping EPR problems to the appropriate, contextually relevant knowledge resources answering these problems is one key to informed clinical decision-making and hence better healthcare outcomes [5]. These knowledge resources can be present on the Web or elsewhere, e.g., on a private Intranet.

The ideal online medical library/ knowledge service should act as a "contextual medical knowledge provider" within a Clinical Information System, linking problem-specific knowledge to real patient data (EPR). This can support clinical decision making, improve patient care and educate the student, professional and patient.

10.1.1 Clinical Codes in HealthCyberMap

By tagging a medical Web resource (or a metadata record of it) with clinical codes, we are automatically establishing the relationships (as defined by the coding scheme in use) between this resource and other related (tagged) resources, and also similarly coded EPRs. This is the basis of PKL.

HealthCyberMap has been designed to make use of existing clinical coding schemes as a common backbone for:

- language-neutral topical resource indexing (concept-based/ code-based description of resource subjects)—Chapters 6, 7 and 8;
- topical resource visualisation and navigation using suitable metaphors, e.g., hierarchical human body diagrams—Chapter 9; and
- enhanced information retrieval and linking, e.g., EPR to HealthCyberMap resources to support clinical decision making.

10.1.2 Towards Semantically Superior Interoperability: The Semantic Web and Its Services

The Semantic Web initiative (<http://www.w3.org/2001/sw/>) aims at creating a Web where information semantics are represented in a form "understandable" by machines as well as by humans. In the case of medical information, this can be achieved using clinical codes. This will pave the way for more "intelligent" machine-to-machine communication, e.g., between an EPR system and a digital medical library or online information service, and ultimately empower humans.

A Web Service is a unit of application logic providing data and services to other applications accessing the service via ubiquitous Web protocols and data formats such as HTTP (HyperText Transfer Protocol). In the case of HealthCyberMap, the author thought of a very useful Semantic Web Service, namely a PKL Service to be consumed by different EPR clients. Linking is possible as both the EPR and HealthCyberMap will be using the same clinical coding system or two different schemes with reliable cross mapping.

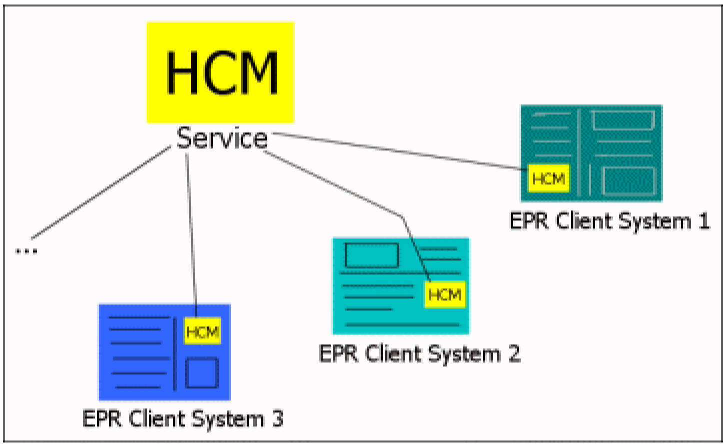


Figure 10.1. One Problem to Knowledge Linking service and different clients/ interfaces (HCM=HealthCyberMap).

10.2 Pilot Implementation and Online Demonstrator

HealthCyberMap relies on stand-alone metadata (in a central database—cf. peripheral metadata embedded in the Web resources themselves) based on the Dublin Core (DC) metadata set [22] with HealthCyberMap's own extensions for resource quality and geographical provenance (see Chapter 7).

Clinical codes are used to populate the DC subject field in HealthCyberMap's metadata base. HealthCyberMap currently allows three DC subject fields per resource record for richer descriptions. The current HealthCyberMap pilot Web implementation uses ICD-9-CM codes (International Classification of Diseases, ninth revision, Clinical Modification—[83]).

The database is registered on HealthCyberMap server (a Windows 2000 Server) as an ODBC Data Source (Open Database Connectivity). ASP Server Pages (Active Server Pages) are used to query the database and present the results on the Web. The ASP pages returned to the calling clients (e.g., an EPR) only contain query results. The actual SQL (Structured Query Language) and ASP code as found in the ASP pages stored on the server is never sent to peripheral clients. Below is the SQL query the author is using.

```
SELECT * FROM hcm WHERE ((hcm.[dc:Subject:1]) LIKE
'%"+Request.querystring("Searchtext")+"%') OR ((hcm.[dc:Subject:2])
LIKE '%"+Request.querystring("Searchtext")+"%') OR
((hcm.[dc:Subject:3]) LIKE '%"+Request.querystring("Searchtext")+"%')
```

In the above code, `hcm` is the name of the database table and `Searchtext` is replaced at runtime with the ICD code that is passed from the EPR as an SQL query argument to HealthCyberMap. The search is done in all three DC subject fields.

In the pilot HealthCyberMap PKL Web demonstrator (<http://healthcybermap.semanticweb.org/pk.htm>—Figure 10.2), the author assumed an EPR that codes diagnoses in ICD-9-CM, and since HealthCyberMap crisply describes the subjects of the Web resources stored in its metadata base using the same codes, a link between the two systems in the form of:

`http://healthcybermap.semanticweb.org/icd.asp?SearchText=PUT_SINGLE_ICD-9-CM_CODE_HERE`

(Acne vulgaris – ICD code: 706.1 in the example shown in Figure 10.2) will be all what we need to perform the linking query and dynamically link the EPR to contextually relevant knowledge and guidelines. The ICD code from the EPR is passed as an SQL query argument to HealthCyberMap (as the value of "`SearchText`").

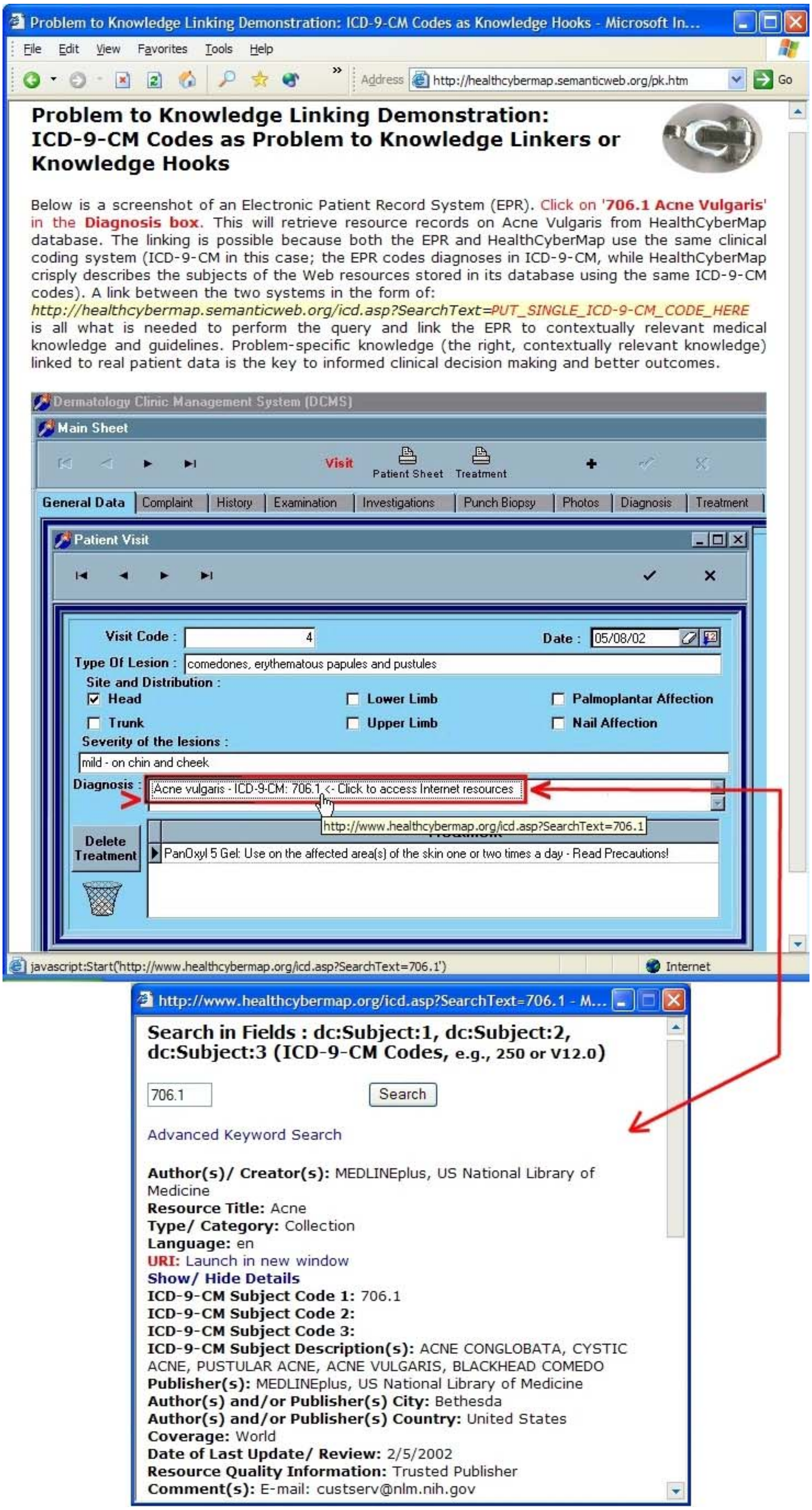


Figure 10.2. Screenshot of HealthCyberMap's Problem to Knowledge Linking Demonstrator (<http://healthcybermap.semanticweb.org/pk.htm>).

Only resources with a DC subject field containing the argument are retrieved. HealthCyberMap's solution is flexible and dynamic. We can keep adding/ deleting resources to HealthCyberMap and have all new changes instantly reflected on our output without modifying HealthCyberMap architecture or code (or the client's calling code). The same query will always retrieve the latest updated data from the underlying database.

The Web demonstrator described in this chapter shows the feasibility of EPR to HealthCyberMap problem to knowledge linking using ICD codes (in DC subject fields) as crisp problem to knowledge linkers or hooks. Though ICD-9-CM codes are mainly used for EPR billing (US), the author has successfully demonstrated another clever use of these codes using mainstream technology that most interested groups can easily adopt. It is noteworthy that a prototype Emergency Room EPR, InfoMED ER (<http://www.infomed-epr.com>), developed in 2002 by Dr. R.G. Rochin at City University, London, is currently using HealthCyberMap pilot PKL service to provide contextual resource links within patient records.

By minimising irrelevant leads (noise) and reducing the time needed to find relevant information (the right contextually relevant knowledge is linked to real patient data in the EPR), the system is potentially beneficial. Empirical evidence also shows that well-informed physicians and patients are able to make better clinical decisions that positively affect healthcare outcomes [89]. The analysis of HealthCyberMap formative evaluation questionnaire results showed that respondents, especially clinicians, are welcoming the idea of PKL and finding it potentially useful (Chapter 12).

10.3 Conclusion

Many information needs arise during everyday clinical practice. Problem to knowledge linking aims at answering these needs by providing contextually appropriate medical knowledge in the right place and at the right time to help physicians and patients make better clinical decisions that positively affect healthcare outcomes. In this chapter, we have reported on the design and development of a re-usable and flexible Semantic Web problem to knowledge linking service. The service makes use of metadata and clinical codes to contextually link real patient data and problems in disparate EPR clients to relevant resources in an online medical knowledge service (HealthCyberMap). Clinical codes act as crisp knowledge hooks, providing a reliable common backbone language for communication between EPRs and HealthCyberMap. By minimising irrelevant leads and reducing the time needed to find relevant information, the service is potentially beneficial.

Part IV: Formative Evaluation of HealthCyberMap Pilot Service

11 HealthCyberMap Formative Evaluation Methodology

11.1 Introduction

Standard good practice dictates that evaluation plans be developed at the same time as designing a service. This will naturally be an iterative process. Ideas about the object of the service and its shape gradually evolve. At the same time, the expectations of the service emerge, allowing the evaluation requirements to gradually unfold. As it becomes clear what the service is expected to do (and not do), so our notions of relevant questions to be asked in determining whether the service meets its expectations, also become apparent [90, 91].

The author believes that evaluation should run throughout the whole lifetime of any service (not just once or for a limited time) to make sure the service is always delivering what it has promised, and to overcome designers' blindness (deficiencies overlooked by designers and only seen by users).

Crawford [91] provides a short practical guide on the evaluation of library and information services and tips on developing and administering evaluation questionnaires. He sees evaluation as an internal control mechanism that ensures resources dedicated to the evaluated service are used efficiently and effectively to the best interests of users. Evaluation can help justifying/ defending a service plan and planning for future improvements. It can also help identifying the extent to which identified problems, if any, can be solved. Usually problems cannot be solved either due to resource constraints or due to the involvement of parties outside the project's control. Differing or contradictory needs of different user categories, e.g., healthcare professionals and laypersons, might be also highlighted during evaluation (it is wrong to assume all users have similar needs). According to Crawford, evaluation ensures users are involved in the project's development. Ideally, the outcome of any survey work and resulting decisions should be reported back to users (closing the feedback loop) showing them that evaluation is a valid exercise to which they have contributed.

Depending on the development stage of the service being evaluated, two types of evaluation could be described [92]:

- Formative evaluation (during service development—evaluating concepts in their infancy); and

- Summative evaluation (at the end of a project—evaluating more complete/ fully developed services).

The current HealthCyberMap Pilot Service cannot be considered fully developed yet. It offers rather limited Web interface functionality and topical coverage (1640 resources in the current metadata base), and as such it would be impractical to conduct a summative evaluation of the service impact on clinical practice and health outcomes, or the cost savings that might be associated with its use. However, one could still evaluate the concepts behind the different pilot service interfaces to inform and guide any further development of the service (or similar services), e.g., to take reliable decisions about discarding or modifying a certain interface, or refining and expanding the service metadata set and query arguments. Consequently, it was decided to conduct a small-scale public formative evaluation of HealthCyberMap Pilot Service to learn, among other things, how useful users think are the ideas behind, and intended purposes of the different service interfaces.

11.2 Methods

Different evaluation methods have different strengths and weaknesses, and only a good multi-method approach to Web service evaluation can help putting together an adequate, more complete picture of how a Web service is being used and received by users [93, 94]. Evaluation methods include questionnaires, semi-structured interviews, video tracking of a sample of users in a usability lab, analysis of Web server transaction logs, and expert panel critique using a standard checklist of usability heuristics [90, 91, 93, 94]. The choice of the specific methodologies used for an evaluation depends on the resources (human, financial, time, etc.) available for the evaluation, among other things [94].

Log files are automatically generated by Web servers and can offer valuable insight into Web site usage in natural working conditions, compared to the artificial setting of a usability lab (and without the high costs of the latter). They represent the activity of many users, over a potentially long period of time, compared to a limited number of usability lab users for an hour or two each [95], or a small number of respondents to an evaluation survey. Some Web site usage details derived from server log analysis, e.g., information about top referring search engines/ sites or the geographical provenance of site visitors, might be more accurate than similar information obtained from a user questionnaire, because of the much larger number of users covered in server logs.

However, log files also have their limitations (see Chapter 13: Analysis Results of HealthCyberMap Server Log for the Formative Evaluation Study Period). Standard log files contain very incomplete information on the user's goal in visiting the site. Besides, they cannot tell what users have been doing within a page (e.g., did they visit any of the suggested external links on that page). They just track users' paths from one page to another, and even this tracking can be incomplete because of the effects of browser/ proxy caching [95]. With only log files, one cannot tell whether users were satisfied and found what they were looking for or not. Thousands of users might have accessed a particular page or map and then found it useless; logs will not report user-perceived quality, relevance or usefulness of accessed pages. User questionnaires can complement and patch these and other deficiencies of server logs.

A good example of combining users' questionnaires with server logs statistics in the context of evaluating a digital library service is the work carried at the Universities of Wales Aberystwyth and Sheffield to evaluate the UK National electronic Library for Health (NeLH) Pilot Site (<http://www.nelh.nhs.uk/>) [96]. It should be noted that the current NeLH Pilot Site does not use any navigational cybermaps (it is only text based).

For HealthCyberMap formative evaluation, we used an online user questionnaire, in addition to server log analysis.

An evaluation methodology specific to HealthCyberMap Semantic Subject Search Engine based on relevance metrics is described in detail at the end of this chapter, but it was decided that it would be better to postpone this specialised search engine evaluation until after the implementation of SNOMED-CT (Systematised Nomenclature of Medicine—Clinical Terms) and a true terminology server in HealthCyberMap. (See Chapter 8: HealthCyberMap Medical Semantic Subject Search Engine, and Chapter 15, Sections 15.3.1 and 15.3.2.)

11.2.1 HealthCyberMap Formative Evaluation Questionnaire

The author developed a formative evaluation questionnaire for HealthCyberMap pilot implementation and published it as a public Web form to be filled online (<http://healthcybermap.semanticweb.org/questionnaire.asp>—see also Appendix 2:

Formative Evaluation Questionnaire of HealthCyberMap Pilot Service). Respondents' input is written to a database residing on HealthCyberMap server.

The questionnaire went publicly online on 18 April 2002. Participants were solicited in the following manner:

- announcements posted to relevant Internet newsgroups such as <sci.med.informatics> and <bit.listserv.medlib-l>;
- announcements posted to Yahoo! Groups mailing lists of SemanticWeb.org <http://groups.yahoo.com/group/semanticweb/>, BMiS (British Medical Informatics Society) <http://groups.yahoo.com/group/bmis-members/>, and NeLH VBL (National electronic Library for Health, Virtual Branch Libraries) <http://groups.yahoo.com/group/nelh-vbl/>;
- one-time e-mails targeting colleagues and researchers in the fields of medicine and medical informatics, and friends from outside these fields;
- the questionnaire was also advertised in a poster on HealthCyberMap presented by the author on 10 May 2002 (Open Day of the Centre for Measurement and Information in Medicine at City University); plus
- the questionnaire could be accessed by any visitor of HealthCyberMap Web site via a well-noticed link at the top of the front page (<http://healthcybermap.semanticweb.org/>).

11.2.1.1 Questions and Tasks

HealthCyberMap questionnaire development was guided by some existing relevant questionnaires, e.g., surveys described in NIH Web Site Evaluation and Performance Measures Toolkit[‡‡] [94] and the Hanford Decision Mapping System Evaluation Questionnaire[§§], in addition to Crawford [91] and Gillham [93] guidelines for designing questionnaires.

Respondents were told at the beginning of the questionnaire form that they are contributing to an initial evaluation of HealthCyberMap concepts in their infancy rather than to an evaluation of a full-blown service, so that they may adjust their

---

‡‡ See: <http://irm.cit.nih.gov/itmra/weptest/app_a8.htm> and <http://irm.cit.nih.gov/itmra/weptest/app_a5.htm>

§§ See: <https://catalyst.washington.edu/webtools/webq/survey.cgi?user=cdrew&survey=6>

responses accordingly. For example, users should not expect a complete coverage of all clinical and health topics at this stage of HealthCyberMap development.

As with the Hanford Decision Mapping System Evaluation Questionnaire, a short description of HealthCyberMap main aims and a further information link (<http://healthcybermap.semanticweb.org/#underhood>) were also provided at the beginning of the questionnaire form to help respondents (who are not yet familiar with HealthCyberMap) get oriented quickly to the different HealthCyberMap interfaces. The orientation link leads to a section of HealthCyberMap Web site called "Under the Hood" offering a downloadable HealthCyberMap Quick Tour in Microsoft PowerPoint format, plus several links to relevant background information about the service (for interested users).

Also following the Hanford Questionnaire model, respondents were told that HealthCyberMap questionnaire is anonymous (a link to HealthCyberMap Privacy Statement was also provided—<http://healthcybermap.semanticweb.org/privacy.htm>), and that they may leave blank any question they do not want to answer. This gives respondents more flexibility and freedom and could increase the number of respondents.

Respondents were advised at the start of the questionnaire that it should take around 30 minutes to complete the questionnaire (this figure includes the time needed to get oriented to the different HealthCyberMap interfaces, and to complete the tasks associated with some of the questions—see below).

Throughout the questionnaire form, direct links were provided to the different service interfaces whenever these were mentioned in the questions (Figure 11.1). The aim was to help new users who might be still confused about the names and functions of the different interfaces quickly jump to and check these interfaces in order to answer the questions. All links on the questionnaire form open in a separate window without closing the questionnaire window, so that any already selected/ typed answers are not lost.

The questionnaire is composed of 41 questions, including 35 multiple-choice (closed) questions and 6 open questions (including one routing open question, question 20, that depends on user's answer to the question before it). The open questions allow respondents to express and explain their opinions more freely when closed questions with prescribed answers seem too restrictive [93].

11.2.1.1.1 Subject Descriptors

Questions 1-15 cover subject descriptors/ profiles, e.g., gender and age, education, role (member of general public, general practitioner, etc.), computer and Internet experience, type of operating system, Web browser and Internet connection, and attitude towards the Internet as a credible source of health information. These descriptors help us gain a general idea about respondents, assess how representative they are of the wider Internet/ HealthCyberMap audience, and interpret responses to the specific evaluation questions that follow about HealthCyberMap usability and usefulness (questions 16-41).

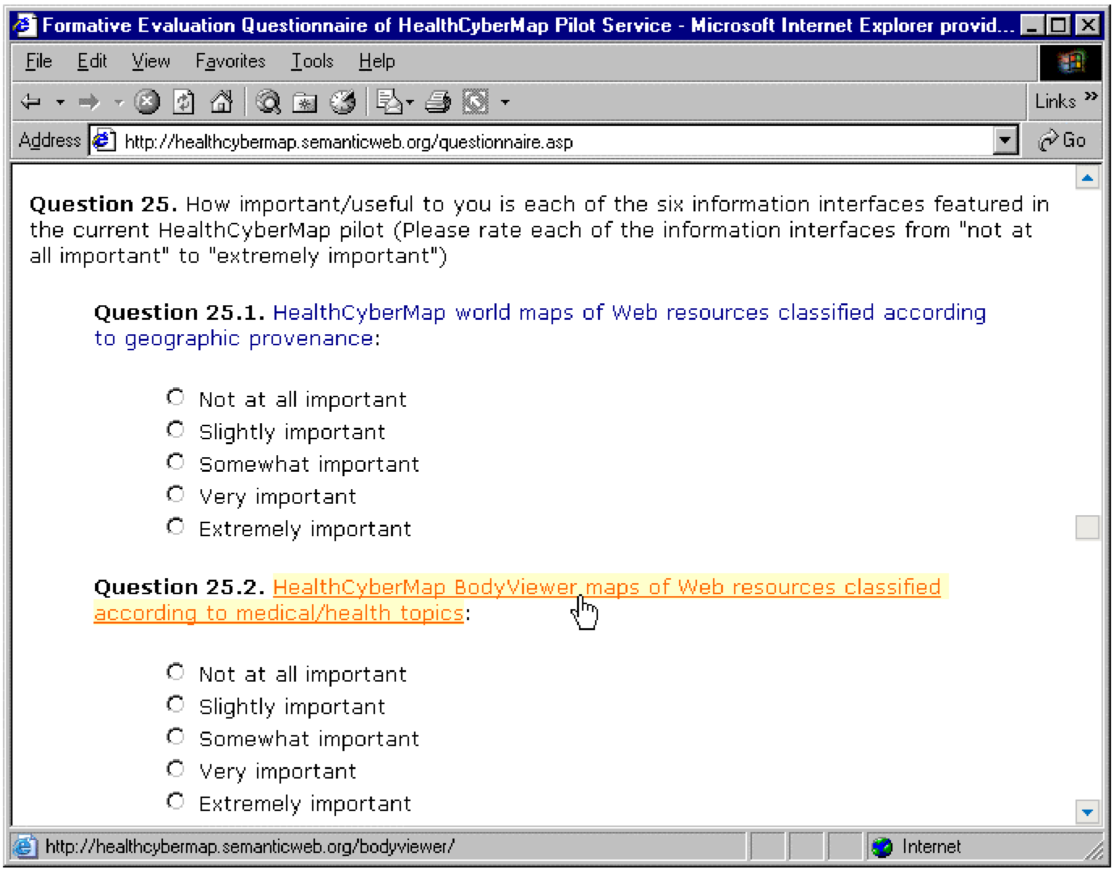


Figure 11.1. Direct links that open in a separate window were provided to the different service interfaces whenever these were mentioned in questions.

11.2.1.1.2 Usability, Usefulness and User Satisfaction Issues

Testing usability and user acceptance is a critical part of any Web-based information service [97]. However, usability alone does not imply usefulness. Not every usable service is useful. Usability is related to the degree of difficulty understanding and using an information service, while usefulness depends on actual user's needs. Useful features enable users to "do things" they want or need by providing the tools necessary to perform specific tasks. In contrast, usable features make "doing" easy

[98]. In extreme cases, low usability might obscure usefulness, if users cannot understand what the service at hand is about.

In HealthCyberMap questionnaire, respondents were asked to rate on a five-step scale how well or usable (question 22.1-22.6) and how useful (question 25.1-25.6) they found the six HealthCyberMap pilot interfaces.[***] They were also asked to rate on a similar five-step scale how useful they found the four proposed HealthCyberMap future interfaces/ possibilities (question 35.1-35.4)[†††]. These latter interfaces were not as developed as the six main interfaces in the evaluated pilot service, but they were adequately explained (online documentation) to help respondents assess their usefulness.

The author also included other usability and usefulness questions, guided by Nielsen's usability criteria [99] and the evaluation criteria provided by Batagelj et al [100]. Issues covered by these questions include:

- whether HealthCyberMap approach and interfaces, including the use of visual maps as a navigational aid, meet users' information retrieval and navigation needs, and avoid detail overload;
- whether the meanings of hotspots/ icons on HealthCyberMap maps are instantly recognisable (i.e., familiar to users) or do they need to be learned and recalled (questions 23 and 24);
- whether online help is needed at all, and if needed is it adequate (question 26);
- the speed at which HealthCyberMap maps load since this can also affect usability and user acceptance (question 27—to be interpreted in light of the answers to the earlier question on the type of user Internet connection); and
- resource relevance (question 30) and quality (question 31) issues.

[***] The six HealthCyberMap pilot interfaces are the world maps of resources by geographic provenance, BodyViewer maps of resources by medical/ health topics, the textual resource index using ICD-9-CM top-level categories of medical/ health topics, map of resources by type (e.g., audio-visual material, interactive resources, digital atlases, fact sheets, etc.), resources by language, and HealthCyberMap Semantic Subject Search Engine (see Chapters 8 and 9).

[†††] The four proposed HealthCyberMap future interfaces/ possibilities are the multi-axial classification of resources based on two or more Dublin Core metadata elements, customisation/ location-based customisation, problem to knowledge linking, and mapping health problems in HealthCyberMap and identifying corresponding information needs and gaps (see Chapters 10 and 14).

11.2.1.1.3 Usability Task

Usability evaluation should also include query scenarios (tasks) based on representative information seeking tasks and real-world data [48, 94, 100]. HealthCyberMap questionnaire includes a usability task (questions 28 and 29). Users are asked to find resources on some topic like “diabetes mellitus” using HealthCyberMap interfaces and to report whether they were successful in completing this task from the first attempt, successful after one or more failed attempts, or not successful at all (question 28), mentioning the interface(s) they found most helpful in in accomplishing this task (question 29). Effectiveness usability studies also need to care for different user profiles. For example, some users might not know in advance that resources on “diabetes mellitus” are classified under “endocrine disorders”, and, for this reason, might not be successful in completing this task from the first attempt (though the intuitive exploratory nature of the maps can help users quickly discover and learn new things and HealthCyberMap Semantic Subject Search Engine can be always used when a user cannot locate a topic using the other interfaces). Therefore, it should be very useful to know more about the background and characteristics of users (e.g., age, previous education, existing knowledge and experience, and browser/ device) while interpreting the results of this task, as these descriptors could affect user’s ability to perceive and/ or to comprehend a map or visual metaphor [38]; this background information is provided by questions 1-15.

It should be noted that monitoring a representative sample of users performing this kind of tasks in a properly-equiped usability lab (when resources permit) can offer valuable extra (quantitative) information. The answers to information seeking tasks can be recorded and checked to ensure that users were really successful. Information about the paths that users took to locate information (e.g., exact time taken to accomplish task, and number and sequence of pages accessed during that time) can be also recorded and can give us an accurate quantitative idea about the different strategies that users took, and also help us identify any areas of the service that were misleading to users.

11.2.1.1.4 Comparative Task (HealthCyberMap vs. Visual Net)

HealthCyberMap questionnaire also includes a comparative task (question 34). Users are asked to compare HealthCyberMap’s mapping approach to that of Visual Net PubMed interface (<http://pubmed.antarcti.ca/start>). The phrasing of question 34 and its six qualitative answer options is based on a similar question that appeared in the

Customer Satisfaction Survey (<http://irm.cit.nih.gov/itmra/weptest/app_a8.htm>) that is included in NIH Web Site Evaluation and Performance Measures Toolkit [94]. A quantitative approach to comparative task analysis is also possible, and is usually carried on in a usability lab [101].

The purpose of this task (question 34) is to learn how users perceive the difference between the two interfaces (HealthCyberMap vs. Visual Net), especially regarding map iconicity (high/ associative and pictorial in HealthCyberMap vs. low/ geometric in Visual Net—Figures 11.2 and 11.3) [102]. Based on Nielsen's well-known usability criterion of "recognition not recall" [99], associative and pictorial icons that enable instant recognition/ comprehension should be preferred to geometric ones.

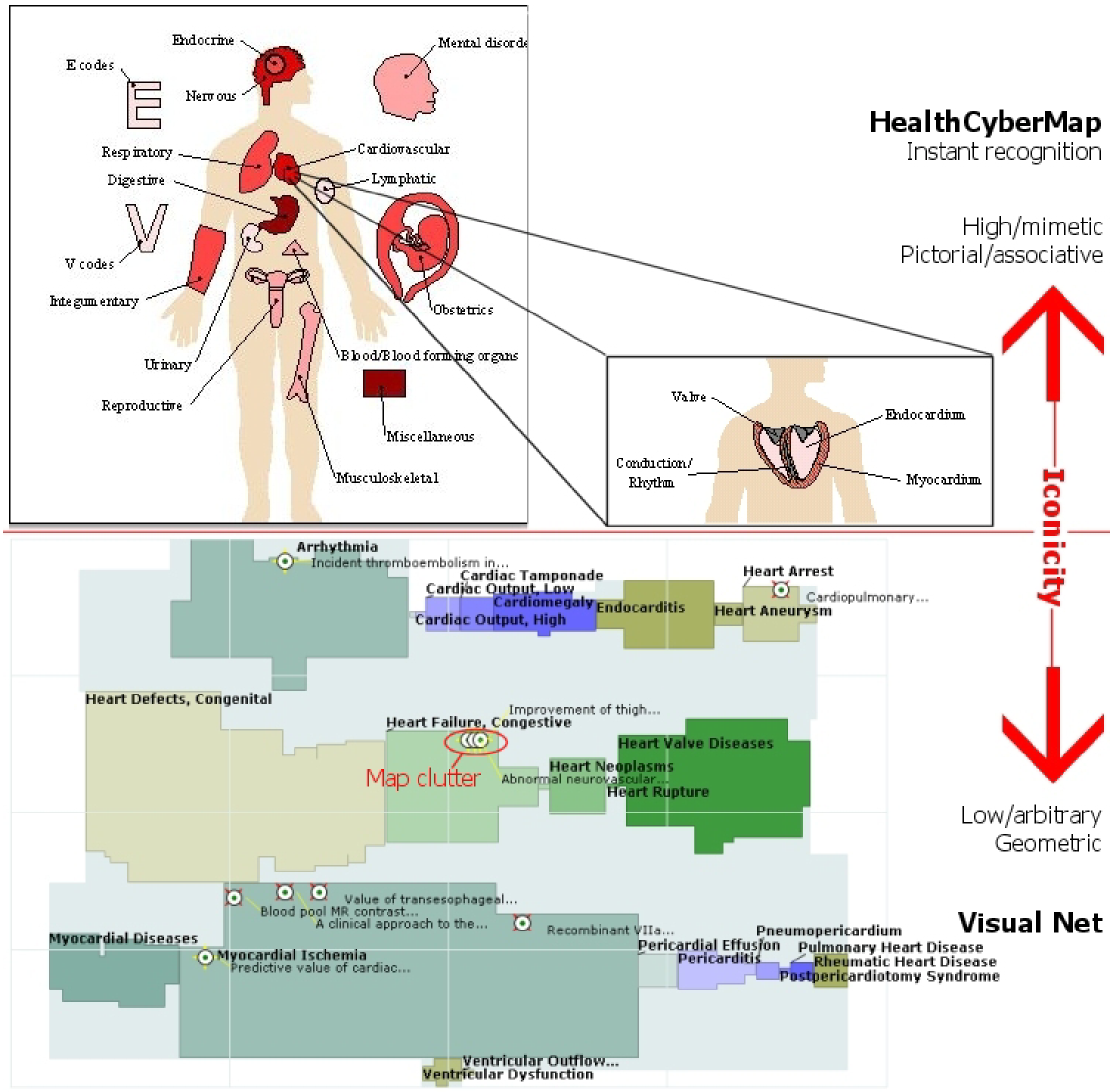


Figure 11.2. Screenshot of parts of HealthCyberMap and Visual Net navigational maps for "heart diseases". Notice the difference in map iconicity between HealthCyberMap and Visual Net approaches, and the map clutter resulting from Visual Net's way of representing each resource directly on the map using a distinct point symbol.

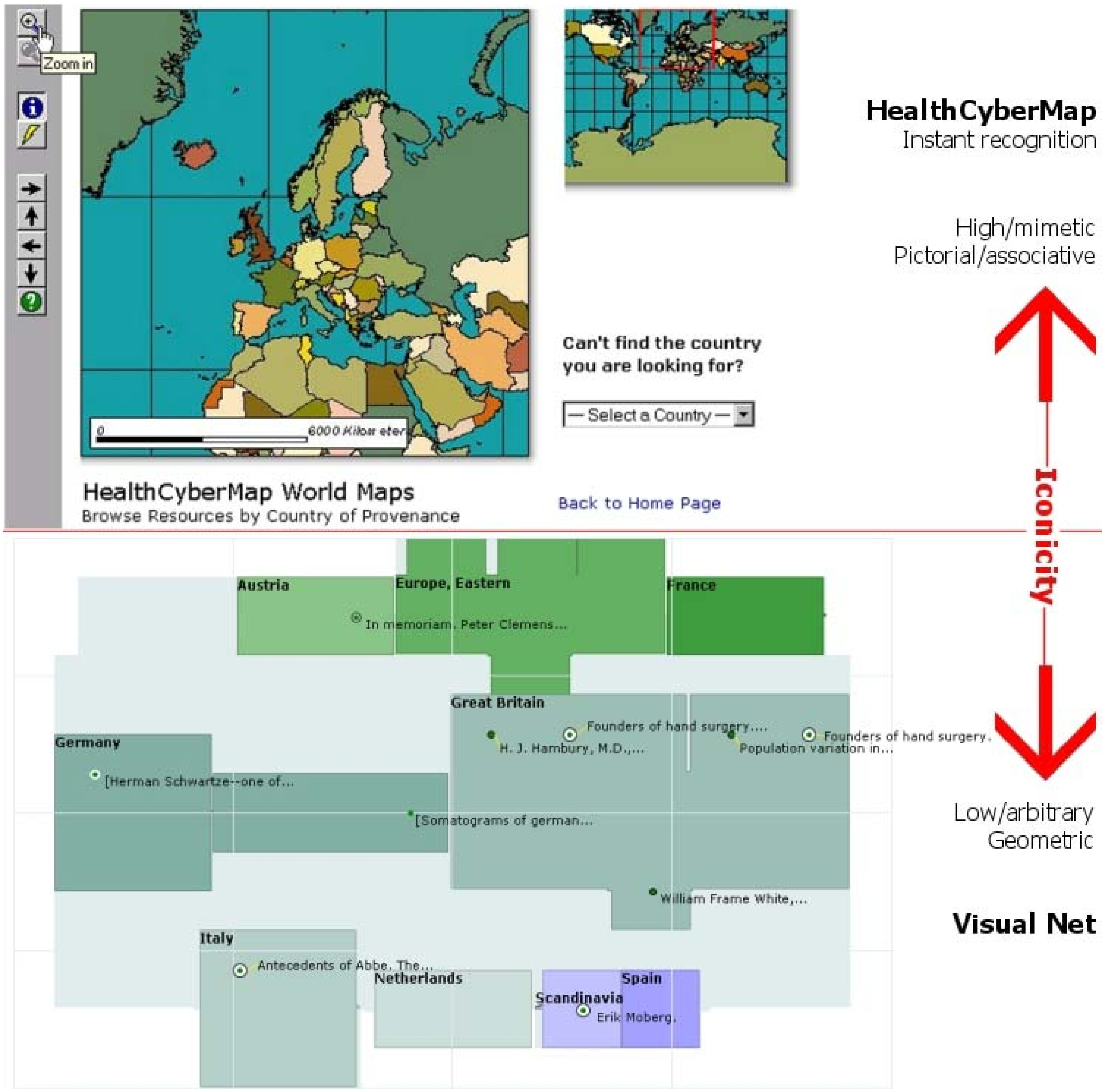


Figure 11.3. Screenshot of parts of HealthCyberMap and Visual Net navigational maps for "geographical places (countries of Europe)". Notice the difference in map iconicity between HealthCyberMap and Visual Net approaches, and how "Great Britain" is oddly mapped (and shaped) east of "Germany" on Visual Net map.

11.2.1.1.5 Other Issues

The intended audience for HealthCyberMap includes healthcare professionals and librarians, and patients and the public in general. However, HealthCyberMap pilot service does not organise information resources by their intended primary audience, following the growing trend of treating medical knowledge as a single blob (or pool of knowledge) that is relevant to both professionals and laypersons, and so should be made accessible to all groups without any distinction. Supporters of this trend think that patients should be empowered and given more information and control of their conditions [103]. Indeed, laypersons sometimes show more knowledge and understanding of their own conditions than their treating doctors do. Nevertheless, a question was included in HealthCyberMap evaluation questionnaire to collect users' opinions on the utility of organising information resources by their intended primary audience in a future implementation of the service (question 36). (A metadata field could be easily added to record the intended primary audience of a resource.)

There are also questions on service access frequency (visitor stickiness—question 16) and how users heard about HealthCyberMap, which could help determining the most successful service advertising method (question 17).

11.2.1.2 Analysis of Results

The analysis of answers of 35 respondents who filled in and submitted the questionnaire during the period between 18 April 2002 and 1 June 2002 is presented in Chapter 13: Analysis Results of HealthCyberMap Formative Evaluation Questionnaire. Crawford [91] suggests a minimum acceptable sample size of 30 questionnaire respondents, so our results based on the responses of 35 subjects should be adequate.

The author used Microsoft Excel 97 to produce the results and graphs presented in Chapter 12, guided by very helpful ideas suggested in Gillham [93], like the scaled response bar graphs used to present responses to questions 22, 25 and 35.

11.2.2 HealthCyberMap Server Log Analysis

HealthCyberMap server logs were analysed for the period from 18 April 2002 to 1 June 2002 (45 days). HealthCyberMap site structure and content was kept unchanged during this 45-day period to ensure consistent server log analysis results. As mentioned before, this is the same period during which HealthCyberMap online formative evaluation has been running. HealthCyberMap online formative evaluation questionnaire was launched and first advertised on the 18th of April 2002, and by the 1st of June 2002, 35 respondents had filled in and submitted the online questionnaire form.

Some of the information obtained using server log analysis is also collected in HealthCyberMap questionnaire, e.g., repeat visitor (retention) statistics (cf. question 16), referring search engines and sites (cf. question 17), and types of Web browsers and operating systems accessing HealthCyberMap (cf. questions 11, 12 and 14). Information coming from server log analysis is usually more quantitative (and potentially more accurate because of the much larger number of users analysed) compared to similar information, usually qualitative, obtained using questionnaires. In areas of overlap, HealthCyberMap server log and questionnaire analyses could be considered complementary, and used to crosscheck each other (e.g., to verify that the sample of respondents who filled-in the questionnaire is a good representative of all site users).

11.2.2.1 About Server Log Analysis

Server log analysis started as a way for server administrators to get a sense of actual server load and bandwidth use/ misuse (e.g., visitors abusing the site), and then use this information to ensure adequate bandwidth and server capacity on their Web sites to meet visitors' demands. Server load and bandwidth information remains today the most reliable type of information one can get from standard server log analysis. Server log files are also used these days to learn about the number of site visitors, visitor retention (measurement of unique users who return to the site over a given time), and Web site usability (did the site successfully meet visitors' specific information-seeking goals). However, the information learned from standard server log analysis on these issues is much less reliable (for a variety of reasons explained in Chapter 13) and cannot be used *alone* to make strong inferences about site success or usability [95, 104].

McDunn [105] provides a primer on Web server log file structure and analysis. Server access log files record Web server activity, providing details about file requests to the server and response of the latter to those requests, including any error codes. In a server access log file, each line describes the source of a request, which might point to the actual geographic location of the visitor, the date and time of the request, the file requested (e.g., HTML page, GIF or JPEG image, etc.), and length of the transferred file. The line also records other data including transaction status code (Success—200 series, Redirect—300 series, Failure—400 series, Server Error—500 series), the identity of referring pages, and the browser and operating system used to send the request [95, 105]. Referring pages might be other pages on the same site or external pages (e.g., an external Web search engine results page).

Following is one line from HealthCyberMap server log files:

Source of Request (Host) | Date and Time of Request | Page Requested (HTTP protocol) | Status Code | Number of Bytes | Referring Page | Browser | Platform

```
s8n38.hfx.eastlink.ca - - [21/Apr/2002:09:47:01 +0500] "GET /bodyviewer/Default.htm
HTTP/1.1" 200 2126 "http://healthcybermap.semanticweb.org" "Mozilla/4.0 (compatible;
MSIE 6.0; Windows 98)"
```

Requesting a single Web page can result in multiple hits to the server. The HTML page will show as a hit in the log file, and each graphic on that page will also count as a hit in the log. So an HTML page with three graphics will be recorded as four individual hits in the log. This set of four hits is described as a page view, all the hits necessary to display a single Web page (Figure 11.4) [105]. (A page using frames counts as multiple page views.)

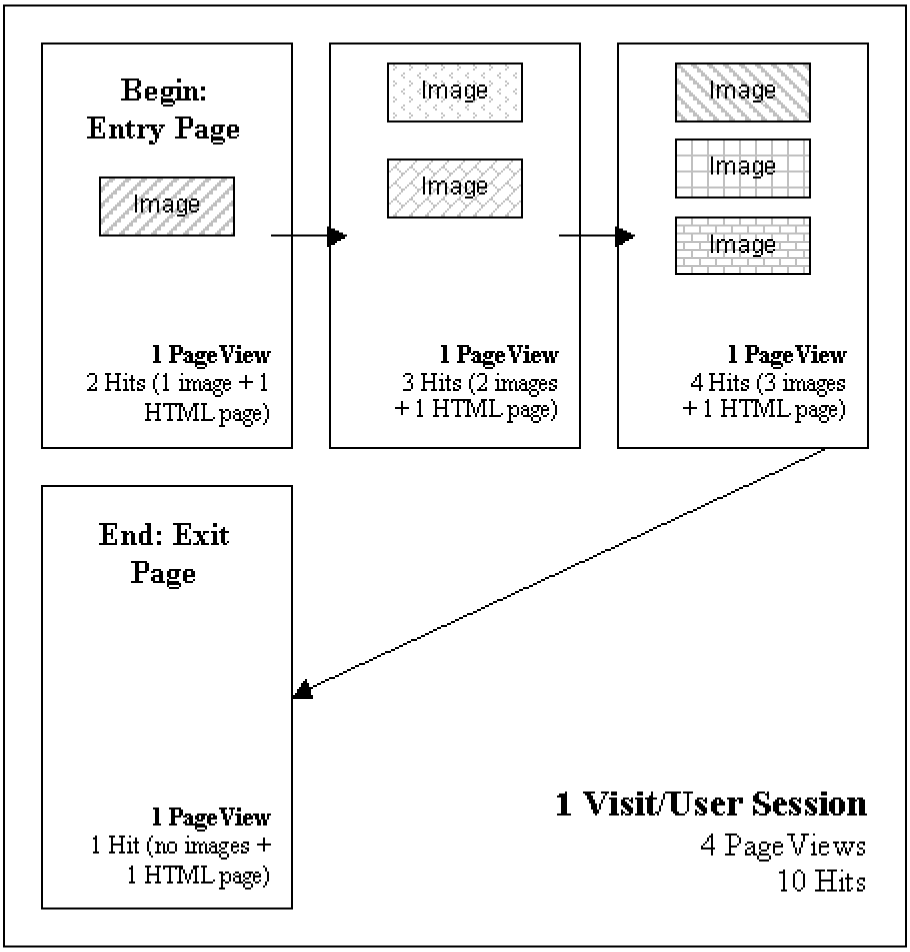


Figure 11.4. The relation between hits, page views and visits (sessions).

11.2.2.2 Tool Used (Sawmill 6.3.8) and Settings

The author reviewed several server log analysis tools running under Windows, including NetTracker 5.5 Professional from Sane Solutions (<http://www.sane.com/>), 123LogAnalyzer 2.10 from ZY Computing

(<http://www.123LogAnalyzer.com/>), and Sawmill 6.3.8 server log analysis tool from Flowerfire (<http://www.sawmill.net/>). No two tools among those tested produced exactly the same results (using the same HealthCyberMap server log files downloaded from HealthCyberMap server and covering the 45-day period of this study). For example, they reported different numbers of unique visitors for the same 45-day period of the analysis (though some of these numbers were close to each other). Ferguson [106] describes a similar situation. This is due to the different methods used by these tools to compute their statistics (e.g., some tools ignore the last number(s) of an IP address), combined with the inherent limitations of standard log files (see Chapter 13).

The author first excluded those tools producing very different results compared to the rest of the tools that were tested, thus narrowing the number of possible choices. The final choice was based on several criteria like ease of use, configurability, and quality and details of generated reports and graphs, and was also guided by Bannan's excellent review of many of these tools [107].

Sawmill 6.3.8 was ultimately selected to perform the analysis of 26,365 individual entries (lines) in HealthCyberMap server log files corresponding to the 45-day period of this study. Maximum session duration was set to 7200 seconds (2 hours); this is Sawmill's default value (Figure 11.5). This controls the maximum length of a session (visit) in compiling session information. This option is useful because some large ISPs (Internet Service Providers), e.g., AOL, and other large companies/ academic bodies use Web caches/ proxies that make all hits from their clients appear to be coming from one or just a few computers. When many people are using these caches at the same time, this can result in the intermixing of several true sessions in a single apparent session, resulting in incorrect session information. By discarding long sessions, which are probably the result of these caches, this problem is reduced (but not completely fixed).

Session timeout controls the amount of time a session can be idle before it is considered complete. Sessions are considered ended when hits from a visitor have not hit the site in the number of second specified under this setting. For HealthCyberMap server log analysis, session timeout was set to 1800 seconds (30 minutes); again this is Sawmill's default setting (Figure 11.5). This means that if a visitor did not hit HealthCyberMap site for half an hour, the previous hits are considered a completed single session, and any subsequent hits are considered a new session.

The results of HealthCyberMap server log analysis are presented and discussed in Chapter 13: Analysis Results of HealthCyberMap Server Log for the Formative Evaluation Study Period.

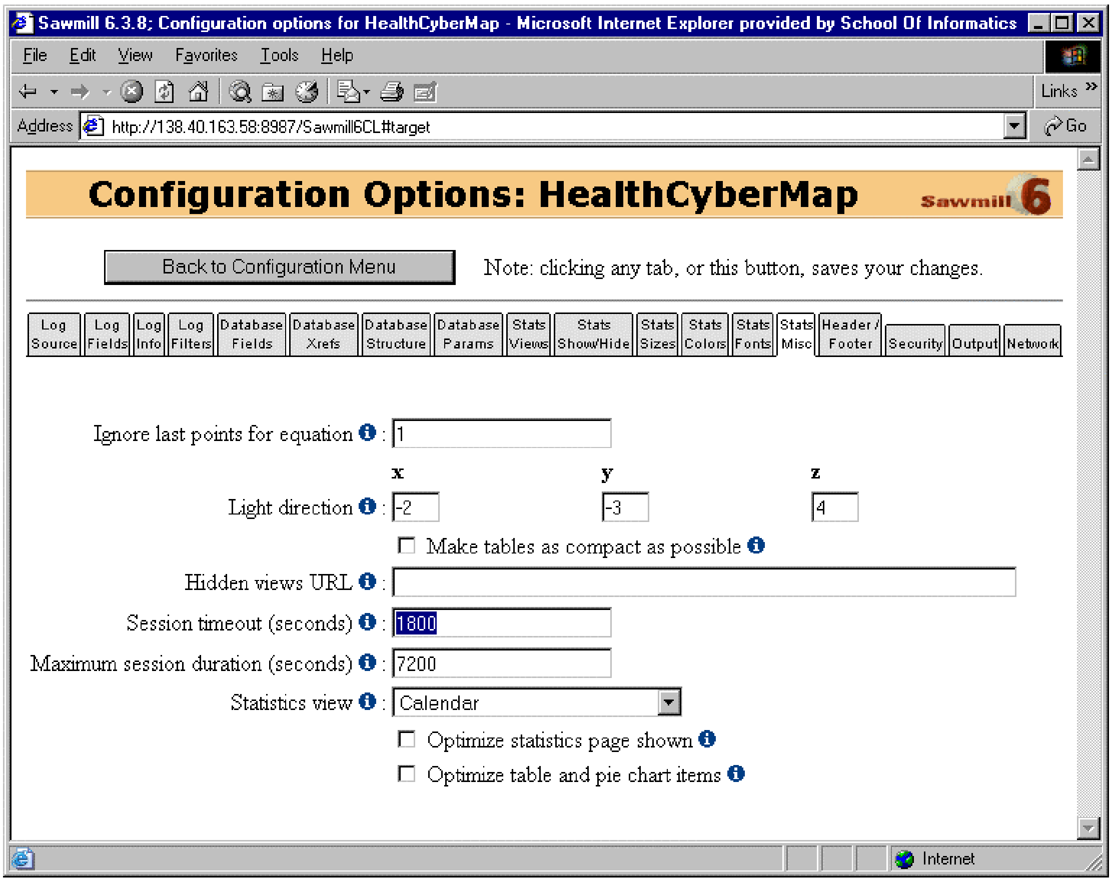


Figure 11.5. Setting maximum session duration and session timeout in Sawmill 6.3.8.

11.2.3 A Proposed Evaluation Methodology for HealthCyberMap Medical Semantic Subject Search Engine

11.2.3.1 Relevance Metrics

The concept of "relevance" is a fundamental concept of information retrieval [108]. Search engine/ information retrieval evaluation based on relevance metrics, namely recall and precision, is widely mentioned in the literature [109, 110, 111].

The measurement of precision requires a qualified individual, or group of individuals, to inspect the output from a search and to sort the output into two groups of resources—relevant and not relevant (precision = number of relevant resources in query result/total number of resources, relevant + irrelevant, in query result). The measurement of recall requires that the individual or group of individuals also have access to the complete set of resources that was searched (recall = number of relevant resources in query result/number of relevant resources in the queried resource pool) [109].

The classic Cranfield-model evaluation requires three (well-defined) things: a collection of resources, a set of queries, and a set of relevance judgements for those queries and resources [112].

Comparative search-engine evaluation studies are well known in the literature [109, 113]. A comparative evaluation study is very helpful in comparing a new search engine/ methodology (e.g., manual topic indexing and concept-based retrieval as in HealthCyberMap) to a more traditional one (e.g., free, full text search of spider-indexed resources). Besides precision (noise) and recall (silence) rates, search engines can be compared regarding other aspects like speed, user interface, processing of user input (e.g., spelling correction, term completion and stemming), the way search results are presented and ranked, and, in case of engines querying distinct resource pools, currency and coverage.

11.2.3.2 A Comparative Evaluation of HealthCyberMap Semantic Search Engine Based on Relevance Metrics

The author proposes comparing results of a defined set of searches performed on HealthCyberMap's resource pool (or a defined representative subset of it) using HealthCyberMap's method with results of same queries obtained using a standard free, full text search engine/ spider (searching the actual text of same resources) regarding silence (missed relevant hits) and noise (irrelevant hits).

As almost 70% of the Web is non-textual [114], a good representative resource pool for this kind of studies must include some multimedia/ interactive resources like MEDLINE*plus* patient education tutorials in Macromedia Flash (also manually indexed in the current HealthCyberMap metadata base, e.g., <http://www.nlm.nih.gov/medlineplus/tutorials/ulcerativecolitisloader.html>). (Full text spiders usually fail picking up the topics of this latter type of resources if these topics are not mentioned in the loading HTML document.)

Queries should be repeated with different synonyms (e.g., "renal failure" and "kidney failure" or "myocardial infarction" and "coronary thrombosis"). The best engine will be the one that returns exactly the same set of results with all equivalent synonyms.

11.4 Conclusion

HealthCyberMap formative evaluation aims at evaluating the concepts behind the current pilot service and learning how it is being used and received by users. The ultimate goal is to inform and guide any further development of the service (or similar services). The author adopted a two-method evaluation approach using an online user evaluation questionnaire, in addition to analysis of HealthCyberMap server transaction log. The questionnaire covers issues related to service usability, usefulness, and user satisfaction. It includes a usability task and a comparative task, comparing HealthCyberMap to another approach (Visual Net). Both tasks are assessed qualitatively. Sawmill 6.3.8 server log analysis tool was selected from among several other tools to perform the analysis of HealthCyberMap server transaction log. HealthCyberMap server log analysis can offer valuable quantitative information on server activity and traffic, the geographic provenance of visitors, and referring search engines and sites, among other things. The two chosen evaluation methods (questionnaire and server log) act in a complementary synergistic way to provide a more complete picture about the evaluated pilot service.

A comparative evaluation study of HealthCyberMap Semantic Subject Search Engine based on relevance metrics was also proposed in this chapter, but it was decided that it would be better to postpone this specialised search engine evaluation until after the implementation of a more comprehensive domain ontology like SNOMED-CT and a true terminology server in HealthCyberMap.

12 Analysis Results of HealthCyberMap Formative Evaluation Questionnaire

12.1 Subject Descriptors

12.1.1 Age and Gender

During the period between 18 April 2002 and 1 June 2002, 35 respondents filled in and submitted an anonymous online HealthCyberMap formative evaluation questionnaire. Of those subjects, 57.14% (20 respondents) were male and 42.86% (15 respondents) were female. Figure 12.1 shows the age and gender distribution of respondents.

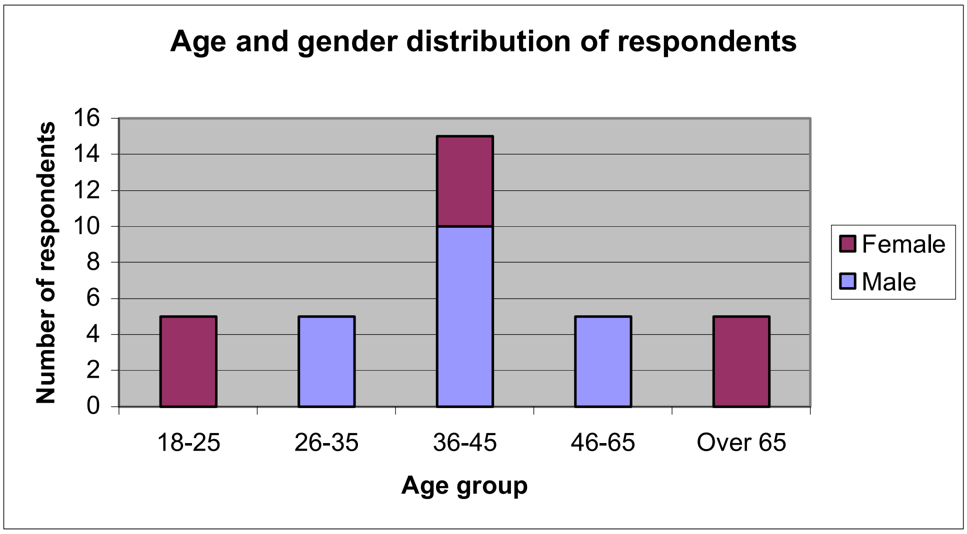


Figure 12.1. Age and gender distribution of respondents.

12.1.2 Highest Level of Education and Role

Figure 12.2 shows the distribution of the highest level of education achieved by respondents and gender.

Five subjects (14.29%) described their role as “Member of the general public seeking health-related information”. Another five respondents (14.29%) identified themselves as “General practitioner”. Fifteen respondents (42.86%) selected “Scientist/ researcher” to describe their role. A further five respondents (14.29%) identified themselves as “Medical librarian”, and only five respondents (14.29%) chose “Other” as the option that best describes their role (Figures 12.3 and 12.4). There were no “Member of the general public touched by disease”, “Hospital doctor”, “Nurse”, or

"AHP (Allied Health Professional)/ other healthcare professional" among respondents.

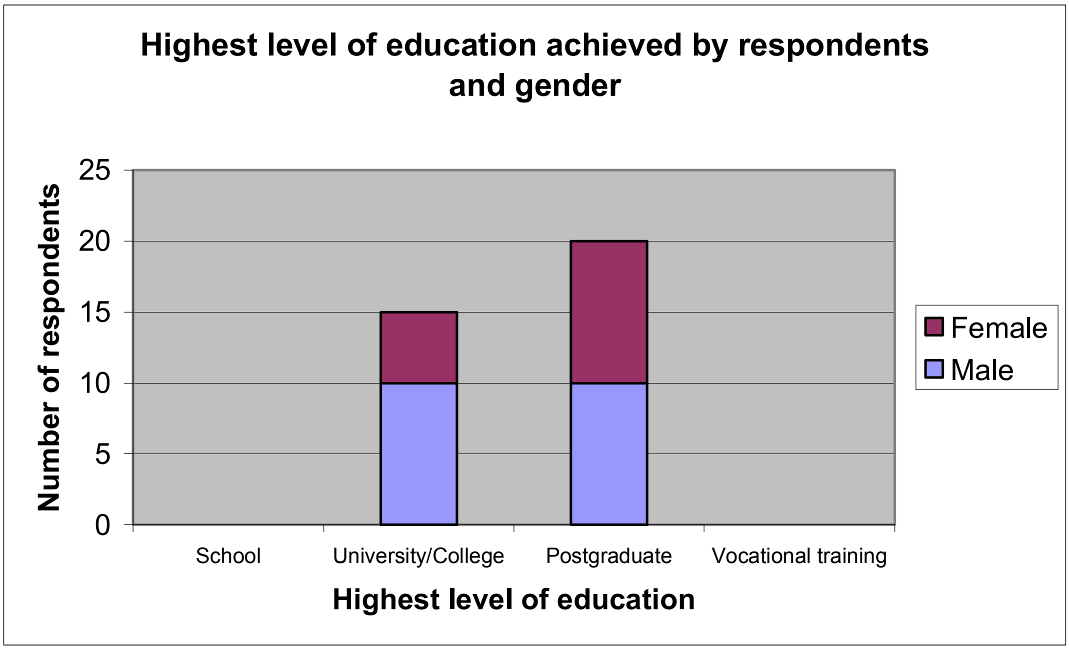


Figure 12.2. Distribution of the highest level of education achieved by respondents and gender.

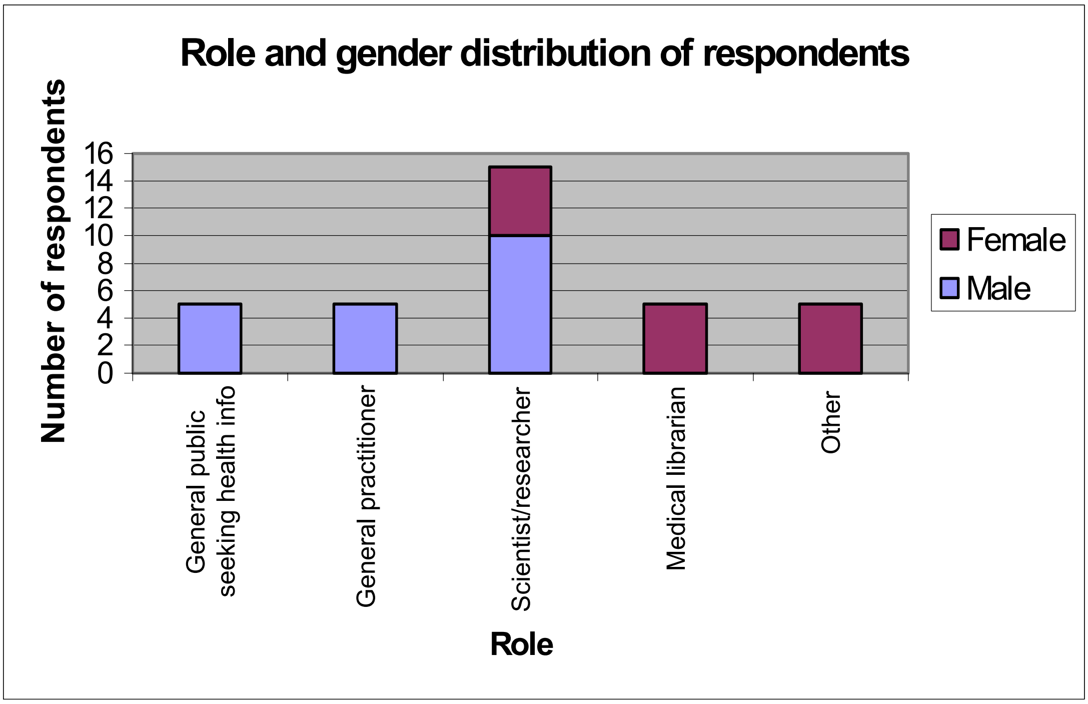


Figure 12.3. Role and gender distribution of respondents.

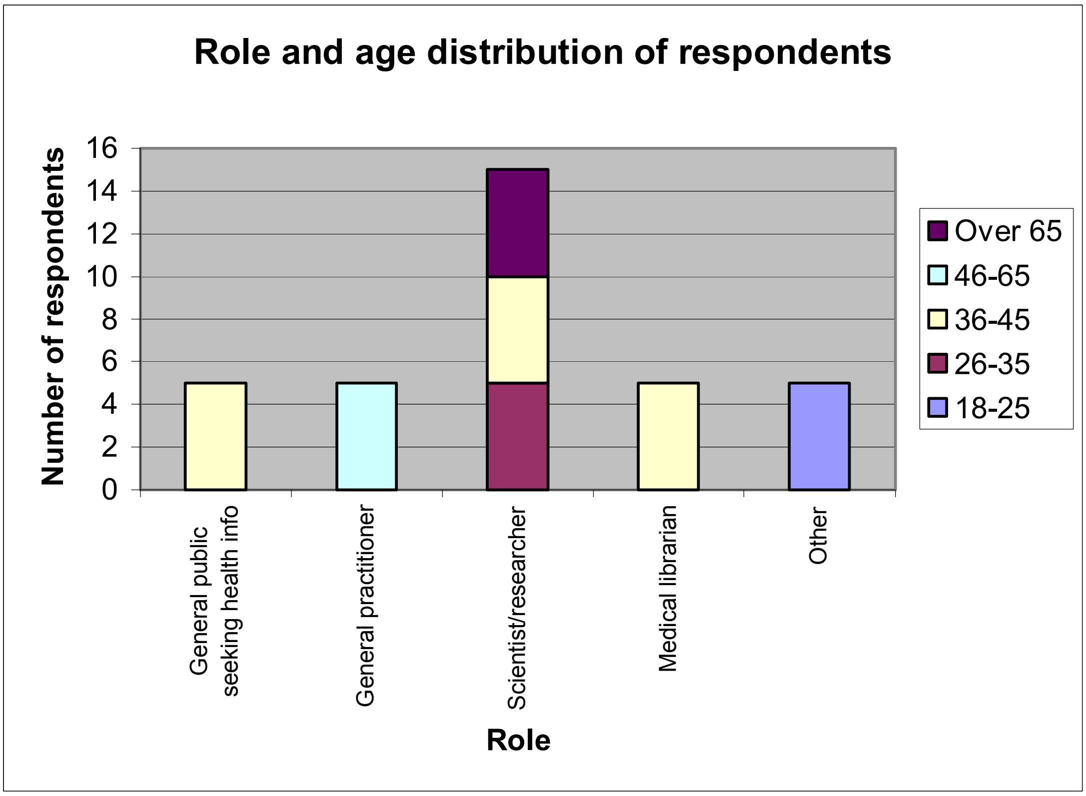


Figure 12.4. Role and age distribution of respondents.

12.1.3 Computer and Internet Use

Thirty respondents (85.71%) spend more than 5 hours per week (including work and home) using the Internet, while only five respondents (14.29%) spend less than 5 hours per week.

Thirty respondents (85.71%) have regularly used a computer and an Internet browser for more than 2 years, while five respondents (14.29%) have been doing so on a regular basis for only 2-6 months.

All respondents (35—100%) have a working personal computer at home. Fifteen subjects (42.86%) reported that they are most likely to use the Web from home, while 20 respondents (57.14%) use it most likely from work.

12.1.3.1 Type of Internet Connection

Fifteen subjects (42.86%) connect to the Internet using a modem, while 20 respondents (57.14%) have an ISDN (Integrated Services Digital Network), ADSL (Asymmetric Digital Subscriber Line), or broadband connection type (i.e., always connected—Figure 12.5).

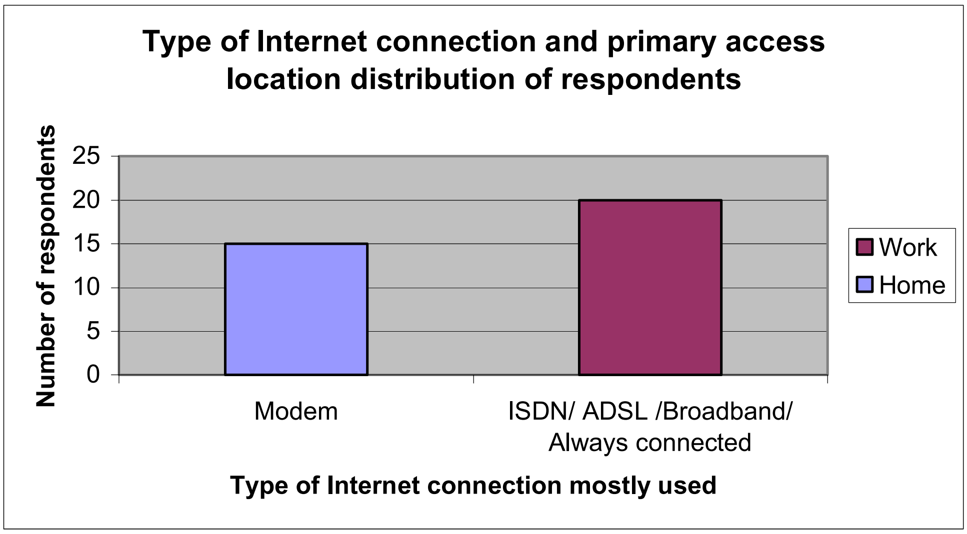


Figure 12.5. Type of Internet connection and primary access location distribution of respondents.

12.1.3.2 Web Browsers and Operating Systems

The most frequently used Web browser by respondents is Microsoft Internet Explorer (30 respondents—85.71%). Only five respondents (14.29%) are using Netscape Navigator or Communicator as their main browser (Figure 12.6). This is largely consistent with server log analysis findings (Chapter 13).

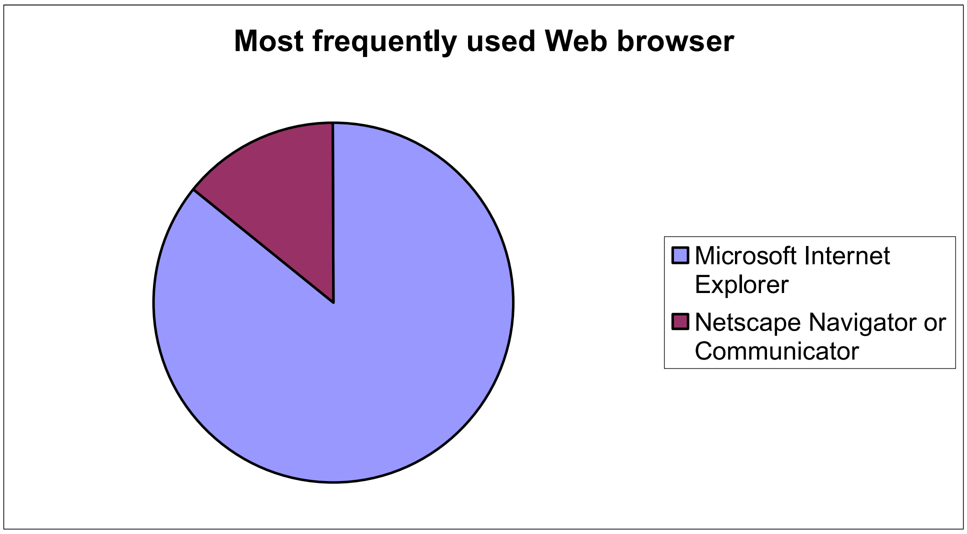


Figure 12.6. Most frequently used Web browser.

The big majority of respondents (25—71.42%) are using a version 6.x browser. Thirty respondents (85.71%) are currently running Microsoft Windows on their primary (most frequently used) PC, while only five respondents (14.29%) are running an Apple Macintosh operating system (Figure 12.7). (No respondents selected “Unix/ Linux”, “WebTV”, or “Other” in answering this question.) Again these figures (for browser version and primary operating system) are largely in agreement with server log analysis findings (Chapter 13).

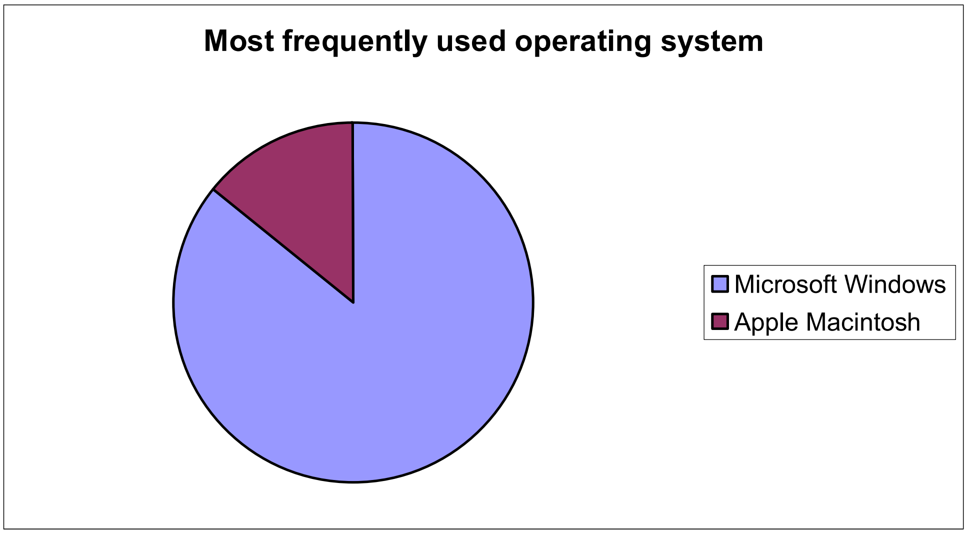
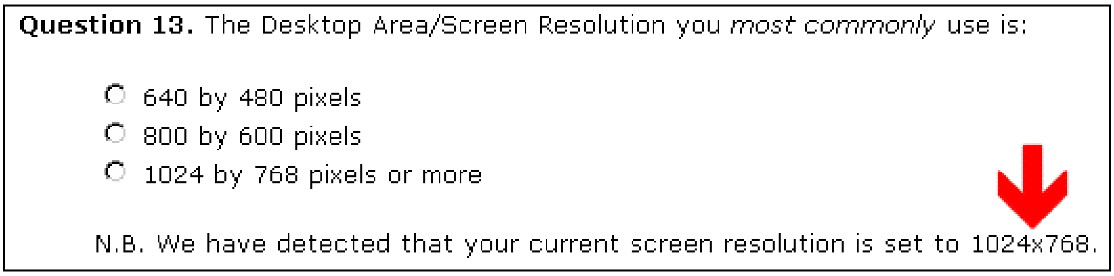


Figure 12.7. Most frequently used operating system.

12.1.3.3 Desktop Area/ Screen Resolution

Fifteen subjects (42.86%) reported “800 by 600 pixels” as the most commonly used Desktop area/ screen resolution, while the remaining 20 respondents (57.14%) reported having their Desktop area set to “1024 by 768 pixels or more” most of the time. (Few years ago, during the early days of Windows operating systems and Web browsing, “640 by 480 pixels” was a very common screen resolution.) To help respondents select the proper answer, the online questionnaire form has a built-in script that displays a respondent’s current screen resolution (Figure 12.8).



Question 13. The Desktop Area/Screen Resolution you *most commonly* use is:

- 640 by 480 pixels
- 800 by 600 pixels
- 1024 by 768 pixels or more

N.B. We have detected that your current screen resolution is set to 1024x768.

Figure 12.8. The online questionnaire form has a built-in script that displays a respondent’s current screen resolution.

It is noteworthy that all HealthCyberMap pilot interfaces display perfectly well with a Web browser window size of 800 by 600 pixels and even less, with no need for scrolling.

12.1.4 Attitude towards the Internet as a Credible Source of Health Information

The Internet is rapidly gaining popularity and credibility as a source of health information. In response to the question “Do you consider the Internet an important source of reliable medical/ health-related information”, 15 subjects (42.86%) replied “To some extent”, while the remaining 20 respondents (57.14%) were more positive and selected “Definitely”. No one selected the “Not at all” answer.

12.2 Usability, Usefulness and User Satisfaction Issues

12.2.1 Usability of the Six HealthCyberMap Pilot Interfaces

Respondents were asked in the questionnaire to rate how well each of the current six pilot information interfaces is fulfilling its intended purpose on a five-step scale from “Extremely well” to “Not at all well” (taking into consideration ease of access, use and navigation, and logical arrangement of information on the pages of concerned interfaces). The results (number of respondents) are shown in Table 12.1.

	Extremely well/Very well	Somewhat well/Slightly well	Not at all well
World maps (geographic provenance)	12	18	5
BodyViewer topical maps	24	11	0
Textual index (ICD-9-CM categories)	21	14	0
Map of resources by type	21	14	0
Resources by language	12	13	10
Semantic subject search engine	18	17	0

Table 12.1. Distribution of respondents’ opinions on the usability of the six information interfaces featured in HealthCyberMap pilot service (by number of respondents for each opinion class).

HealthCyberMap BodyViewer maps of Web resources classified according to medical/ health topics received the best rates (see scaled response bar graph in Figure 12.9). Ten respondents (28.57%) rated HealthCyberMap classification of Web resources according to resource language “Not at all well”. One reason for this might be that respondents were only interested in resources in one language (most probably in English—this is a user-perceived usefulness issue rather than a usability issue). Web resource classification according to resource language is a main ingredient of HealthCyberMap’s proposed content customisation. Ideally, this functionality (classification by language) should be transparently integrated into other interfaces rather than being offered as a separate interface as is the case in the current pilot service. In this way, when the different types of HealthCyberMap maps, e.g., BodyViewer maps, are clicked, only Web resources that are in user's preferred language will be retrieved (or displayed first before resources in other languages if the latter are also retrieved).

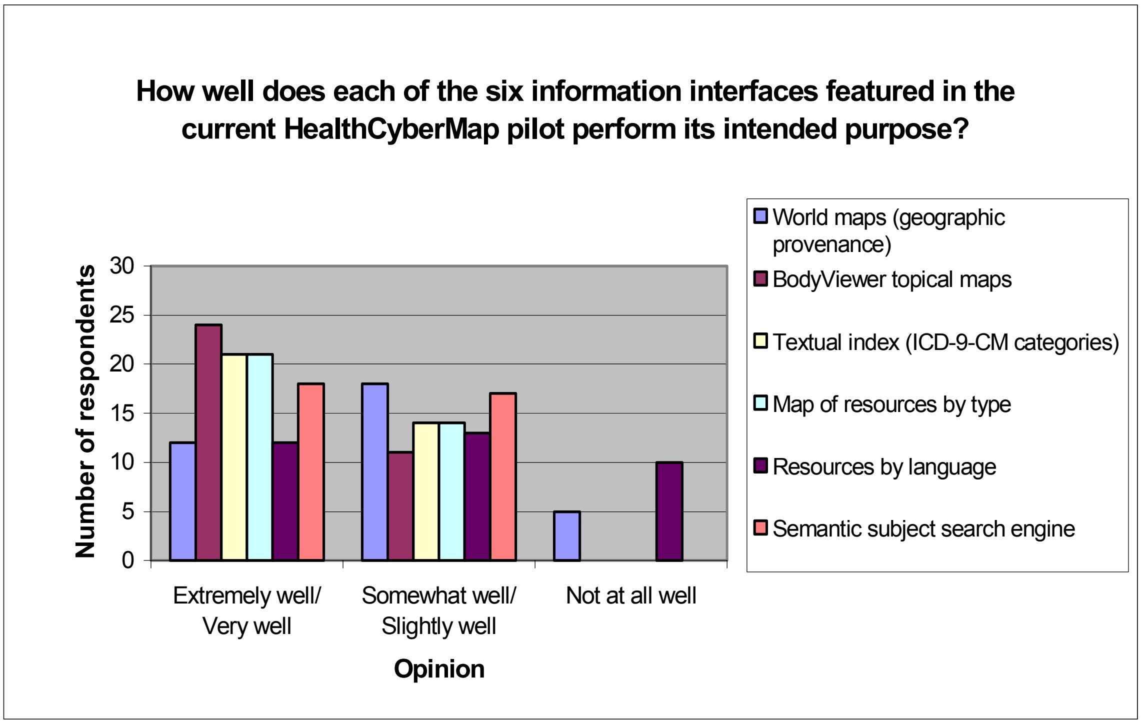


Figure 12.9. Scaled response bar graph showing the distribution of respondents' opinions on the usability of the six information interfaces featured in HealthCyberMap pilot service.

HealthCyberMap world maps (resources classified by geographic provenance) were rated "Not at all well" by 5 respondents (14.29%), most probably because in the current pilot implementation most of the benefits of this kind of classification (finding pointers to *location-specific* health services, disease rates or guidelines) are not available. Another reason might be that clicking a country on the current research pilot world maps with the Identify button  selected displays only (placeholder) information on population, surface area and monetary unit of that country, while some users might have been expecting more health-related information to be displayed instead or in addition. (Both reasons are actually more related to service usefulness.)

Twenty-three respondents (65.714%) described "selecting the right words for search, and finding the right areas/ icons to click on HealthCyberMap maps" as "Very easy". A further 6 respondents (17.143%) selected "Easy" in response to the same question, while the remaining 6 subjects (17.143%) chose "Of moderate (acceptable) difficulty". No one selected "Difficult" or "Very difficult" (Figure 12.10). More than

82% of respondents (29 subjects) described their experience as either “Very easy” or “Easy”, which is an indicator of high service usability.

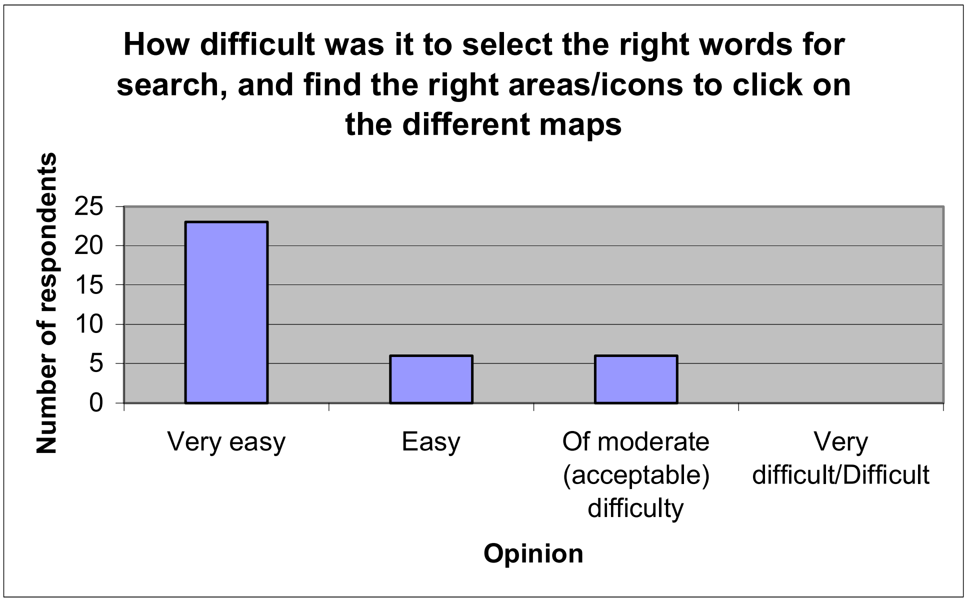


Figure 12.10. Distribution of respondents’ opinions on how difficult is it to select the right words for search and find the right areas/ icons to click on maps.

12.2.2 Specific Usability Issues

12.2.2.1 Metaphor Comprehension

In response to the question “Did you find the meanings of the human body icons on HealthCyberMap BodyViewer maps familiar to you”, 12 subjects (34.29%) answered “Yes”, while the remaining 23 respondents (65.71%) selected “Somewhat (not always)” (Figure 12.11). No one replied “No”. All general practitioners (medically qualified) among respondents chose “Yes”, while all respondents who identified themselves as members of the general public (laypersons) selected “Somewhat (not always)”.

Better icons and maps with enhanced details should be designed for enhanced metaphor comprehension and to care for different user roles and backgrounds.

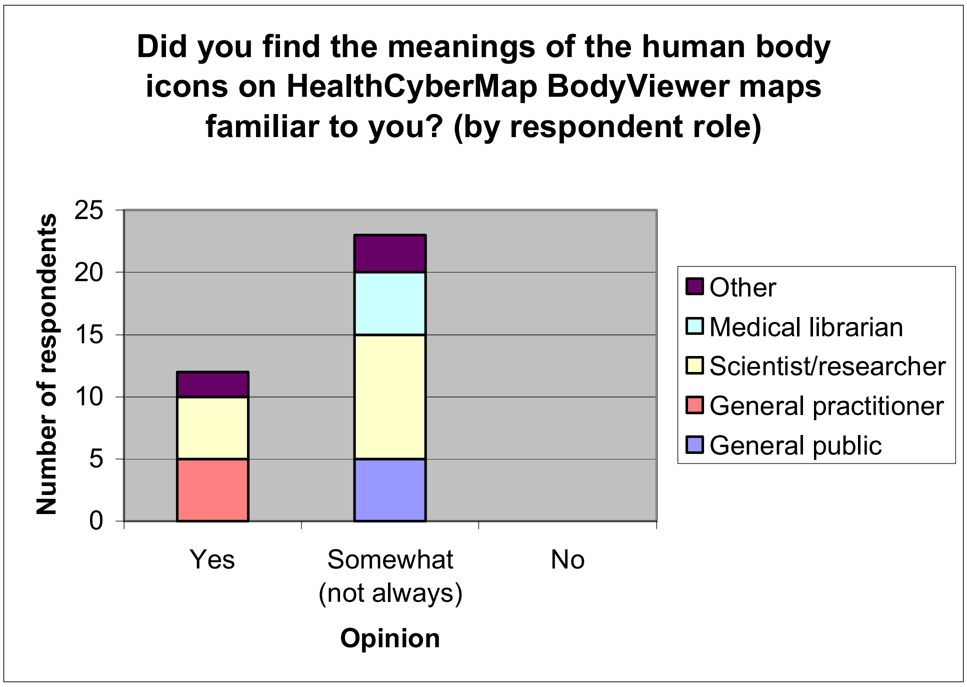


Figure 12.11. Distribution of opinions on metaphor comprehension in BodyViewer maps (by respondent role).

12.2.2.2 Online Help

Most respondents (29—82.86%) either found HealthCyberMap online help and instructions adequate or did not need any help at all (Figure 12.12).

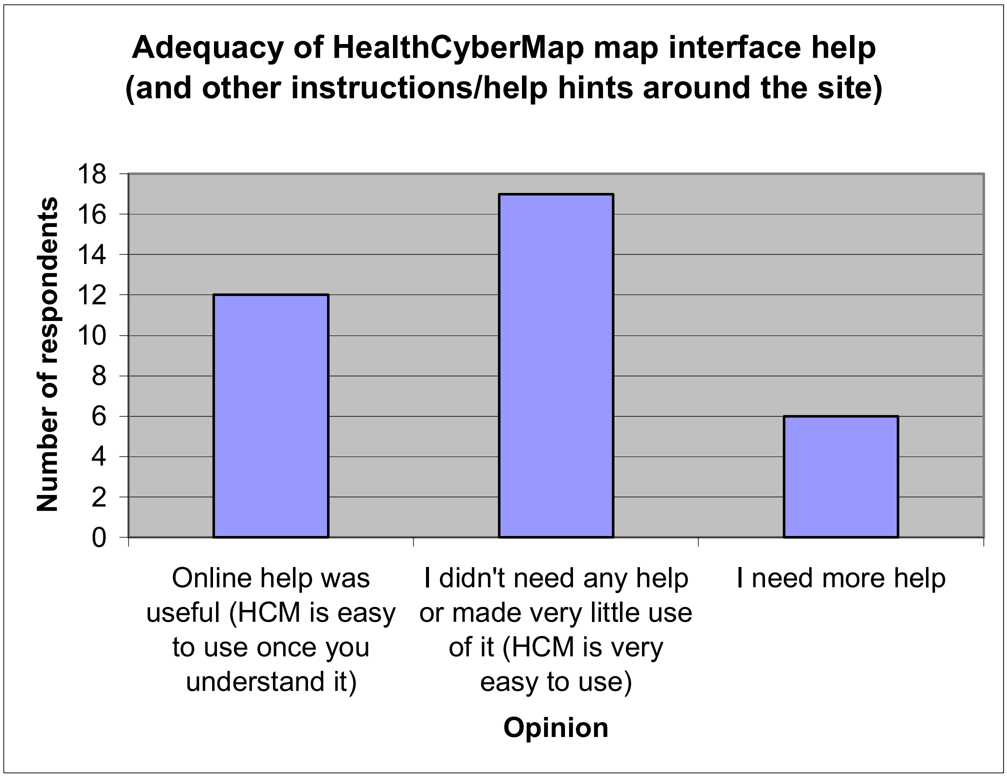


Figure 12.12. Distribution of respondents' opinions on HealthCyberMap map interface help (HCM=HealthCyberMap).

12.2.2.3 Loading Speed of HealthCyberMap Maps

When asked about the speed at which HealthCyberMap maps load on their Internet connection, most respondents (29—82.86%) answered “Reasonably fast”. Only 6 subjects (17.14%) selected “Slow” in response to this question. Ironically, 5 among those 6 who chose “Slow” had an ISDN, ADSL, or broadband Internet connection type (supposed to be faster) suggesting that their reported slow speed was most probably due to Internet traffic congestion rather than to problems with HealthCyberMap design (like file size of the different map images—Figure 12.13).

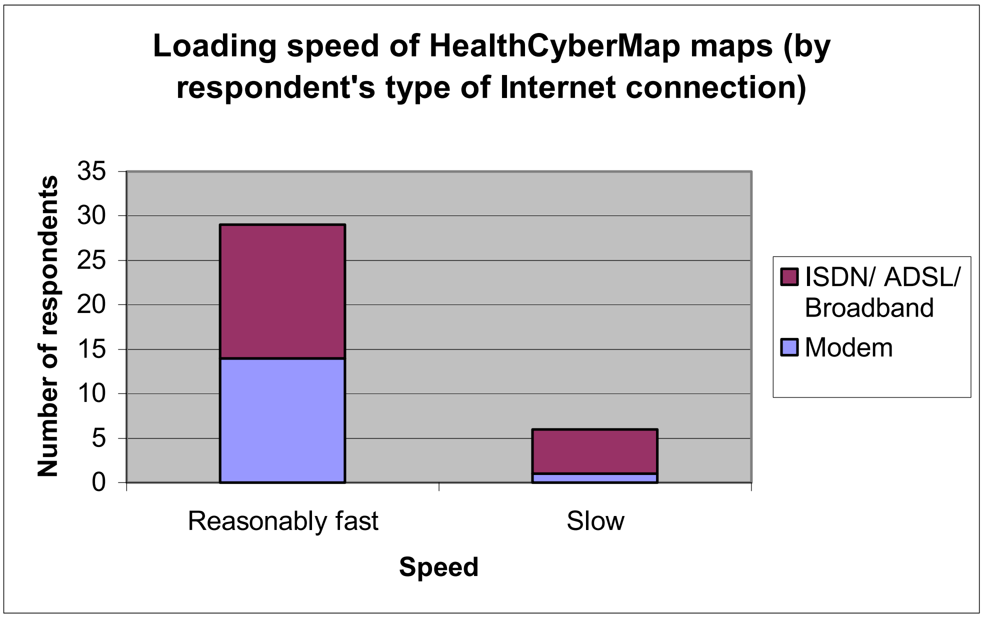


Figure 12.13. Distribution of map loading speed by respondent’s type of Internet connection.

12.2.2.4 Detail Overload

In HealthCyberMap, map and search engine query results (resources) are listed in a separate text window to avoid map clutter. The latter would have been unavoidable had we opted to represent each resource using a distinct point symbol on the map (cf. Visual Net PubMed interface—<http://pubmed.antarcti.ca/start>—Figure 11.2). Also the “Show/ Hide Details” option for each returned resource record in query results is collapsed by default (Figure 9.1). Respondents were asked about the current HealthCyberMap presentation format of query results/ resource details if it causes them to get overloaded quickly with too much detail. The answers varied from “not at all a problem” or “a slight problem” (18 subjects—51.43%), to “a moderate problem” (8 subjects—22.86%) and “a significant problem” (9 subjects—25.71%—Figure 12.14).

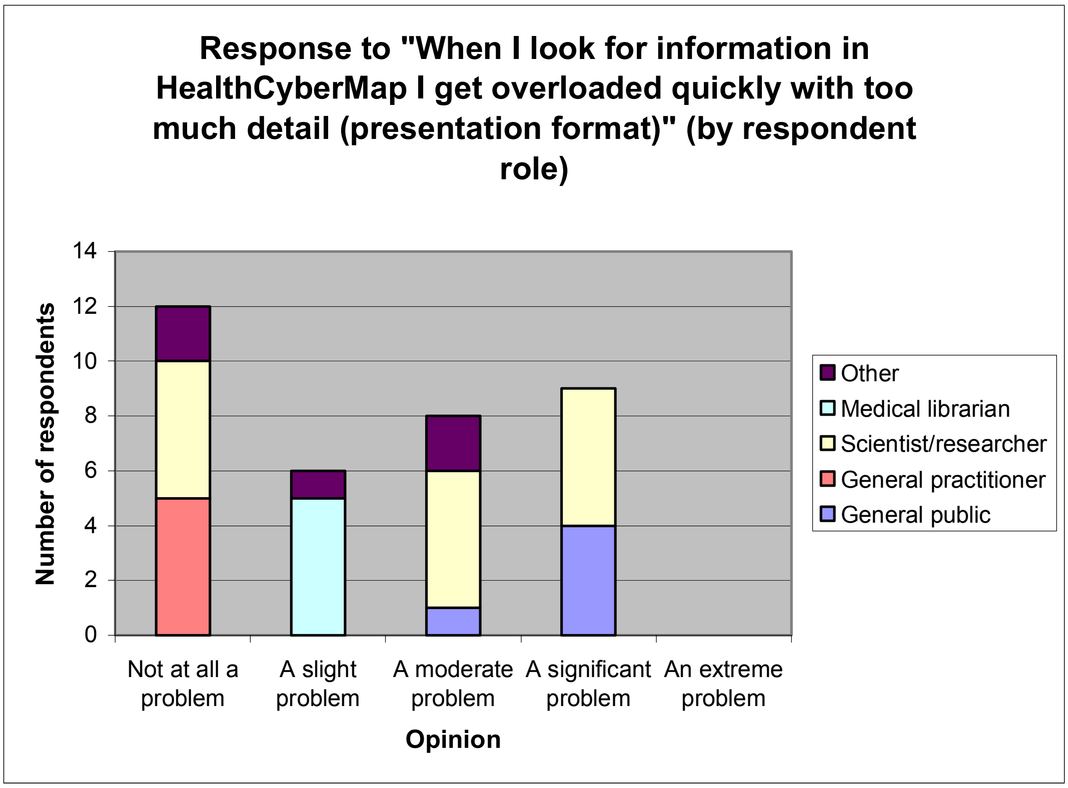


Figure 12.14. Distribution of respondents' opinions on detail overload in HealthCyberMap.

These results could be due to the fact that at this pilot stage, HealthCyberMap interfaces are not fully implemented or optimised for their intended purpose (they are just "concepts in their infancy"). For example, the current textual resource index is only based on ICD-9-CM top-level categories; clicking one of those categories triggers a query that returns a large number of resources (e.g., all "Endocrine Diseases"). However, it is very unlikely that a user would be interested in browsing all resources belonging to a given ICD-9-CM top-level category at once, and if interested, a large number of resources returned could make browsing very difficult. Ideally, the query should be focused on smaller groups of resources/ narrower topics by enabling users to browse and select hierarchical sub-levels (subcategories) of the main (top-level) ICD-9-CM categories (Figure 15.4).

Two of the proposed HealthCyberMap future interfaces/ enhancements could also improve the current response to this question (see below). These are HealthCyberMap customisation and multi-axial classification of resources based on two or more Dublin Core metadata elements (according to user needs). Focusing on individual user needs is the key to minimising the number of results returned and maximising their relevance and suitability for a particular user situation (managing information overload). For example, the service should be able to retrieve only "English language"

(DC language element) “Guidelines” (DC type element) on say “diabetes mellitus” (DC subject element) rather than every resource type in any language on that condition if all what the user is currently interested in are just the English language guidelines.

12.2.3 Relevance and Quality of HealthCyberMap Resource Pointers

The big majority of respondents (24—68.57%), including *all* members of the general public, medical practitioners and librarians in the study sample, either “Strongly agree” or “Agree” with the statement “The pointers to information resources returned by HealthCyberMap were *relevant* to my queries”. Six respondents (17.14%) were neutral (i.e., “Neither disagree nor agree”). The remaining 5 subjects (14.29%) “Disagree” with the statement (Figure 12.15).

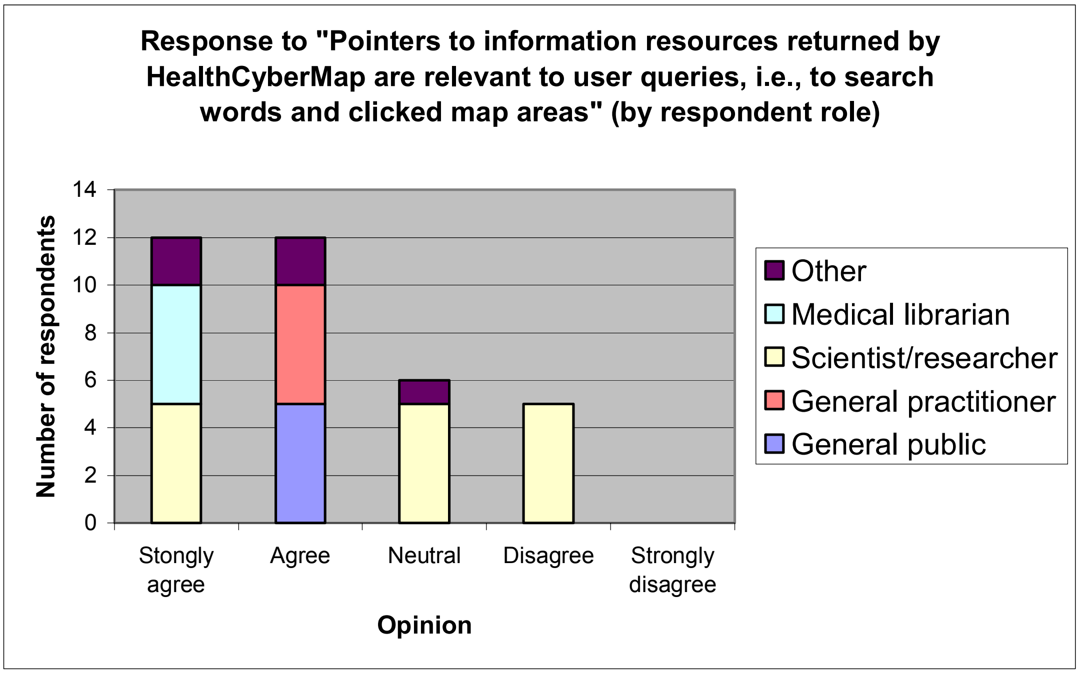


Figure 12.15. Distribution of respondents’ opinions on the relevance of HealthCyberMap resource pointers to user queries (by respondent role).

Eighteen respondents (51.4%), including *all* medical practitioners and librarians in the study sample, think that HealthCyberMap pointers to information resources are of good quality, accurate, up-to-date and useful. The remaining 17 subjects (48.6%), including *all* respondents who identified themselves as members of the general public, chose a neutral position (i.e., “Neither disagree nor agree”—Figure 12.16).

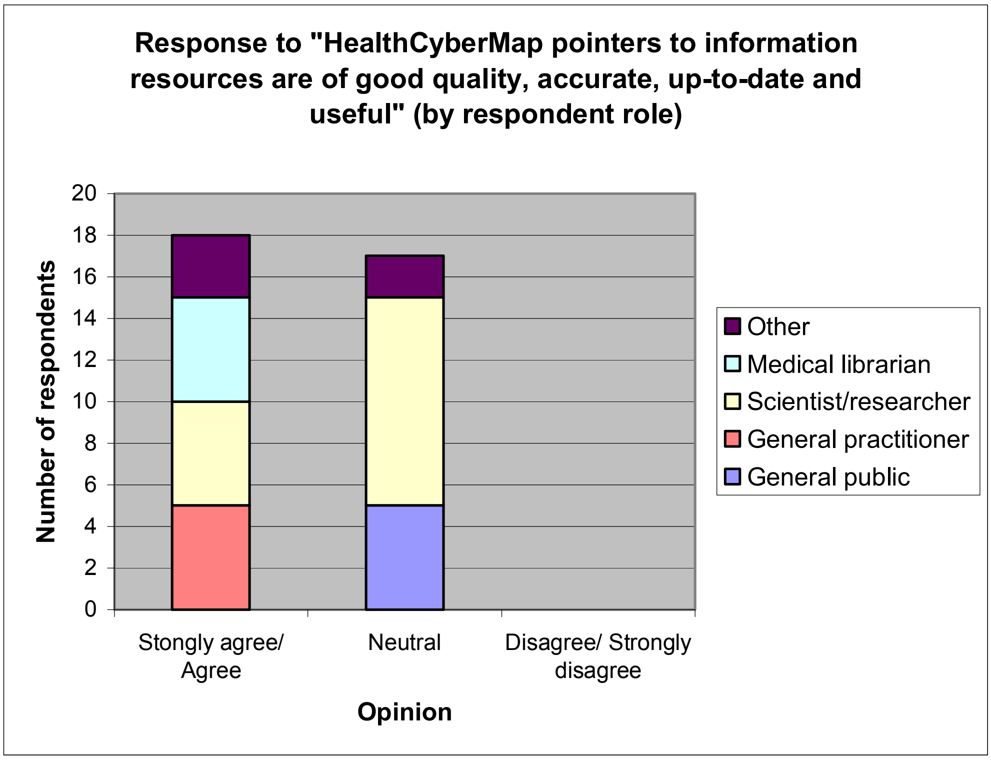


Figure 12.16. Distribution of respondents' opinions on the quality of HealthCyberMap resource pointers (by respondent role).

12.2.4 Usability Task

Respondents were given a usability task to accomplish. The task goal was to find information resources about specific diseases using HealthCyberMap. Respondents were asked to choose at least one disease from the following list: prostate cancer, Crohn's disease, Guillain-Barre syndrome, amyotrophic lateral sclerosis, diabetes mellitus, Mitral valve prolapse, ectopic pregnancy (or any other medical/ health topic(s) of their choice). Eleven respondents (31.43%) were successful from the first attempt, while 22 subjects (62.86%) were successful after one or more failed attempts. There were no respondents who were not successful in completing their task, but 2 subjects (5.71%) chose not to select any answer to this question (Figure 12.17).

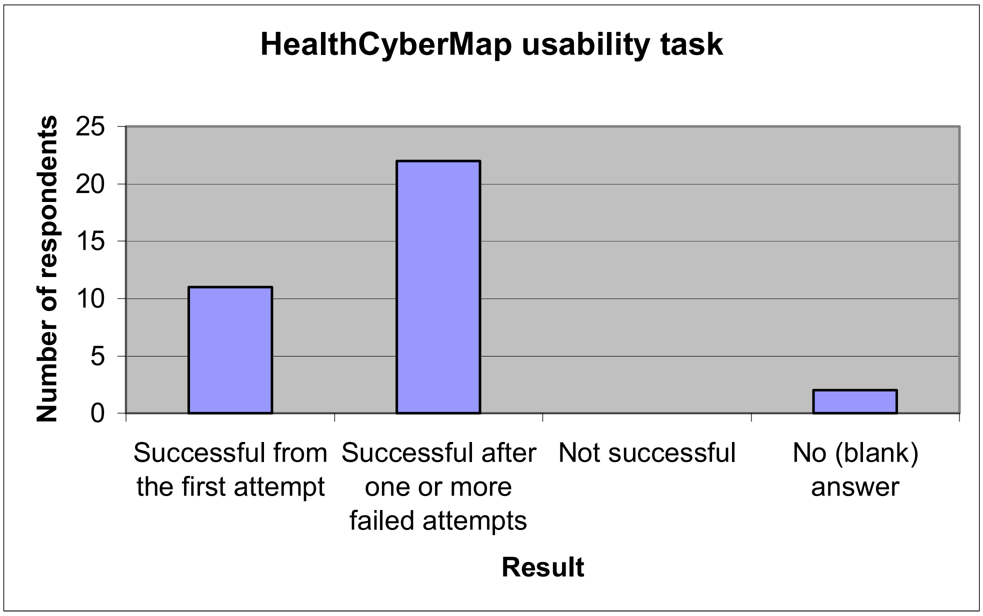


Figure 12.17. Distribution of HealthCyberMap usability task results.

It should be noted that at this pilot stage of HealthCyberMap development, not all medical/ health topics are covered and the textual index of resource categories currently only covers top-level categories. Also BodyViewer child maps (the maps that should appear when clicking icons on the main BodyViewer map) are not fully implemented in the current Web version (that respondents were evaluating) compared to the same maps in HealthCyberMap original project in ArcView GIS. Given these limitations in the current pilot HealthCyberMap service, the results of this usability task seem very encouraging.

Of the 33 respondents who have successfully accomplished HealthCyberMap usability task, 6 subjects (6/33 = 18.2%) selected "BodyViewer maps" as the interface used *most of the time* to accomplish that task. A further 6 subjects (6/33 = 18.2%) reported using the "semantic subject search engine (advanced keyword search)" *most of the time* for the same task. Five respondents (5/33 = 15.1%) chose the "textual index (ICD-9-CM top-level categories)" as the interface used *most often* for the usability task, while 16 subjects (16/33 = 48.5%) found "all three interfaces equally useful" in accomplishing this task (Figure 12.18).

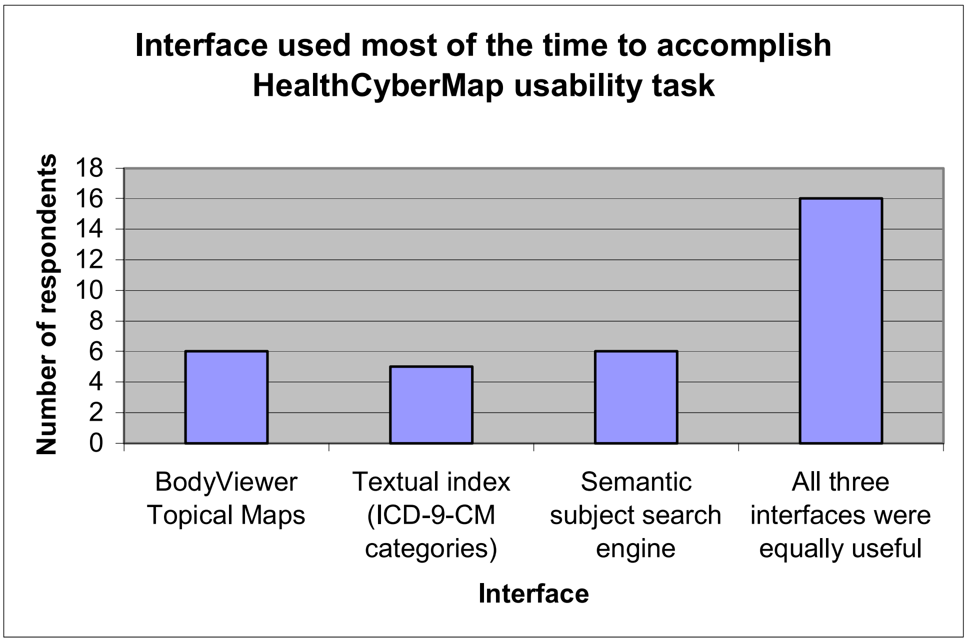


Figure 12.18. Distribution of the interfaces that respondents used most of the time to accomplish HealthCyberMap usability task.

12.2.5 Usefulness of the Six HealthCyberMap Pilot Interfaces

Respondents were asked in the questionnaire to rate how important/ useful they found each of the current six pilot information interfaces on a five-step scale from "Extremely important" to "Not at all important". The results (number of respondents) are shown in Table 12.2.

	Extremely important/ Very important	Somewhat important/ Slightly important	Not at all important
World maps (geographic provenance)	17	18	0
BodyViewer topical maps	24	11	0
Textual index (ICD-9-CM categories)	18	17	0
Map of resources by type	12	23	0
Resources by language	6	24	5
Semantic subject search engine	18	12	5

Table 12.2. Distribution of respondents' opinions on the usefulness of the six information interfaces featured in HealthCyberMap pilot service (by number of respondents for each opinion class).

This time again, HealthCyberMap BodyViewer maps of Web resources classified according to medical/ health topics scored highest while the classification of Web resources according to resource language received the worst ratings (see scaled response bar graph in Figure 12.19).

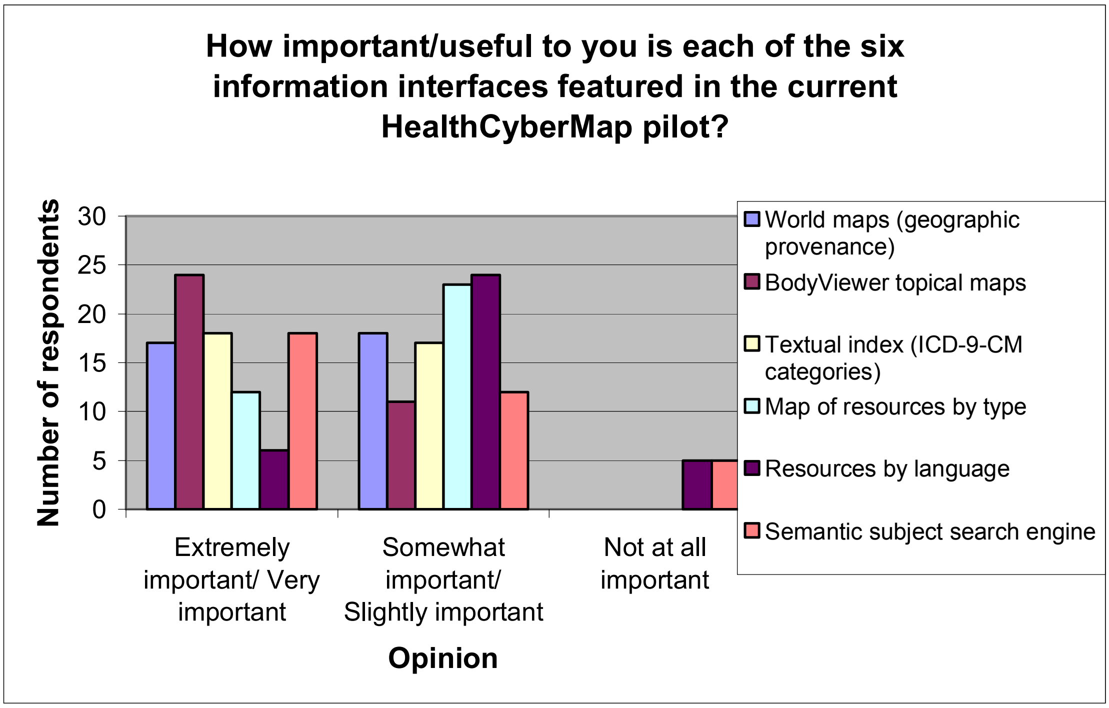


Figure 12.19. Scaled response bar graph showing the distribution of respondents' opinions on the usefulness of the six information interfaces featured in HealthCyberMap pilot service.

The fact that BodyViewer maps (a visual navigation interface) ranked higher than HealthCyberMap Semantic Subject Search Engine in this analysis could be explained in light of Fuller and de Graaff statement on Web search engines [115]. With search engines, users must know what they want *before* they could find it (to type it in the search box) [115]. In contrast to this, the exploratory nature of visual maps invites users to discover and learn new things they did not know or plan to learn before.

12.2.6 Usefulness of the Proposed HealthCyberMap Future Directions

Respondents were asked "how important/ useful they think are the proposed HealthCyberMap future interfaces/ directions". For each proposed interface, they had to select an answer on a five-step scale from "Extremely important" to "Not at all important". The results (number of respondents) are shown in Table 12.3.

	Extremely important/ Very important	Somewhat important / Slightly important	Not at all important
Multi-axial classifications	13	18	4
Customisation (including location-based)	18	17	0
Problem to knowledge linking	19	12	4
Mapping real-world health problems	28	7	0

Table 12.3. Distribution of respondents' opinions on the usefulness of the proposed HealthCyberMap future interfaces/ directions (by number of respondents for each opinion class).

All four future directions received favourable ratings by respondents. However, "mapping health problems in HealthCyberMap and identifying information needs and gaps" was given the best rates by the highest number of respondents (see scaled response bar graph in Figure 12.20).

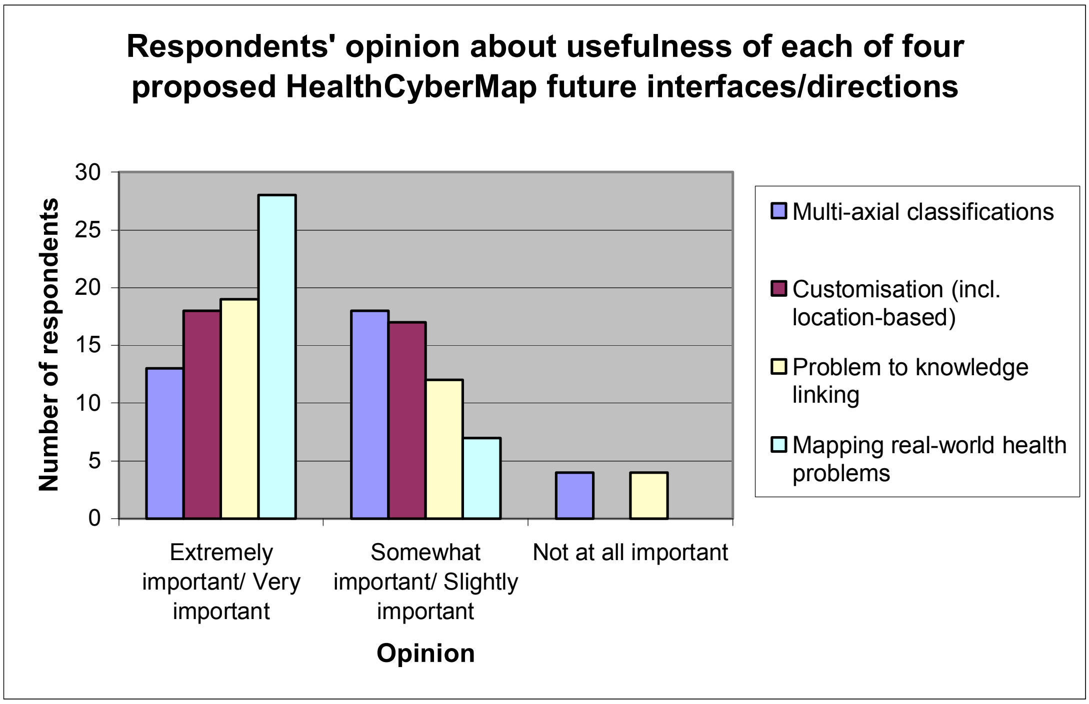


Figure 12.20. Scaled response bar graph showing the distribution of respondents' opinions on the usefulness of the proposed HealthCyberMap future interfaces/ directions.

Mapping real-world health problems remains after all the "native" application of GIS in health (compared to mapping cyberspace), and besides its classic benefits [47], it can also inform health information providers to develop any required resources or modify existing ones in light of any identified real-world needs.

Two future directions, "customisation/ location-based customisation" and "problem to knowledge linking", were ranked next after "mapping health problems". "Problem to knowledge linking" was rated by all five general practitioners among respondents as "Extremely important". This is a very good result since "problem to knowledge linking" is primarily intended for healthcare professionals accessing the electronic patient record.

"Multi-axial classification of resources based on two or more Dublin Core metadata elements" scored lowest compared to the three other HealthCyberMap future directions evaluated in this question. One explanation for this result could be that much like the "resource classification by language" interface (see above), this functionality is also most useful when transparently integrated into other interfaces, which was possibly not very clear in the current standalone demonstrator that respondents evaluated (<http://www.healthcybermap.org/multiaxial.htm>). When integrated into other interfaces (e.g., the semantic search engine or BodyViewer maps), multi-axial classification could help filtering and focusing query results to a much smaller, more relevant and more easily manageable set.

12.2.7 User Acceptance and Satisfaction

In response to the question "Based on what you have seen so far, if HealthCyberMap was fully developed, would you use it as an information portal and one-stop access point to health resources", 27 subjects (77.14%) replied "Yes", 5 (14.29%) replied "No", while the remaining 3 (8.57%) respondents did not select an answer (i.e., blank response, could be possibly interpreted as "Not sure").

Nineteen respondents (54.29%) "Strongly agree" or "Agree" with the statement "HealthCyberMap concepts and interfaces meet my information retrieval and navigation needs (not topical coverage) better than other medical/ health information portals/ gateways". The remaining 16 respondents (45.71%) "Neither disagree nor agree". No one disagreed with the statement (Figure 12.21).

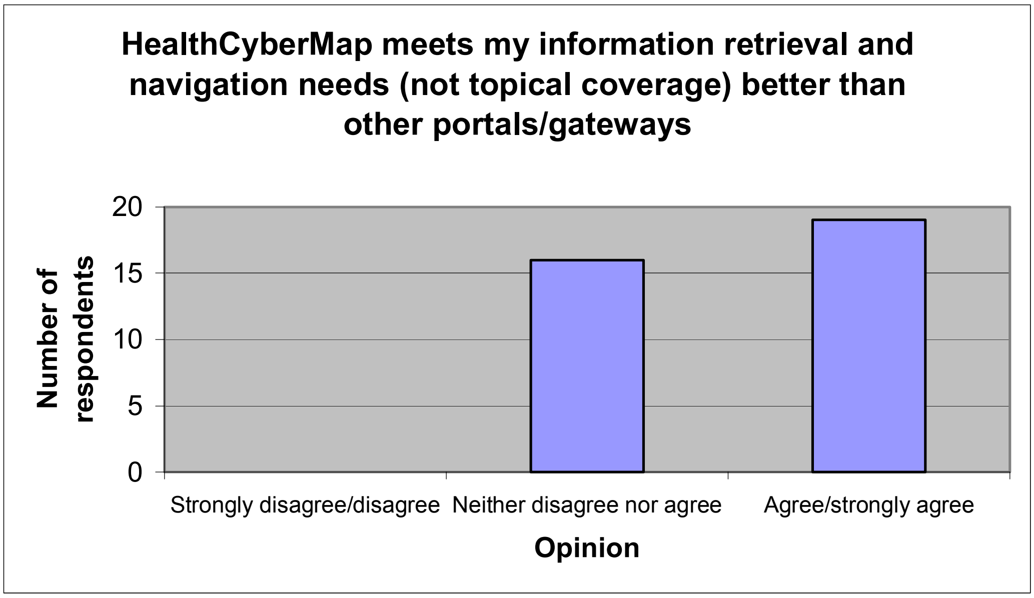


Figure 12.21. Respondents' opinions on whether HealthCyberMap is meeting their information retrieval and navigation needs (not topical coverage in this pilot service) better than other portals gateways.

One respondent, a general practitioner, typed <http://www.doctors.net.uk>, a portal for doctors, in response to the next open routing question "(*Only if you did not agree with* *the above statement*) At this time, what other medical/ health portals/ gateways are better than HealthCyberMap regarding information retrieval and navigation (not topical coverage)", though he had answered the previous question with "Neither disagree nor agree". All other respondents left this routing question blank.

In response to the statement "I prefer to use a conventional search engine (e.g., Google) to gather information or to have a librarian, staff member, or family member gather information for me", 17 subjects (48.57%) selected "True", and 18 (51.43%) chose "False".

12.2.8 Attitude towards Visual Maps as Navigational Aid

Most respondents (29—82.86%) selected "Yes, definitely" in response to the question "Do you think visual maps are a useful addition to/ complementary improvement over conventional text-based Web portal interfaces". The remaining 6 subjects (17.14%) chose "Could be an improvement (but not always)" in response to the same question. The answer "No, visual interfaces are useless" was not selected by any respondent, so one can assume that all respondents generally had a positive and welcoming attitude towards visual maps as one useful form of Web portal interfaces (Figure 12.22).

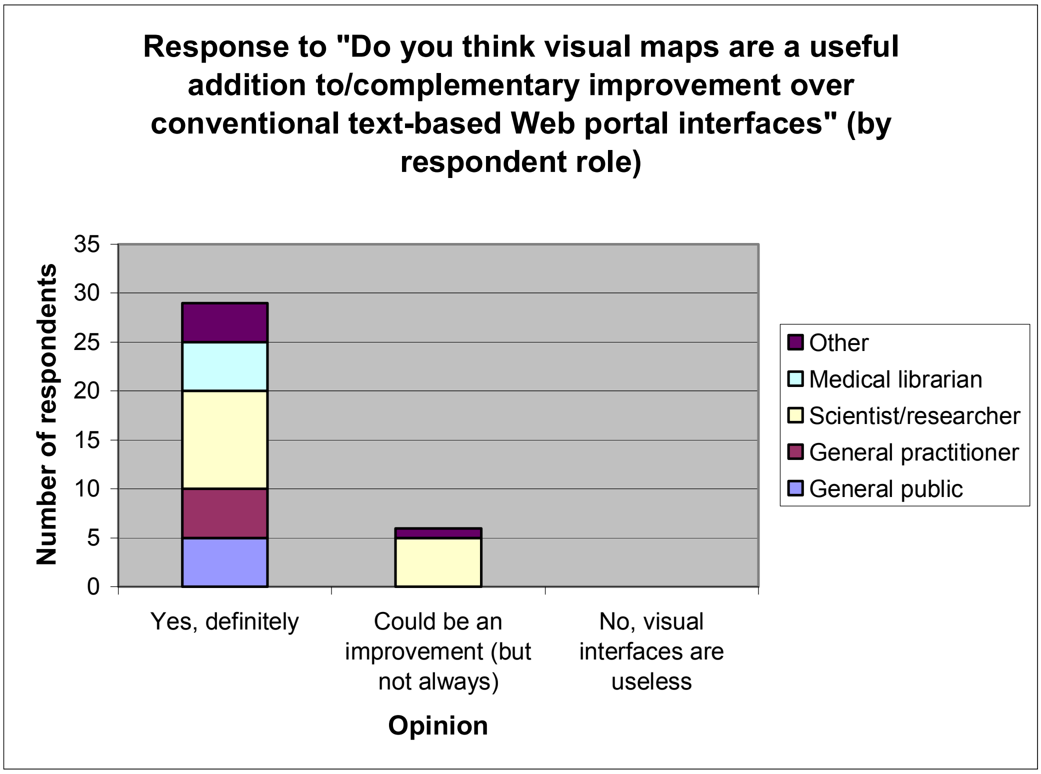


Figure 12.22. Attitude towards visual maps as navigational aid (by respondent role).

12.2.9 Comparative Task (HealthCyberMap vs. Visual Net)

Respondents were asked to compare HealthCyberMap mapping approach to that of Visual Net PubMed interface (link to the latter was given on the online questionnaire form—<http://pubmed.antarcti.ca/start>) regarding the use of familiar metaphors and presentation format of resource details and links (management of detail overload). Eighteen respondents (51.4%) found HealthCyberMap approach to be either “superior” or “somewhat better” than that of Visual Net. Twelve subjects (34.3%) found both approaches to be about the same, while only 5 respondents (14.3%) felt HealthCyberMap approach is inferior (Figure 12.23).

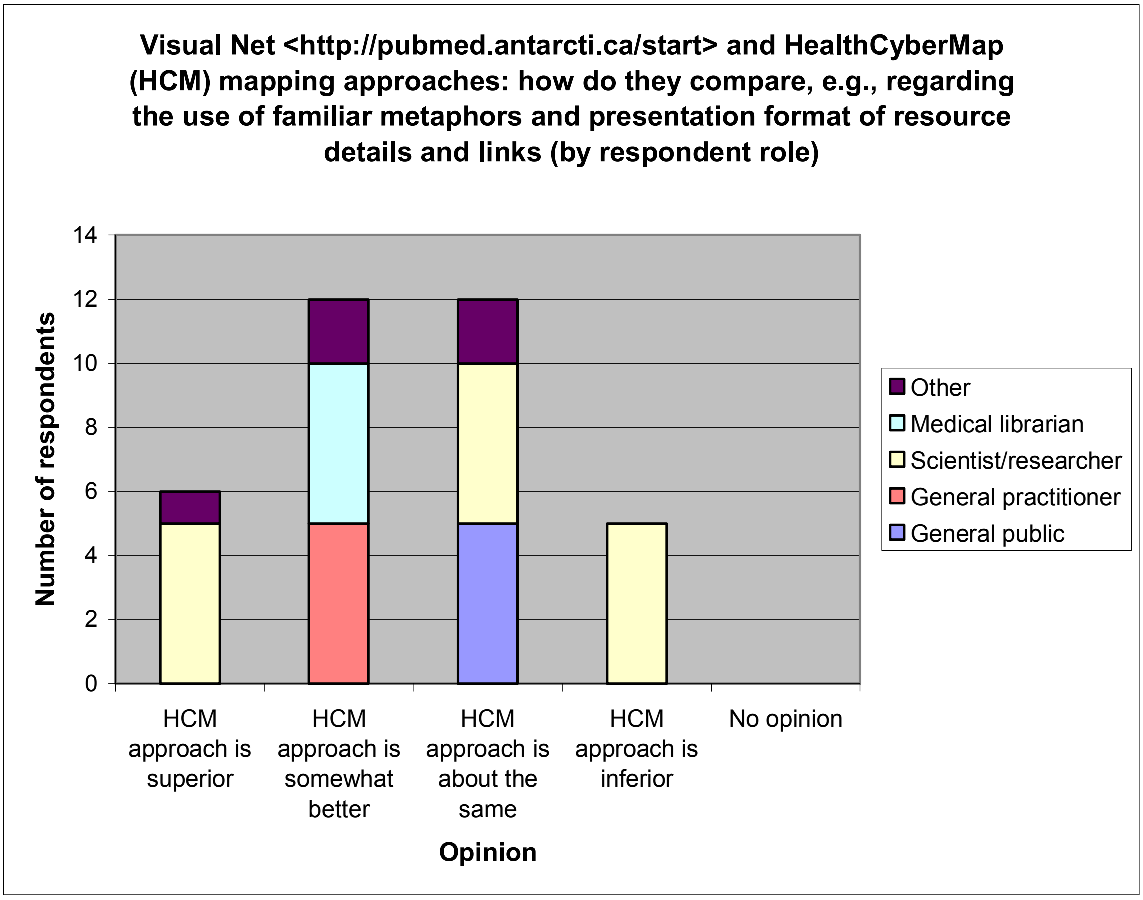


Figure 12.23. Comparative task results (by respondent role).

12.3 Other Issues

12.3.1 Organising Resources by Intended Primary Audience

Twenty-five respondents (71.43%) thought it would be useful in a future implementation of HealthCyberMap to also organise information resources by user needs/ intended primary audience as patients, health professionals, or basic researchers (i.e., answered “Yes”). Those who answered “Yes” included *all* members of the general public and general practitioners in the study sample. Only 6 subjects (17.14%) answered "No" (including all five medical librarians among the study sample), while the remaining 4 respondents (11.43%) did not select an answer (i.e., blank response, could be possibly interpreted as "Not sure").

It should be noted that this sample of respondents did not include subjects with only school level education or vocational training, and did not include any patients at all (respondents identifying themselves as "members of the general public touched by disease"). It only included a small percentage (14.3%) of subjects identifying themselves as "members of the general public seeking health-related information". The general practitioners, scientists and researchers who answered "Yes" to this question might just be holding the classic viewpoint that patients should not be offered access to the same level of knowledge as physicians (also it has been said physicians sometimes feel "threatened" by a well-informed patient).

We believe information should be made accessible to *all*, but preferably in two distinct forms. A simplified jargon-free form for laypersons and patients who cannot cope with medical jargon (and also for healthcare professionals looking for this kind of presentation maybe to compile and print a patient information leaflet for their patients), and the full professional form intended for healthcare professionals, but also made freely accessible to interested members of the general public and patients who can understand it.

12.3.2 Service Access Frequency (Visitor Retention)

Ten respondents (28.57%—all of them describing themselves as "Scientist/ researcher") have accessed HealthCyberMap pilot Web site on a weekly basis over the past three months (before filling in the questionnaire). A further ten respondents (28.57%) reported accessing HealthCyberMap Web site every month over the past three months (before filling in the questionnaire). Fifteen subjects (42.86%) selected "This is the first time I have accessed the Web site" at the time they were filling in the questionnaire (Figure 12.24). No one reported accessing the Web site on a "Daily" basis, most probably because at this early pilot stage, HealthCyberMap does not yet provide enough topical coverage to make the service a daily stop for information seekers.

Scientists/ researchers were probably accessing the site more frequently (weekly) compared to other role groups because they are interested in HealthCyberMap methodology ("under the hood") rather than for any real health information seeking purposes (this hypothesis is supported by their role as scientists/ researchers).

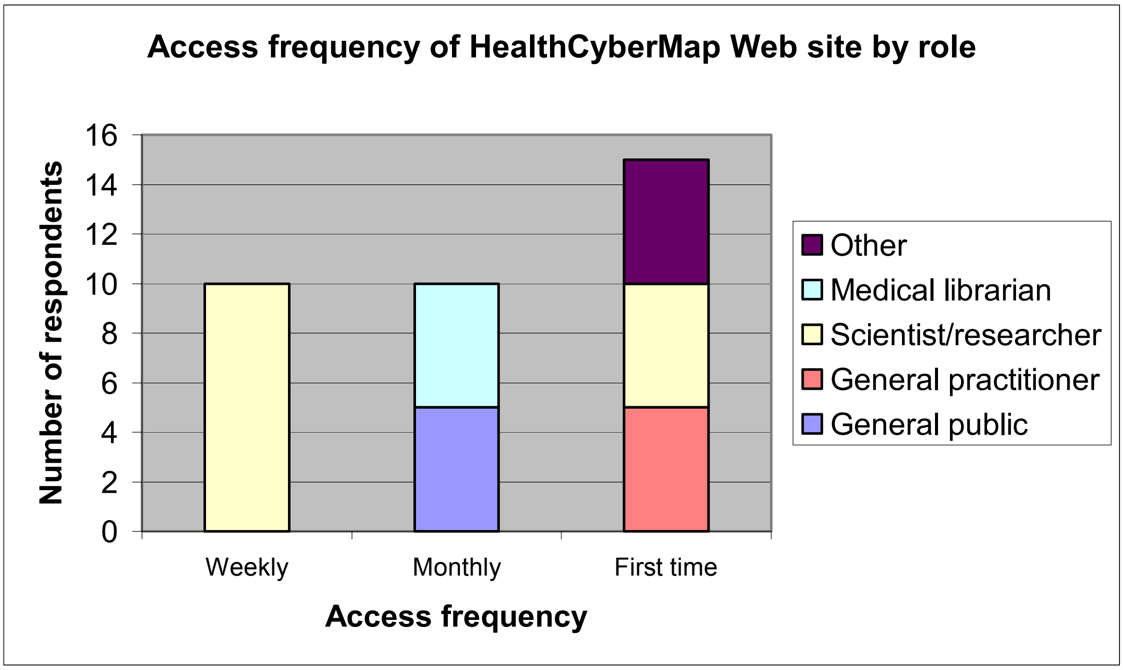


Figure 12.24. Access frequency of HealthCyberMap Web site by role.

12.3.3 Service Advertising Methods

(Cf. Referring search engines/ external sites in server log analysis—Chapter 13)

Five respondents (14.29%) first learned about HealthCyberMap by listening to a presentation and/ or reading a paper on the project. Ten respondents (28.57%) first heard about HealthCyberMap site through an e-mail or e-mail list message (e.g., message posted on SemanticWeb.org Group <http://groups.yahoo.com/group/semanticweb/message/479>). Fifteen subjects (42.86%) first came to HealthCyberMap site via an Internet search engine or a link from another site. Search engines and referrer links proved to be the most effective method of dissemination in our case, even better than mass e-mail methods. HealthCyberMap is currently fairly well exposed on the Web. Google (an Internet search engine, <http://www.google.com/>) for example returns 98 results for "healthcybermap" from HealthCyberMap Web site and also from external sites referring to HealthCyberMap (as of June 2002). The remaining five respondents (14.29%) learned about HealthCyberMap by being personal friends or relatives of researcher involved in this project (Figure 12.25). Although this latter category of respondents (friends and relatives) could be a potential source of bias, we decided not to exclude it from our study. No one selected "A colleague told me about it" or "Other".

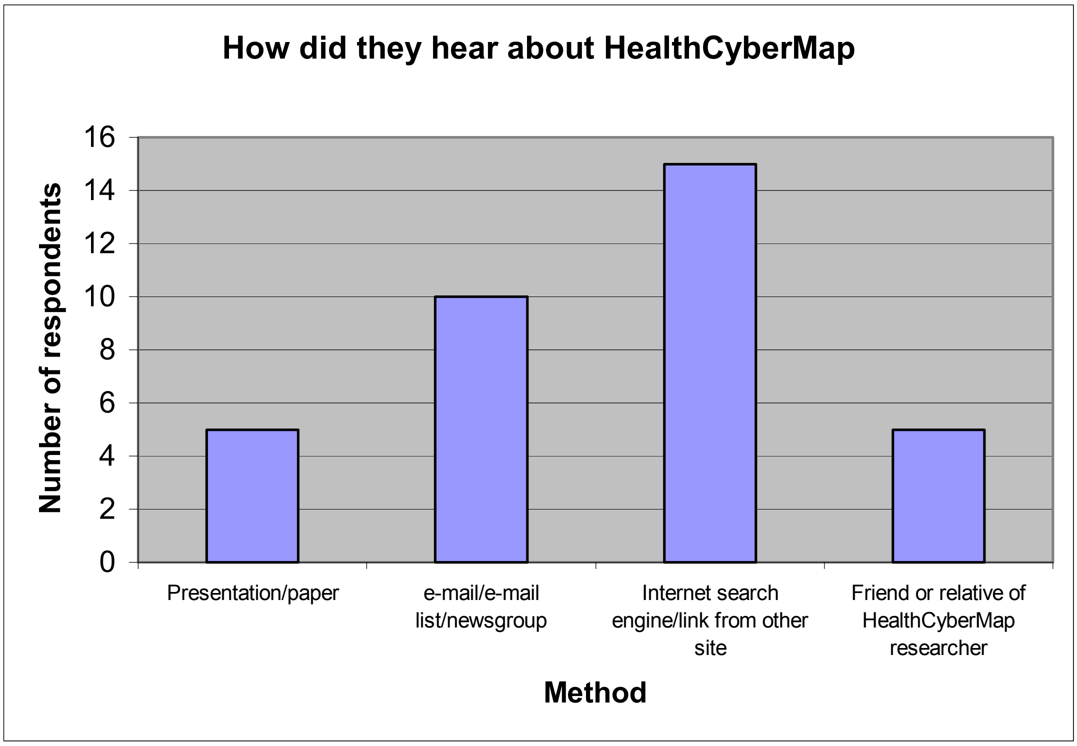


Figure 12.25. Distribution of the different advertising methods through which respondents first learned about HealthCyberMap.

12.3.4 Responses to Open Questions

The open questions part of the questionnaire (questions 37-41) yielded very little results. These questions are a good way of finishing a questionnaire. They give respondents the freedom to express and explain in free, non-restrictive text what they like or dislike in the service, as well as any personal comments or opinions they might have, which closed questions with prescribed answers do not allow [93].

Like all other questions in this questionnaire, answering the open questions was optional, but this alone does not explain the low response rates to open questions (because the answer rate of closed questions was nearly 100% and respondents had the option to leave them blank as well). One plausible explanation could be that typing a free-text answer takes time and is not seen as easy as selecting an answer from a pre-defined set of answers. Some respondents might think there is a possibility of identifying them based on their writing style or particular opinions and might not like this. Gillham [93] mentions that it is easier for human beings to talk than to write.

For this reason, interviews (talking) might be more fruitful compared to open questions (writing).

Only two respondents (5.7%) answered the question "Were there any parts of the service that you found especially helpful? What do you like most about HealthCyberMap and why?" One subject answered "HealthCyberMap Web resources classified according to health topics" and the other typed "Yes, HealthCyberMap Web resources classified according to ICD-9 and according to resource type".

Only one subject (2.85%) answered the question "Were there any parts of the service that you found especially difficult to use or understand? What do you dislike most about HealthCyberMap and why?" by typing a simple "None".

Only four respondents (11.43%) answered the question "Should resources (if available) be invested to continue developing and implementing HealthCyberMap? Why or why not?" All four answers were positive:

"Yes, because it collects a wide variety of information concerning medical problems"

"Yes, as it is a great help to medical research and clinicians"

"Yes, this is a very promising service"

"Of course, if you can show that you offer a unique service"

In response to the question "What are your suggestions or comments about what would make the service better", only two respondents (5.7%) typed their comments, which included a wish to have the service "frequently updated" and a remark that the "world national 'stats' seemed to be minimal" in this pilot. Concerning this latter observation, the respondent might have been expecting more details and health-related information to be displayed when clicking a country on the world maps with the Identify button  selected (currently, only "placeholder" information on population, surface area and monetary unit of that country are displayed).

Three respondents (8.57%) gave their feedback on this formative evaluation questionnaire in response to the last question "we would like to hear your thoughts about this questionnaire. For example, are we asking the right questions? Are we asking the questions in the right way?" Two respondents answered very positively: "it is a well structured questionnaire" and "the questionnaire is very well designed", while one subject (role: scientist/ researcher) commented "too long, too many questions".

12.4 Conclusion

Thirty-five subjects responded to HealthCyberMap online evaluation questionnaire during the 45-day period from 18 April 2002 to 1 June 2002. The sample of respondents was in general a good representative of the wider Internet/ HealthCyberMap audience regarding age, sex, and computer and Internet use (and equipment). All respondents hold a university/ college or postgraduate degree. There were no subjects with lower levels of education. General practitioners, scientists/ researchers and medical librarians were well represented in the sample, which did not include any nurses or patients at all, though it included subjects identifying themselves as “members of the general public seeking health-related formation”. This could be due to the fact that at this pilot stage of HealthCyberMap (with rather limited Web interface functionality and only 1640 resources in the underlying metadata base) the service is of more interest to the professional and scientific community than to the general public.

Most respondents hold a very positive attitude towards the Internet as a credible source of health information and towards visual maps as a navigational aid for medical and health-related Internet resources.

HealthCyberMap scored high on all general and specific questions covering user satisfaction and service usability, e.g., metaphor comprehension, ease of use and adequacy of online help, map loading speed and management of detail overload. More than 94% of respondents were successful in completing the usability task.

HealthCyberMap BodyViewer maps of Web resources classified according to medical/ health topics received the highest usability and usefulness rates among the evaluated pilot interfaces.

The big majority of respondents thinks HealthCyberMap resource pointers are of good quality and relevant to user queries.

All proposed HealthCyberMap future directions received favourable ratings by respondents, with “mapping health problems in HealthCyberMap and identifying information needs and gaps” scoring highest. “Problem to knowledge linking” (intended primarily for healthcare professionals accessing the electronic patient record) was rated “Extremely important” by all general practitioners among respondents.

Most respondents also think it would be useful in a future implementation of HealthCyberMap to also organise information resources by intended primary audience as patients, health professionals, or basic researchers.

Internet search engines and links from other sites proved to be more effective than targeted e-mail and postings to mailing lists in advertising the service.

The results of the comparative task were very encouraging, with most respondents finding HealthCyberMap approach to be better than or as good as that of Visual Net PubMed interface.

Specific ideas for improving the service in light of the formative evaluation questionnaire results include:

- Designing better icons and maps with enhanced details for better metaphor comprehension, caring for different user roles and backgrounds;
- Using HealthCyberMap world maps as an interface to *location-specific* health services, disease rates and guidelines in addition to the current use as an interface to resources classified by the geographic provenance of author(s)/ publisher(s);
- Transparently integrating some of the evaluated interfaces ("classification of resources by language" and "multi-axial classification of resources") that were never intended to function as standalone interfaces into other HealthCyberMap interfaces.

13 Analysis Results of HealthCyberMap Server Log for the Formative Evaluation Study Period

The author used Sawmill 6.3.8 server log analysis tool to perform the analysis of 26,365 individual entries (lines) in HealthCyberMap server log files corresponding to the 45-day period beginning 18 April 2002 and ending 1 June 2002 (Figure 13.1). (The logs have been downloaded from HealthCyberMap server for analysis on a local Windows NT 4.0 machine.)

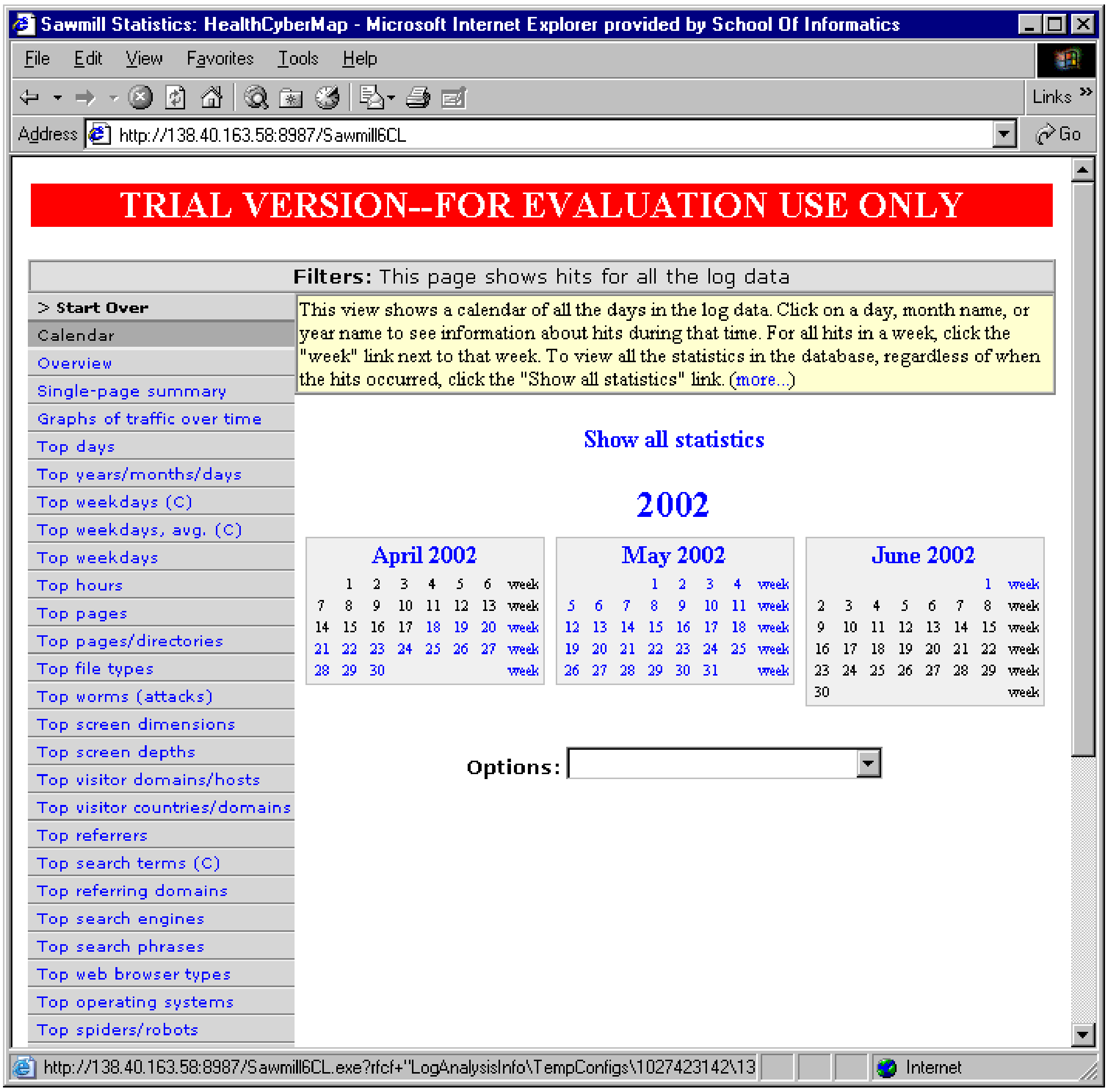


Figure 13.1. Front page of Sawmill's statistics for HealthCyberMap. Note how Sawmill has established itself as a Web server on port 8987 on the local Windows NT 4.0 machine used for this analysis (not HealthCyberMap Web server).

13.1 General Statistics

13.1.1 Overview

An overview of HealthCyberMap server log statistics for the 45-day period of this study is presented in Table 13.1.

Total hits:	26365
Total page views (including map views):	12479
Total unique visitors:	1410
Total bytes transferred:	642.31 Meg
Starting day:	18/Apr/2002
Ending day:	01/Jun/2002
Total days covered:	45
Average hits per day:	585
Average page views per day:	277
Average unique visitors per day:	31
Average bytes transferred per day:	14.27 Meg

Table 13.1. An overview of HealthCyberMap server log statistics for the 45-day period of this study.

13.1.1.1 Notes on the Computation of Total Page Views and Unique Visitors

When a visitor requests a page that uses frames (e.g., HealthCyberMap World Maps <http://healthcybermap.semanticweb.org/world_map/>), the server log records one request for the page itself, plus one additional request for each frame. Log file analysis tools count each of these requests as a separate page view, even though only one page is displayed to the user, resulting in some overcounting.

The number of visitors shown represents *unique* visitors. Sawmill defines visitors to be unique hosts. Unique hosts (or sites) count indicates how many unique IP addresses have made requests to HealthCyberMap server during the reporting period. This does not accurately reflect the actual number of unique individual users (real people or robots) that visited HealthCyberMap, which is impossible to determine with server logs and HTTP protocol alone.

By using unique hosts to determine the number of unique visitors, a visitor is only counted once irrespective of how many visits he/ she has made to the site or how many pages he/ she has browsed during any visit (as long as this visitor's IP address remains constant). However, because subscribers to dial-up Internet services share IP addresses, it is possible for multiple people to connect at different times through the same IP address and thus to be counted as a single visitor. It is also possible for the same person to connect at different times through different IP addresses each counting as a different visitor. And even with a fixed IP address, it is still possible that different users might be using the same computer/ IP address at different times [95]. Things get even more complicated with caches and proxies that many providers now use to improve performance, making many concurrent distinct users/ visits appear as one visitor/ visit.

Using cookies does not improve the situation very much, as different users might share the same computer and appear as a single repeat visitor, and a repeat visitor might delete his/ her cookie file and appear as a different new visitor [95]. Requesting users to create logging accounts with a unique user name/ password combination for each user to access the site can greatly improve results. However, many users are put off by sites requesting them to register before accessing the material they are interested in, and a single person might create multiple accounts (e.g., when they forget their password).

13.1.2 Sessions Summary and Visitor Retention Statistics

Table 13.2 shows a summary of visitor sessions (visits), including at least 59 visits by non-human robot visitors (see later below). A session starts when a visitor enters the site, and ends when that visitor leaves (Figure 11.4). Of all unique HealthCyberMap visitors during the period covered by this analysis (1410), 24.3% (343) could be identified as repeat visitors who have visited HealthCyberMap more than once during that period; 64.7% (1958) of all counted sessions (3025) are attributed to repeat visitors.

Total unique visitors/hosts:	1410
Total sessions:	3025
Total days covered:	45
Average sessions per day:	67
Repeat visitors:	343
Sessions by repeat visitors:	1958
One-time visitors:	1067
Two-time visitors:	151
Three-time visitors:	40
Four-time visitors:	29
Five-time visitors:	19
Six+-time visitors:	104
Average visits per visitor:	2.15
Median visits per visitor:	1
Average session duration :	00:02:48
Sum of all session durations:	5 days, 21:59:58

Table 13.2. Sessions summary and repeat visitor figures for the 45-day period of this study.

Better visitor retention (repeat visitor) figures could be expected if the service is developed past the current pilot stage to offer a wider, more comprehensive coverage of health and clinical topics.

13.2 Activity/ Traffic Statistics

Figure 13.2 shows graphs of the traffic (hits, page/ map views and sessions) over time.

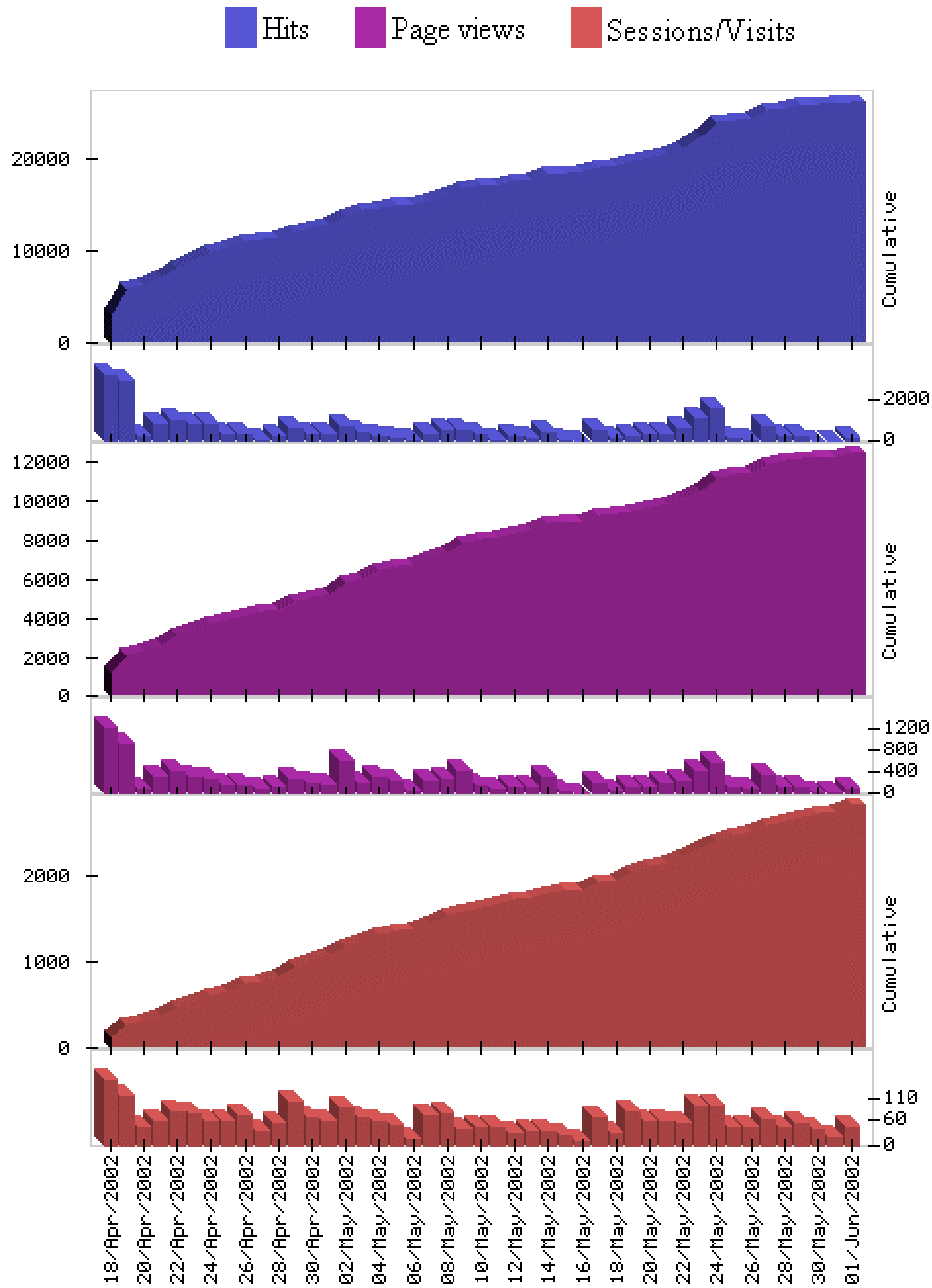
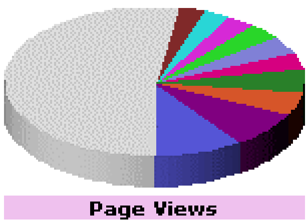


Figure 13.2. Graphs of HealthCyberMap server traffic over time for the 45-day period of this study.

The bar graphs show the traffic for each day, and the increasing area graphs show the total cumulative traffic throughout the 45-day period of the analysis. The overall trend is that traffic and popularity have been growing steadily over the period of the study. Notice the peaks on the 18th and 19th of April 2002 corresponding to the announcement of the launch of HealthCyberMap evaluation by targeted one-time e-mails and in several mailing lists and newsgroups. Some of the troughs in the graphs above are due to known server downtimes (e.g., for scheduled and unscheduled server maintenance tasks by the hosting company).

13.2.1 Top Ten Days

Figure 13.3 shows the top ten days by traffic (page views and unique visitors) for the 45-day period of the study. Again notice the peaks on the 18th and 19th of April 2002 corresponding to the announcement of the launch of HealthCyberMap evaluation.



	Date/time range	Page views	Visitors	Page views bar
1	18/Apr/2002 (Thursday)	1230	155	
2	19/Apr/2002 (Friday)	974	122	
3	02/May/2002 (Thursday)	637	93	
4	24/May/2002 (Friday)	602	99	
5	09/May/2002 (Thursday)	470	43	
6	22/Apr/2002 (Monday)	463	85	
7	23/May/2002 (Thursday)	460	98	
8	27/May/2002 (Monday)	396	65	
9	23/Apr/2002 (Tuesday)	355	78	
10	04/May/2002 (Saturday)	354	63	
	35 other items	6538		
	Average	277	31	
	Total	**12479**		

Figure 13.3. Top ten days by traffic for 45-day period of this study.

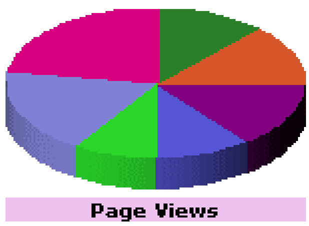
13.2.2 Average Traffic on Each Weekday

Figure 13.4 shows the *average* traffic on each day of the week over the analysed 45-day period. This can be useful for getting an idea of which days are peak days, but the results in our case were partially influenced by the exceptional traffic peaks on Thursday 18th and Friday 19th of April 2002.

13.2.3 Value of Traffic Statistics

The goal of this kind of statistics shown in Figures 13.3 and 13.4 (and also in Figure 13.2) is to help the Web site administrator(s) get a sense of the actual load on the server [104]. This is useful for server diagnostics and planning, and for detecting unusual behaviour and demands that may require planning action (e.g., increasing server bandwidth capacity or reducing the size/ minimising the use of certain file types that consume much bandwidth like large PowerPoint presentations and zipped archives).

Note: Ignoring data from 01/Jun/2002 because that's the latest day in the data, and may not be complete

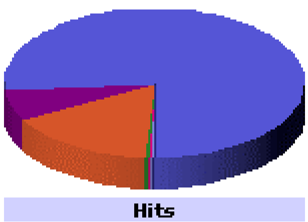


	Day of week (instances)	Page views	Page views (%)	Page views bar
1	Sunday (6)	206	10.72%	
2	Monday (6)	261	13.58%	
3	Tuesday (6)	262	13.63%	
4	Wednesday (6)	226	11.76%	
5	Thursday (7)	438	22.79%	
6	Friday (7)	358	18.63%	
7	Saturday (6)	171	8.90%	
	Total	**1922**	**100%**	

Figure 13.4. Average traffic on each day of the week over the analysed 45-day period.

13.2.4 Server Response and Top File Types Statistics

Sawmill also computes other statistics for bandwidth use, server response/ error codes (e.g., 50x—server errors; Figure 13.5), top hours of the day (by traffic), and top file types (Figure 13.6) that can also help monitoring and tuning server performance.



	Server response	Hits (%)	Hits bar
1	200 (Successful Transfer)	76.62%	
2	404 (Not Found)	6.65%	
3	304 (Not Modified)	15.77%	
4	500 (Internal Server Error)	0.43%	
5	302 (Moved Temporarily)	0.19%	
6	206 (Partial Content)	0.25%	
7	406 (Not Acceptable)	0.06%	
8	405 (Method Not Allowed)	0.01%	
9	403 (Forbidden)	0.01%	
10	401 (Unauthorized)	0.00%	
11	501 (Not Implemented)	0.00%	
	Total	**100%**	

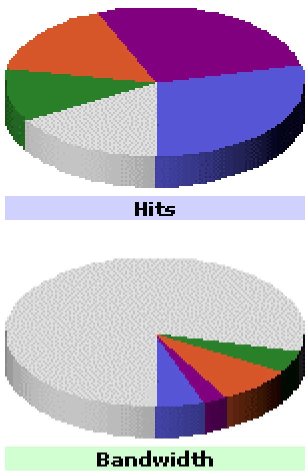
Figure 13.5. HealthCyberMap server response to hits.

The 404 responses shown in Figure 13.5 are due to users trying to access some of HealthCyberMap older content via outdated search engine/ referring sites links or bookmarks dating back to the period before the launch of formative evaluation on 18

April 2002, during which the site was undergoing continuous development and restructuring (files deleted, moved or renamed).

The 304 (not modified since last request) response occurs when the client browser performs a conditional GET request and finds that the requested file has not been modified since the date on the file in the browser's cache. In this case the copy in the browser's cache is displayed and the server does not need to retransmit the file. It should be noted that 304 responses do not represent all cached hits that might have occurred, and there might be more page views in reality than what has been recorded in server logs. A frequently requested document may be served directly from the cache without the server having any record of it having been viewed again. The server records instances only when the cached document is compared with the server version for currency; if, or how often, this occurs depends on browser settings [116].

If transfer of a large file is commonly interrupted/ incomplete (also recorded in the log file as an error), then one can ascertain that visitors are not patient enough to view the file, and something should be done to reduce its size [95].



	File type	Hits	Hits (%)	Bandwidth	Bandwidth (%)	Hits bar / Bandwidth bar
1	HTM	7249	28.80%	35.31M	5.50%	
2	GIF	6924	27.51%	16.33M	2.54%	
3	JPG	3966	15.76%	51.80M	8.07%	
4	ASP	2909	11.56%	31.27M	4.87%	
	other items	4124	16.38%	507.31M	79.02%	
	Total	**25172**	**100%**	**642.02M**	**100%**	

Figure 13.6. Top file types accessed by HealthCyberMap visitors during the 45-day period of this study (by number of hits).

There are other file types that are not shown in detail in Figure 13.6, which received fewer hits but consumed much bandwidth (grey zones in the pie charts in Figure 13.6). These include some PowerPoint presentations and zipped archives that are available for downloading from HealthCyberMap Web site like "HealthCyberMap Quick Tour", a zipped PowerPoint presentation that introduces the service to new users and the scientific community (4.84 MB in size—
<http://healthcybermap.semanticweb.org/mnkb_HCM_poster_20020510.zip>).

13.3 HealthCyberMap Interfaces by Popularity

To compute the most popular HealthCyberMap interfaces, the number of unique visitors accessing each interface was used rather than the number of page/ map views recorded for each interface, since the number of pages that form each interface is very variable (Figure 13.7). The unique visitor values for BodyViewer topical maps and the textual index of resources according to ICD-9-CM categories were aggregated into a single value since the two interfaces in the evaluated pilot implementation shared the same dynamic query pages (*.asp). Please note that the Semantic Subject Search Engine is also indirectly accessible from within the World maps and BodyViewer interfaces (when users click the "Find all resources having the same primary subject as this one" link).

BodyViewer topical maps received the highest number of unique visitors, while the classification of Web resources according to resource language scored lowest (cf. formative evaluation questionnaire results—Chapter 12).

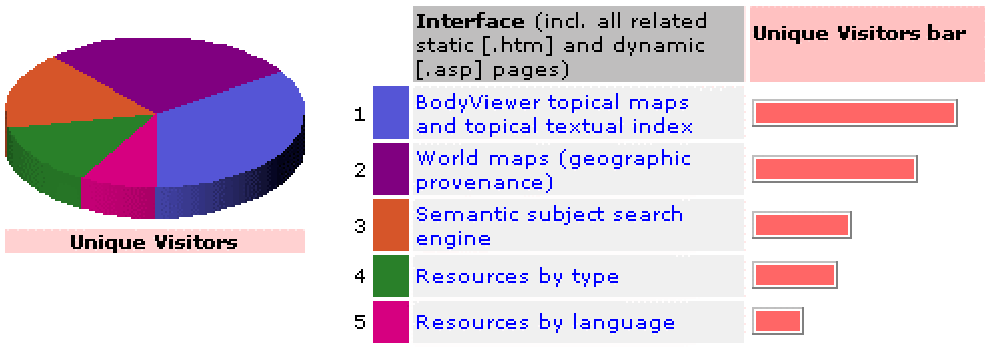


Figure 13.7. HealthCyberMap interfaces by popularity (number of unique visitors accessing each interface).

It should be noted that it is the most often visited pages that are also cached and retrieved (from cache) more frequently, and because not all cached page views are logged in server logs [116], the most popular pages might end up being underestimated.

13.4 Questionnaire Page Access Statistics

HealthCyberMap Formative Evaluation Questionnaire page (<http://healthcybermap.semanticweb.org/questionnaire.asp>) has been viewed 328 times (page views) by 190 unique visitors in 208 visits (sessions). The average time spent on the questionnaire page was 3 minutes 36 seconds per session. The 328 page views included 35 "POST /questionnaire.asp" HTTP requests corresponding to the 35 filled-in and submitted questionnaires (see Chapter 12). To fill-in the questionnaire, a user first requests the questionnaire form ("GET /questionnaire.asp" HTTP request—counts as one page view), then selects/ types the answers and clicks the "Save" button at the bottom of the form to post answers ("POST /questionnaire.asp" HTTP request). When HealthCyberMap server receives a "POST /questionnaire.asp" HTTP request, it saves the filled-in questionnaire to the evaluation database on the server and sends to the user another questionnaire.asp page with a "Thank you" message instead of the questionnaire form fields (counts as a second page view).

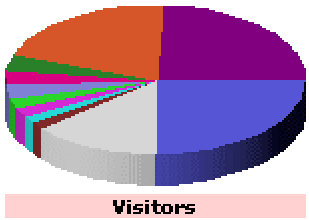
13.5 Visitor Information

13.5.1 Geographical Provenance of HealthCyberMap Visitors

Using Sawmill 6.3.8, HealthCyberMap visitors during the 45-day period of this study could be traced to more than 50 countries (Table 13.3). The most active countries/ domains (top ten) accessing HealthCyberMap (non-human robot visitors included, e.g., *.Fastsearch.net) are presented in Figure 13.8 sorted by number of unique visitors from each country/ domain and in Figure 13.9 sorted by number of page/ map views for each country/ domain. More sophisticated server log analysis tools utilise their own database to resolve visitors' IP addresses into more recognisable and useful provenance data (see for example, WebTrends' GeoTrends Database: <http://www.netiq.com/support/wrc/geotrends.asp>).

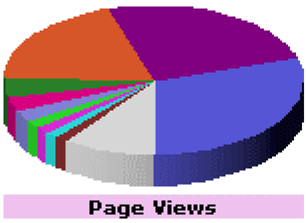
Australia (au)	Finland (fi)	Malaysia (my)	Slovak Republic (sk)
Austria (at)	France (fr)	Mexico (mx)	South Korea (kr)
Barbados (bb)	Germany (de)	Military (mil, United States)	Spain (es)
Belgium (be)	Government (gov, United States)	Netherlands (nl)	Sweden (se)
Brazil (br)	Greece (gr)	Network (net)	Switzerland (ch)
Canada (ca)	Hong Kong (hk)	New Zealand (nz)	Taiwan (tw)
Commercial (com)	India (in)	Non-profit (org)	Thailand (th)
Croatia (hr)	Indonesia (id)	Poland (pl)	Turkey (tr)
Czech Republic (cz)	International (int) [WHO]	Portugal (pt)	Ukraine (ua)
Denmark (dk)	Ireland (ie)	Romania (ro)	United Kingdom (uk)
Dominican Republic (do)	Israel (il)	Russia (ru)	United States (us)
Educational (edu, United States)	Italy (it)	Saudi Arabia (sa)	Unresolved IP address (354 visitors)
Egypt (eg)	Japan (jp)	Singapore (sg)	Venezuela (ve)

Table 13.3. A list of countries/ domains that have accessed HealthCyberMap during the 45-day period of this study.



	Domain description	Visitors	Visitors bar
1	Commercial (com)	370	
2	IP Address	354	
3	Network (net)	299	
4	Educational (edu)	51	
5	United Kingdom (uk)	45	
6	France (fr)	42	
7	Germany (de)	34	
8	Canada (ca)	31	
9	Belgium (be)	23	
10	Japan (jp)	21	
	41 other items		

Figure 13.8. Most active countries/ domains (top ten) accessing HealthCyberMap during the 45-day period of this study sorted by number of unique visitors from each country/ domain (non-human robot visitors included).



	Domain description	Page views	Page views bar
1	IP Address	3727	
2	Network (net)	3122	
3	Commercial (com)	2405	
4	United Kingdom (uk)	565	
5	Educational (edu)	370	
6	Japan (jp)	340	
7	Canada (ca)	199	
8	France (fr)	172	
9	Netherlands (nl)	171	
10	Germany (de)	171	
	41 other items		

Figure 13.9. Most active countries/ domains (top ten) accessing HealthCyberMap during the 45-day period of this study sorted by number of page/ map views for each country/ domain (page views by non-human robot visitors included).

13.5.1.1 Value of Determining User Location

Determining the most active countries accessing a site can help prioritising development plans of new interfaces and content in other languages, e.g., French language content and user interface, besides English (giving higher development priority to languages of most active countries). This becomes especially important when development resources (human and financial) are scarce, and to ensure rapid and sustained leadership over competitor services that might be also exploiting the potential of multilingual interfaces and content.

The author has noticed that language and country-specific pages like /es.asp (resources in Spanish, with a Spanish HealthCyberMap user interface) and /spain.asp (resources from Spain, not necessarily in Spanish) and /italy.asp (resources from Italy, not necessarily in Italian) were mostly accessed by visitors from corresponding countries (Spain and Italy in this example). Nearly all Italian hosts who visited HealthCyberMap during the period of this study, e.g., guide.unipv.it, accessed /italy.asp during their visits. The same applies to Spanish hosts, e.g., 213-96-80-80.uc.nombres.ttd.es, who were also very keen to visit /es.asp and /spain.asp. It seems that users have a tendency to explore resources in their native language and/ or country if available.

13.5.2 Top Web Browsers and Operating Systems

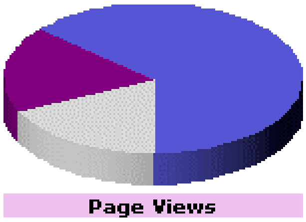
Table 13.4 presents the top Web browser types/ versions accessing HealthCyberMap by number of unique visitors (robots excluded).

Web browser type/version	Unique human visitors using it
Internet Explorer (all versions)	932—about 63.4% of all human visitors
Internet Explorer/5.x	562—60.3% of those using Internet Explorer
Internet Explorer/6.x	362—38.8% of those using Internet Explorer
Netscape Navigator (all versions)	247—about 16.8% of all human visitors

Table 13.4. Top Web browser types/ versions accessing HealthCyberMap during the 45-day period of this study by number of unique visitors (robots excluded).

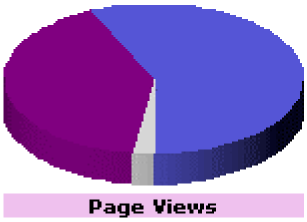
Figure 13.10 shows the top Web browser types accessing HealthCyberMap by page views (excluding 2009 page views by spiders/ robots—see Figure 13.14). Microsoft Internet Explorer (all versions aggregated) ranks first (6656 page views—63.57%) followed by Netscape Navigator (all versions aggregated; 1950 page views—18.62%).

Figure 13.11 shows the version distribution of Microsoft Internet Explorer browsers accessing HealthCyberMap. More than 97% of Internet Explorer browsers accessing HealthCyberMap during the period of this analysis were either version 5.x (3782 page views—56.82%) or version 6.x (2727 page views—40.97%).



	Web browser	Page views
1	Internet Explorer/	6656
2	Netscape Navigator/	1950
	23 other items	1864
	Total	**10470**

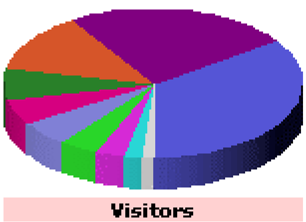
Figure 13.10. Top Web browser types accessing HealthCyberMap during the 45-day period of this study sorted by page views (excluding 2009 page views by spiders/ robots).



	Web browser	Page views
1	Internet Explorer/5.	3782
2	Internet Explorer/6.	2727
	4 other items	147
	Total	**6656**

Figure 13.11. Version distribution of Microsoft Internet Explorer browsers accessing HealthCyberMap during the 45-day period of this study.

Figure 13.12 shows the top ten operating systems accessing HealthCyberMap during the analysed 45-day period sorted by the number of unique visitors using each operating system. Spider/ robot visitors and visitors with unspecified/ undetected operating system were excluded from the list. Only one visitor used WebTV platform to access the site during the same period (not shown in Figure 13.12).



	Operating system	Visitors	Visitors bar
1	Windows 2000	439	
2	Windows 98	311	
3	Windows XP	162	
4	Windows ME	84	
5	Windows NT 4.0	72	
6	Windows NT	70	
7	Windows 95	54	
8	Power Macintosh	38	
9	Linux	25	
10	SunOS	15	

Figure 13.12. Top ten operating systems accessing HealthCyberMap during the 45-day period of this study sorted by the number of unique visitors using each operating system.

13.5.2.1 Note on WebTV

This pilot HealthCyberMap service is not fully compatible with WebTV because of the very limited support of imagemaps in the latter. Only rectangular hotspots are supported in WebTV (see <http://developer.msntv.com/Develop/seamimgmap.asp>); complex shapes, e.g., the outline of Canada on the map in Figure 13.13 are not supported. However, all equivalent textual interfaces (e.g., the country drop-down list shown in Figure 13.13) work as intended and can be used as an alternative way to access the same material accessible via the hypermaps.

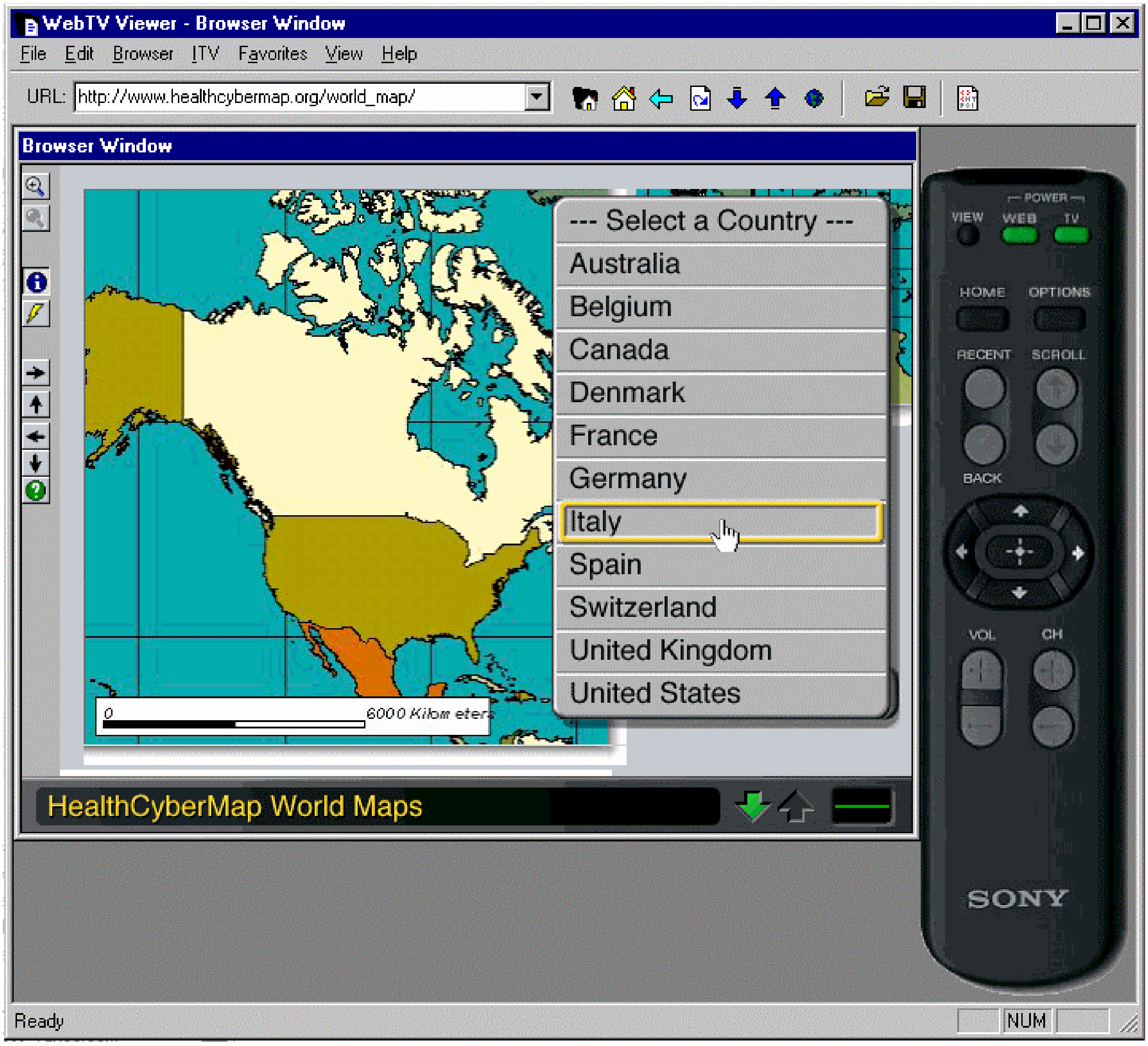


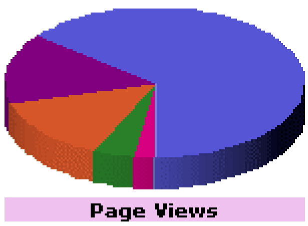
Figure 13.13. HealthCyberMap in WebTV viewer (this viewer can be downloaded from <http://developer.msntv.com/Tools/WebTVVwr.asp>).

13.5.2.2 Value of Browser and Operating System Statistics

Browser/ operating system server log statistics could be important in optimising and prioritising development/ testing decisions and efforts, especially with limited development resources. For example, given the results shown in Figures 13.10, 13.11 and 13.12, if one was to focus development efforts on the few most prevalent platforms/ browsers, these would be Windows 2000/XP and Windows 98/Me operating systems, and Internet Explorer browser version 5.x or later. From the questionnaire analysis (Chapter 12), it was also learned that developing for a Web browser window size of a little less than 800 by 600 pixels remains the safest choice today to avoid the need for scrolling (544 by 372 pixels if specifically targeting WebTV).

13.5.3 Non-human Visitors

Figure 13.14 shows the top spiders (robots) that hit HealthCyberMap site during the 45-day period of this analysis. Spiders are programs that automatically crawl or walk through a site, indexing site contents for search engines. Spider hits are not "human" hits; they are automatically generated and do not represent human traffic. Of the total 12479 page views served by HealthCyberMap server from 18 April 2002 to 1 June 2002, 2009 page views (16%) were served to robots and the rest to humans. The top two spiders that crawled and indexed HealthCyberMap during the period of this study were FAST-WebCrawler (<http://www.fastsearch.net>) and Googlebot, Google's Web-crawling robot (<http://www.google.com/bot.html>).



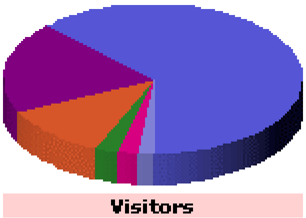
	Spider	Page views
1	FAST-WebCrawler	1293
2	Googlebot	301
3	larbin	280
4	Scooter	88
5	Inktomi Slurp	45
6	Northern Light Gulliver	2
	Total	2009

Figure 13.14. Top spiders (robots) that hit HealthCyberMap site during the 45-day period of this analysis

13.6 Referring Search Engines and Sites

Figure 13.15 shows the Web search engines people used to find HealthCyberMap site. Google and Yahoo! (also powered by Google) ranked top, bringing a total of 242 unique human visitors to HealthCyberMap.

Top search phrases included "multiple sclerosis", "urinary system diseases", "diseases of the reproductive system", and "diseases of the circulatory system". It could be very useful in the future to also log what visitors search for *within* HealthCyberMap, e.g., keywords they type in the search boxes of HealthCyberMap Semantic Subject Search Engine, to gain insight into what users want (and whether they found it using HealthCyberMap or not). This can help improving the service and in planning future content [117].



	Search engine	Visitors	Visitors bar
1	Google	185	
2	Yahoo	57	
3	AskJeeves	35	
4	Lycos	7	
5	MSN Search	6	
6	Netscape Search	5	

Figure 13.15. Web search engines people used to find HealthCyberMap site.

13.6.1 Value of Determining Referring Search Engines and Sites

By determining which sites are referring visitors to HealthCyberMap, one can determine which online advertising activities are most successful at driving traffic to the service. For example, only four visitors during the 45-day period of this study were referred to HealthCyberMap by following a newsgroup link, e.g., <http://groups.google.com/groups?dq=&hl=en&group=**sci.med.informatics**&selm=a9moqm%24jr7%241%40canard.ulcc.ac.uk> (message advertising the launch of HealthCyberMap Formative Evaluation Questionnaire). Six visitors were referred to HealthCyberMap during the same period from a Yahoo! Groups message announcing HealthCyberMap, e.g., <http://groups.yahoo.com/group/semanticweb/message/479>. (This might not be the actual number of users who came to HealthCyberMap via newsgroups/ Yahoo! Groups links, as other users might have copied and pasted HealthCyberMap addresses from the announcement messages posted to newsgroups.) On the other hand, search engines brought hundreds of users to HealthCyberMap during the same period. This is largely consistent with questionnaire findings (Chapter 12—Figure 12.25).

13.7 Navigation Paths and Top Entry Page

Regarding the navigation paths taken by visitors through the site, Sawmill showed no particular, recurring path pattern. One could say that nearly each of the 3025 recorded sessions represented a unique navigation path through the site. (The actual paths cannot be retraced accurately given the browser and proxy cache problems outlined above.) This finding was expected given the widely varied profiles of users who have accessed HealthCyberMap during the period of this analysis (user roles, backgrounds, geographical provenance, etc.), and also the different clinical questions and visit purposes they should have had in their minds while using the service. However, 1409 sessions (46.58%) started at HealthCyberMap main front page (<http://healthcybermap.semanticweb.org/Default.htm>—top entry page).

Server logs also (at least theoretically) record the last page viewed during a given user session, but there are two powerful reasons why it might not be the actual exit page. First, the last page viewed might have been displayed from cache and not recorded in any way in the log file. Second, the user might have navigated to an external site or left his/ her workstation for a period of time that exceeds what Sawmill software regards as a session (30 minutes in our analysis) before returning back to HealthCyberMap.

13.8 Measuring User Interest and Content Usefulness Using Server Logs: Is It Possible?

Counting page views and hits alone is not an adequate measure of user interest. Consider a site where one has to go through pages A, B, C and D to reach page E. Server logs will show just as many page views for pages A through D as there are for page E by visitors who actually wanted to access only the latter. What is needed is a path independent measure of user interest [115].

Sawmill computed an average session duration of 2 minutes 48 seconds during the 45-day period of this analysis. (Some sessions were nearly two hours long, while others lasted only few seconds.) Sawmill can also calculate the average time spent on a given page per session. In theory, these figures could help determining which parts of the service were found most interesting/ useful to visitors. In fact, back in 1996, Fuller and de Graaff [115] suggested using the length of time users spend looking at a page as an estimate of user interest in that page (time-based metric of user interest).

Medians should be much less subject to variability than means, and hence provide more reliable estimates [95], but unfortunately, Sawmill does not compute any median times (only averages). However, even when medians are used, there remain many limitations to any time-based metric of user interest or content usefulness (based on standard server logs):

- Visits by robots tend to distort these figures (a robot is a very fast visitor that can crawl hundreds of pages in a very short duration);
- A visitor might have been accessing other HealthCyberMap pages from cache between two "successive" page requests that have been logged in HealthCyberMap server log with nothing else recorded in between;
- The time to download the same page (after a GET request has been logged in server logs) can vary widely depending on server and network load (time of the day), and user's connection. Server logs only record the time when a data transmission was initiated, but not the time when the transfer was completed [95];
- A long time spent on a page might have been actually spent doing something else unrelated to the page in question and which is not logged in HealthCyberMap server logs, e.g., answering a telephone call, getting a cup of coffee, or browsing an external link from that page (e.g., clicking an external resource link in HealthCyberMap query results window);
- In case of Web search engines, directories and portals to external resources (like HealthCyberMap), a short visit duration or short time spent on a search results page could be equally interpreted as indicators of a successful service (users have quickly found/ located what they have been looking for in just few clicks); and
- A properly designed page taking into consideration human factors issues related to navigation and ease of comprehension could take *less* time to browse, while remaining very interesting and useful to readers.

The author believes that a time-based metric of user interest or content usefulness based on standard server logs is not adequate. We need to survey users (or a representative sample of them) if we are to clearly determine whether or not they found the information presented to them useful and found all what they wanted (see Chapter 12: Analysis Results of HealthCyberMap Formative Evaluation Questionnaire).

Another solution has been proposed by Scholtz et al [118] who developed WebVIP (Visual Instrumenter Program), a tool for testing Web sites that requires tracking code to be added to all site pages, and which they claim overcomes many of the limitations of standard server log file analysis.

13.9 Conclusion

Sawmill 6.3.8 was used to analyse HealthCyberMap server log files for the 45-day period beginning 18 April 2002 and ending 1 June 2002. The estimated total number of page views during this period is 12479, of them 16% were served to robot visitors. Of 1410 estimated unique HealthCyberMap visitors (humans and robots) during the period covered by this analysis, 24.3% were identified as repeat visitors and contributed collectively to 64.7% of the estimated 3025 visits that took place during the same period. The overall trend is that traffic and popularity have been growing steadily over the analysed period, with adequate server performance throughout the same period. Better visitor retention figures could be expected if the service is developed past the current pilot stage to offer a wider, more comprehensive coverage of health and clinical topics and more functional Web interfaces.

BodyViewer topical maps received the highest number of unique visitors, while the classification of Web resources according to resource language scored lowest. To be most useful, the classification of resources by language should be transparently integrated into other interfaces in any future implementation of the service (as part of the proposed content customisation).

HealthCyberMap visitors during the 45-day period of this study could be traced to more than 50 countries around the world. More sophisticated server log analysis tools are needed that utilise their own database to resolve more visitors' IP addresses into more recognisable and useful provenance data. Determining the most active countries accessing a site can help prioritising development plans of new interfaces and content in other languages. This in turn will ensure rapid and sustained leadership over any competitor services that might be also exploiting the potential of multilingual interfaces and content. Interestingly, the author noticed from HealthCyberMap logs that language and country-specific pages were mostly accessed by visitors from corresponding countries. It seems that users have a tendency to explore resources in their native language and/ or country if available.

Based on the top Web browsers and operating systems accessing HealthCyberMap during the period of this analysis, if one is to focus future development efforts on the few most prevalent platforms/ browsers, these would be Windows 2000/XP and Windows 98/Me operating systems, and Internet Explorer browser version 5.x or later. Moreover, from the questionnaire analysis it was also learned that developing for a Web browser window size of a little less than 800 by 600 pixels remains the safest choice today to avoid the need for scrolling.

Google and Yahoo! (also powered by Google) ranked top among referring search engines, newsgroups/Yahoo! Groups and external sites, bringing an estimated total of 242 unique human visitors to HealthCyberMap.

Sawmill showed no particular recurring navigation path pattern taken by visitors through the site, though nearly half of all recorded visitor sessions started at HealthCyberMap main front page.

Counting page views and hits alone is not an accurate measure of user interest. The author believes that a time-based metric of user interest or content usefulness based on standard server logs is also not adequate because of browser/ proxy caching effects, network latency, and other reasons. We need to survey a representative sample of service users if we are to clearly determine whether or not they found the information presented to them useful and were able to easily find all what they wanted.

Part V: Future Work, Discussion and Recommendations

14 Future Direction: HealthCyberMap as a Customisable, Location-based Medical/ Health Information Service

14.1 Introduction

The concept that location can influence health is well known in medicine. Certain diseases tend to occur in some places and not others. Health information needs and services also vary with location. In fact, different places on Earth are usually associated with different profiles that can also change with time: physical, biological, environmental, economic, social, cultural, and sometimes even spiritual profiles, that do affect and are affected by health, disease, and healthcare [47].

Caregivers need to know not only the history of patients they treat but also information about the social and environmental context within which those patients live. The Internet offers a wealth of health information resources that can answer most of the knowledge needs of clinicians and their patients, and the public in general, but also carries with it the risk of overloading them with unnecessary information. A big challenge remains to find and push location-specific knowledge (e.g., local disease rates and guidelines—Table 14.1) to users based on their location and needs [119].

Examples of Location-specific Medical/ Health Information
- Local disease rates and information*, maps and guidelines; - Targeted health education; - Addresses of local healthcare facilities; - Local health news; - Local health risks and hazards; - Travellers' health information; - Local drugs/ drug trade names and prices (in local currency); - Information whose digital distribution rights are limited to some location(s); - In addition to serving up content (and interface) in language(s) relevant to the viewer's location.
* For example, the most common cause of splenomegaly in Kenya is malaria, while in the UK the most common causes of Slenomegaly are cytomegalovirus infection and toxoplasmosis.

Table 14.1. Examples of location-specific medical/ health information.

Location-based information means information that is *immediately relevant*, which is the essence of the Semantic Web. Research literature discussing health and healthcare-specific potentials and applications of location-based services is currently very scarce (as of August 2002).

14.1.1 The Potential and Applications of Location-based Services in Health and Healthcare

Different user (device) localisation technologies exist today that can locate in real time mobile (wireless) and desktop Internet users to their country, town or city of access, and even to their exact location on Earth (with an accuracy of metres) depending on the technology used. Geobytes, Inc. provides one such technology (see below). Departing from their generic list of applications of user localisation technologies [120], one can think about the following health and healthcare-specific scenarios:

- Geographical customisation of Web services, Web sites, and e-mail newsletters (personalisation by location), for example to serve up language relevant to the viewer's location, to cite only drugs, drug trade names and their prices as found in the viewer's country, and to deliver local health news, local weather and air quality maps, travellers' health information, and other health content specific to the viewer's location.
- Geographically targeted banner and email marketing and advertising. This can prove very useful to commercial online health service providers and pharmaceutical companies.
- Web site traffic analysis to determine the geographical provenance of health sites' visitors—see for example Figures 13.8 and 13.9. This information can then be used to tailor site content to match the needs of actual visitors and the characteristics of their location, as well as to refine site marketing strategies and monitor advertising campaigns (especially in case of commercial online health services).
- Fraud detection and user authentication by confirming that users are actually present where they claim they are.
- Digital rights management to ensure compliance with distribution rights of copyrighted health information and media.

14.1.2 Mobile Location-based Services and their Applications in Health and Healthcare

Besides transmitting real-time information on their precise location, next generation mobile phones currently entering the market (e.g., Microsoft Smartphone—<http://www.microsoft.com/mobile/smartphone/default.asp>) also provide users with wireless Internet access [121]. This makes them very versatile and powerful devices.

Mobile health applications require an understanding of where consumers are, where they have been and where they are going. Wireless mobile devices can continually transmit device (user) location to such applications, which must then make use of the transmitted information in a sensible way.

According to Berkowitz and Lopez [122], *location-awareness* refers to applications or services that make use of location information provided by suitable devices or software (location need not be the primary purpose of the application or service), while *location-sensitivity* refers to location-enabled mobile devices that can be used by location-aware applications and services such as mobile phones, personal digital assistants and pagers. Such devices rely on GPS (Global Positioning System)/ mobile phone-related technologies.

One can also add to the second category of location-sensitive tools any software whose primary function is to detect the locations of "static" Internet users (e.g., in the office or at home) based on their IP (Internet Protocol) address as demonstrated in HealthCyberMap (IP geolocation [123, 124]—see below).

Following are some health and healthcare application examples of mobile location-based technologies.

- Health information and service providers can (if needed) react immediately to the changed location of a mobile user by delivering personalised timely information and services for his/ her new roaming region [122].
- An online healthcare facility locator service can assist users in finding the nearest hospitals or clinics based on their location and health needs, and even provide them with driving directions and real-time traffic information [122]. For example, a mobile patient checking on the availability of a dental clinic in a given city might access geocoding services that identify the location of the patient and nearest clinics, and would cull data from real-time booking services to check for clinics' working hours and book a suitable appointment, and from driving directions and real-time traffic information services to route the patient to the clinic.
- Helping ambulance and rescue teams precisely and quickly locate and track people who are in a medical emergency, injured, or lost, and also for ambulance fleet management. New FCC rules (Federal Communications Commission—<http://www.fcc.gov/911/enhanced/>) mean that GPS receivers will very soon become an integral part of all mobile phones.
- Mobile patient monitoring and automated emergency calls with very precise information on patient location if the system detects any medical problem requiring intervention. It should be also possible to transfer all monitoring data to a secure Web-based patient record accessible from anywhere in the world [121].

Digital Angel Technology (<http://www.digitalangel.net>) belongs to this category of services.

- Sampling real-time data on environmental exposure to irritants and pollutants with information on the individual's physical reactions in different situations and locations. One practical application of such exercise could be to send alarm messages to mobile asthmatic patients if they start moving to low air quality locations within large cities [121].

14.1.3 Important Issues Related to Location-based Services

User devices used to access a service might change with location, e.g., a desktop or laptop computer at home or in the clinic and a more limited mobile device on the road. The drawback of the small size of mobile devices is that display is considerably smaller and input much more difficult (e.g., no full-scale keyboard). Location-based services must take into consideration the input and output characteristics of different user devices by carefully choosing, personalising and formatting the content to display on such devices [122].

Location capability also poses service providers with the challenge of responsibly handling consumers' personal privacy, especially if they use cookies to track their users [121, 122, 125]. Services should publish their Privacy Policy and respect consumers' choices in this regard. See for example HealthCyberMap's Privacy Statement <http://healthcybermap.semanticweb.org/privacy.htm>.

14.2 HealthCyberMap First Steps toward Being a Customisable Location-based Medical/ Health Information Service

HealthCyberMap (<http://healthcybermap.semanticweb.org>) can be also thought of a geographic information system (GIS)-driven Web-based directory service of health resources that explores new means to organise and present health and healthcare related information on the Internet based on consumer and provider locations. It tries to develop an online information service that should ultimately allow better presentation of the distribution of health and healthcare needs and Internet resources answering them across a geographical area. The service is aimed to provide better support for informed decision-making by the public, patients and their caregivers.

14.2.1 HealthCyberMap Geographical Mapping of Medical/ Health Information Resources

Location-based services draw heavily on GIS and geoinformatics [121] as illustrated in the "dental clinic" example above. Another novel use of GIS is to map conceptual spaces occupied by collections of medical/ health information resources as demonstrated in HealthCyberMap (Chapter 9). Besides mapping the semantic and non-geographical aspects (e.g., subject or topic) of these resources using suitable spatial metaphors (e.g., human body maps—<http://healthcybermap.semanticweb.org/bodyviewer/>), HealthCyberMap also maps some geographical aspects of these resources—Figure 9.1. The resultant maps can be classified as conceptual information space maps and can be used as navigational aid for browsing mapped resources [39].

Two geographical aspects of health information resources are considered in HealthCyberMap, namely coverage and provenance. Coverage deals with the spatial extent or scope of the content of a given resource (this aspect is important if we want to be able to select resources that are appropriate for a particular user location). Provenance refers to the geographical location (city/ country) of a resource publisher or author(s), whichever is more relevant, and can be very useful as an index to information resources and for some kinds of studies [40, 41]—Figure 9.1. Coverage and provenance are not necessarily the same as the physical location of hosting servers, e.g., a British Patient Support Group offering UK-specific advice on some condition might have its Web site hosted on a server in Arizona, US; however, the site remains more relevant to patients and their caregivers living in the UK.

14.2.2 Technologies for Detecting User Location

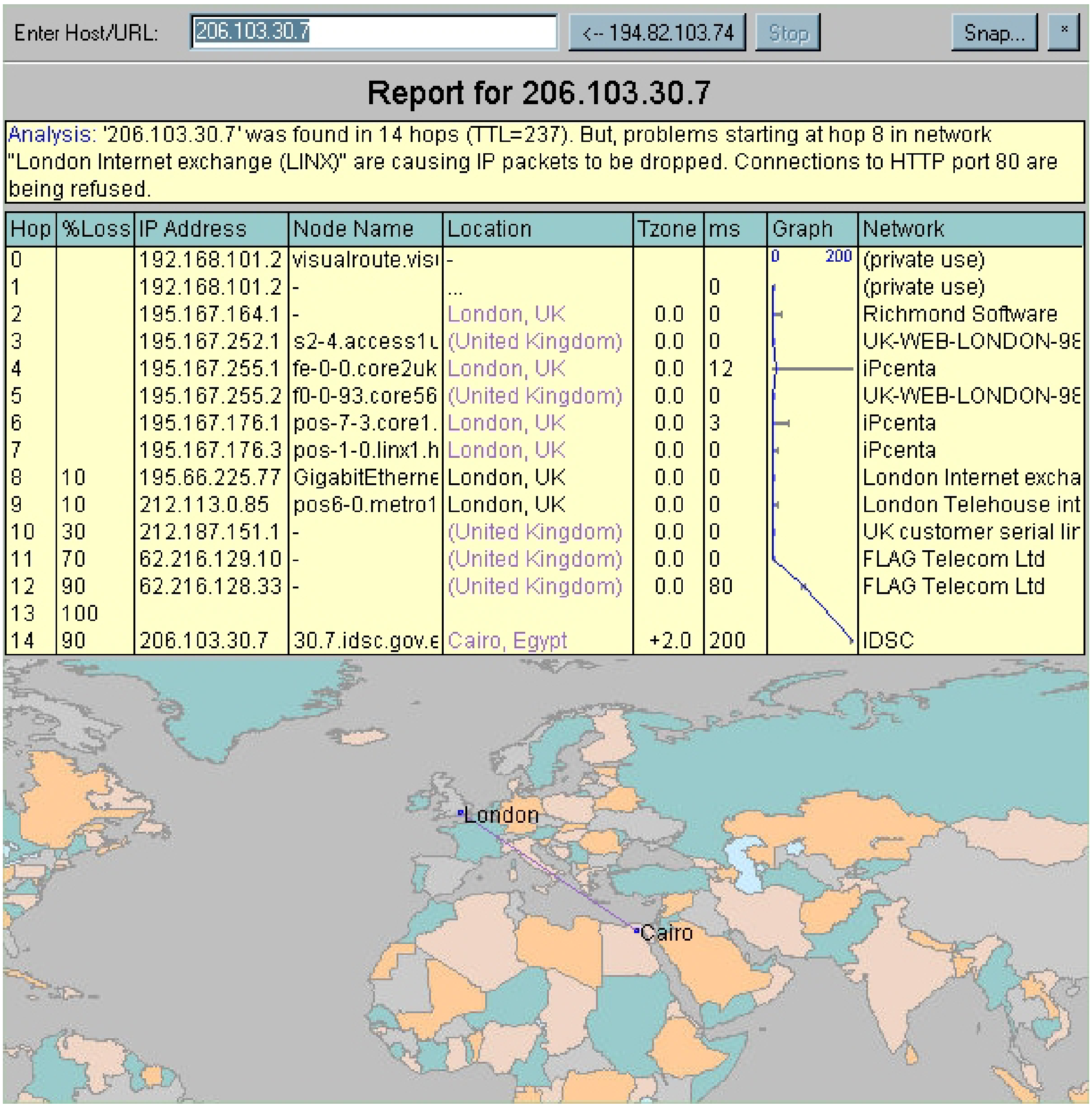
To begin developing a location-based service we have to adopt a method for detecting user location. A user's location can be collected through a Web form that the user fills (can be also used to collect other user information/ preferences besides location), or automatically detected based on the user's IP address (IP geolocation—[123, 124]). Tools exist that allow mapping user IP address to a coarse location (city/ country) on Earth (this should be enough for basic geographical customisation purposes of an online information service like HealthCyberMap). These tools include:

- Geo-IP (<http://geo-ip.com/>);
- Quova GeoPoint (<http://www.quova.com>);
- Geobytes' GeoSelect technology (<http://www.geoselect.com>); and

- VisualRoute, a utility produced by Visualware, Inc. (<http://www.visualware.com>)—Figure 14.1.

More information and a demonstration of the last two tools can be found at: <http://healthcybermap.semanticweb.org/ip.htm>.

IP address to city/ country mapping is not always successful [123]. Available tools sometimes fail to map a user IP address to a city/ country or map it to a wrong location depending on the accuracy and coverage of their underlying lookup databases.



Hop	%Loss	IP Address	Node Name	Location	Tzone	ms	Graph	Network
0		192.168.101.2	visualroute.visu	-			0 200	(private use)
1		192.168.101.2	-	...		0		(private use)
2		195.167.164.1	-	London, UK	0.0	0		Richmond Software
3		195.167.252.1	s2-4.access1u	(United Kingdom)	0.0	0		UK-WEB-LONDON-98
4		195.167.255.1	fe-0-0.core2uk	London, UK	0.0	12		iPcenta
5		195.167.255.2	f0-0-93.core56	(United Kingdom)	0.0	0		UK-WEB-LONDON-98
6		195.167.176.1	pos-7-3.core1.	London, UK	0.0	3		iPcenta
7		195.167.176.3	pos-1-0.linx1.h	London, UK	0.0	0		iPcenta
8	10	195.66.225.77	GigabitEtherne	London, UK	0.0	0		London Internet excha
9	10	212.113.0.85	pos6-0.metro1	London, UK	0.0	0		London Telehouse int
10	30	212.187.151.1	-	(United Kingdom)	0.0	0		UK customer serial lir
11	70	62.216.129.10	-	(United Kingdom)	0.0	0		FLAG Telecom Ltd
12	90	62.216.128.33	-	(United Kingdom)	0.0	80		FLAG Telecom Ltd
13	100							
14	90	206.103.30.7	30.7.idsc.gov.e	Cairo, Egypt	+2.0	200		IDSC

Figure 14.1. Screenshot of VisualRoute, a graphical traceroute utility written in Java by Visualware, Inc. (<http://www.visualware.com/>).

In wireless Internet, mobile devices can continually send their very precise location, e.g., via assisted global positioning services (AGPS—[125]), which can be used in more sophisticated mobile location-based services. However, as outlined above privacy issues must be observed whenever any kind of user location information is gathered (e.g., when a user does not want to reveal his/ her location or accept cookies, this should be respected).

14.2.3 HealthCyberMap Location-based Customisation Possibilities

It should be possible, at least theoretically, to customise (personalise) HealthCyberMap based on a user's geographical location as determined by his/ her IP address used to access HealthCyberMap server. Moreover, HealthCyberMap users should be allowed to override this and manually set their preferences (including personal preferences unrelated to location) following "My Yahoo!" (<http://my.yahoo.com>) and BBC News and Sport (<http://news.bbc.co.uk>—Figure 14.2) examples. A user input form can be used to capture (and store) a user's profile. User's descriptors in this profile can then be used to tailor the content delivered to that user according to some predefined content selection model or rules.

Figure 14.2. BBC News and Sport location-based content customisation.

Two main location-based customisation categories are possible:

1. Language and interface customisation:
    a. Setting HealthCyberMap Web interface language to match user's location language (if a multilingual version of HealthCyberMap is fully developed in the future); and
    b. Retrieving/ giving more importance to Web resources in user's location language (<http://healthcybermap.semanticweb.org/language.htm>)—Figure 9.25. However, it should be noted that some users might move from their native country to another country, e.g., from the UK to France, either temporarily or permanently. It is not always the case that such users will want the language of HealthCyberMap interface and retrieved resources to be changed to reflect their new location, e.g., from English to French.
2. Content customisation:

    Customisation should also (ideally) address any location-specific information needs and match these needs to suitable online resources covering the concerned location and its known health and healthcare makeup (not just its language). This mandates that knowledge about the characteristics and health needs of different locations (location profiles) be available in a form suitable for use by a customisation engine against HealthCyberMap's metadata base of resource pointers to select location-specific information resources.

14.2.4 Customisation Parameters

Customisation depends upon many factors and parameters [15]. Some of these have already been discussed above. Following is an incomplete list of parameters that might be relevant to the customisation of health and healthcare information services.

- Personal and health profiles: user role (patient, physician, nurse, etc.); user's demographic and health profiles (might be automatically inferred from his/ her health record if linked to HealthCyberMap); health and healthcare makeup of user's location (location can influence health; also disease rates, management guidelines and health information needs vary with location); user address (addresses of health services presented to the user should correspond to his/ her address, e.g., address of nearest hospital or pharmacy); user country and region (urban/ rural); user's socio-economic group; lifestyle; user dietary style/ preferences; user culture; language; literacy; user education; previous knowledge;

- Computer and Internet profiles (per user): user computer skills (technophobes/ technophiles); user privacy and security needs; user accessibility needs, e.g., large type for sight-impaired patients; user device type, e.g., WAP (Wireless Access Protocol) phone, personal digital assistant, or PC, and device processing power, screen resolution, colour depth, and other display parameters; user agent (browser), supported character encoding sets, supported scripting languages, supported tag sets and data types, and installed plug-ins and versions; available input modalities, e.g., keyboard vs. mouse/ pen vs. voice, and output modalities, e.g., text vs. images/ video vs. audio only; network capabilities such as bandwidth; and
- Other user preferences, e.g., acceptable language (might differ from actual location language), acceptable cost of content (for paid content) and preferred payment mechanisms, and interface appearance (e.g., choosing a colour theme).

14.2.5 Metadata is Important for Customisation

Pointers to good quality resources need to be described in a central database and organised in such way to allow a content management and customisation engine to easily and suitably recall and re-use them in different customisation scenarios.

Metadata are information about information. HealthCyberMap metadata base of resource pointers uses fields (elements) from the well-known Dublin Core (DC—[22]) metadata set scheme for resource description with HealthCyberMap own extensions.

A DC language field makes possible the selection of resources based on their language to match a user's preferred language. A DC coverage field is used to store the spatial extent or scope of the content of a given resource; information in this field should help selecting those resources that are appropriate for a user's location. DC cannot be used to describe resource quality or the geographical location of a resource publisher or author(s) (to be differentiated from coverage, although both are geographic elements), and so HealthCyberMap had to extend the standard DC elements by introducing its own quality and location elements.

14.2.5.1 Multi-axial Classification of Resources Based on Two or More Dublin Core Elements

Resources can be also selected and classified based on different combinations of two or more DC elements as necessary (Figure 14.3). This multi-axial classification of resources based on two or more DC metadata elements should be transparently integrated into other HealthCyberMap interfaces, e.g., the Semantic Subject Search Engine or BodyViewer's human body maps (rather than offered as a separate interface). When integrated into other HealthCyberMap interfaces, multi-axial classification could help filtering and focusing query results to a much smaller, more relevant and more easily manageable set.

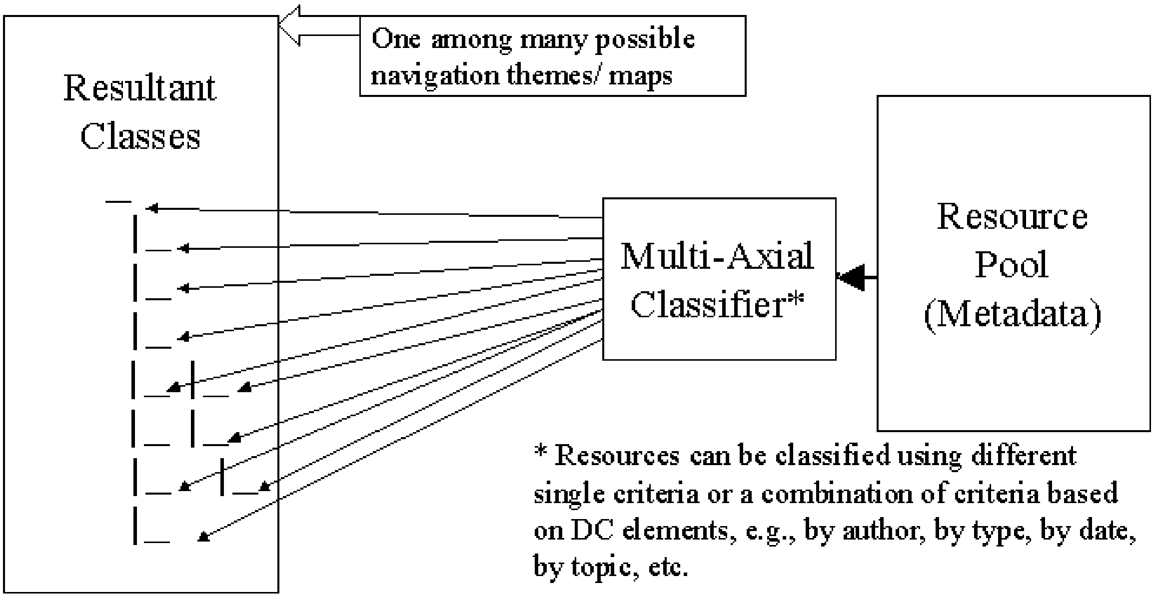


Figure 14.3. Categorising and classifying resources into themed navigation paths according to different fields in their DC metadata records.

An example could be retrieving only "guidelines" in "English language" on "diabetes mellitus", if this is all what the user is asking for, rather than returning all types of resources in all languages on "diabetes mellitus". Following is the SQL (Structured Query Language) query for this example. Note the three DC elements (language, type and subject) that are queried in this example.

```
SQL = "SELECT hcm.[ID], hcm.[dc:Creator], hcm.[dc:Title],
hcm.[dc:Subject:1], hcm.[dc:Subject:2], hcm.[dc:Subject:3],
hcm.[dc:Description], hcm.[dc:Publisher], hcm.[dc:Date],
hcm.[dc:Type], hcm.[dc:Identifier], hcm.[dc:Language],
hcm.[dc:Coverage], hcm.[hcm:Location:city],
hcm.[hcm:Location:country], hcm.[hcm:Quality], hcm.[hcm:Comment]  FROM
hcm  WHERE ((hcm.[dc:Language]) Like 'en') AND ((hcm.[dc:Type]) Like
'Guidelines') AND (((hcm.[dc:Subject:1]) Like '250%') OR
((hcm.[dc:Subject:2]) Like '250%') OR ((hcm.[dc:Subject:3]) Like
'250%'))"
```

In the above code snippet, `hcm` is the name of the database table and 250 is the ICD-9-CM code for "diabetes mellitus".

(See <http://healthcybermap.semanticweb.org/multiaxial.htm> for live demonstrations of this and other examples.)

A visual implementation of the multi-axial classification of resources based on two or more DC metadata elements is also possible. GIS can collate multiple data layers or themes mapped to the same geographic or conceptual spatial framework (like a set of clear transparent overlays) [47]. One possible application of this latter feature in HealthCyberMap could be to map the different resource types (based on the DC type element) as separate layers projected on the human body maps according to resource topics (instead of having all resource types projected as a single layer on the human body maps in the current pilot service). Users can then turn on and off the different DC type layers in different combinations according to their needs, e.g., to only display resources of certain type(s), e.g., guidelines, on the human body topical maps.

14.2.5.2 User Profiles and Device Descriptions

Besides collecting metadata describing information resources, two other types of metadata should be gathered, namely user profiles (including user's location profile which directly affects user's health) and device descriptions. An ideal service should be able to reason with all three types of metadata to personalise and optimise a Web user's experience [126]—Figure 14.4.

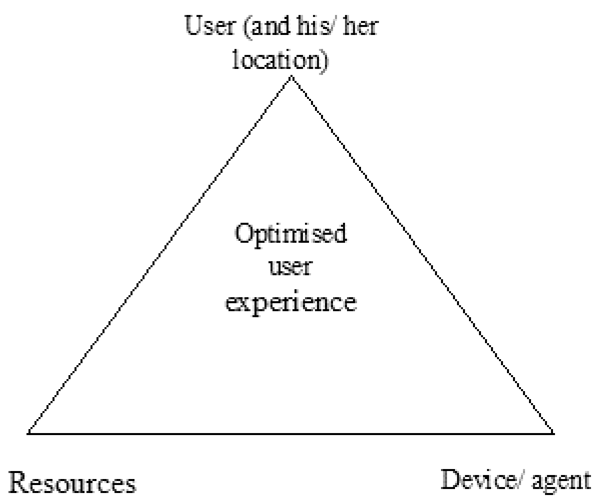


Figure 14.4. Three interrelated categories of metadata are required to optimise (personalise) a user's experience.

14.2.5.3 Customisation Engine

A customisation engine or content management server will use all collected metadata types to select suitable online resources for a particular user and his/ her particular needs at a given time (dynamic match), and also present them in a form that is appropriate for this user and the device he/ she is using to access the resources. The content management and customisation engine will depend on a customisation knowledge base or set of rules to "know" which resource(s) and presentation form are suitable for which situation or profiles. Examples of content management servers and publishing frameworks include Microsoft® Content Management Server (<http://www.microsoft.com/cmserver/>) and Apache Cocoon (<http://xml.apache.org/cocoon/>).

Delivering real-time, location-enhanced and personalised information can help consumers and providers accelerate and optimise their decision-making process in many medical/ health situations and problems. The author believes that the integration of a carefully selected variety of medical/ health Internet information services and resources with users' tasks, needs, preferences and their device capabilities will enable users to focus more on informed decision-making.

14.3 Mapping Health Problems in HealthCyberMap and Identifying Information Needs and Gaps

In addition to mapping medical/ health Web resources, an opportunity exists to map health problems and relate them to available information resources. This can be of great help to healthcare policy makers, planners and managers (helping them make informed decisions). Moreover, any existing *knowledge gaps* in current Web resources can be identified and health information needs can be efficiently and effectively determined and addressed. For example, we might identify a lack of/ a need for adequate Web resources presenting patient education material on a particular health problem. Health and healthcare information providers can then develop any required Web resources or modify existing ones in the light of this information (on current medical/ health information needs and corresponding gaps on the Web).

14.4 Conclusion

HealthCyberMap is a Web-based directory service of medical/ health Internet resources exploring new means to organise and present these resources based on consumer and provider locations. The envisaged location-based medical/ health information service should ultimately allow better presentation of the distribution of health and healthcare needs and Internet resources answering them across a geographical area, with the aim to provide its users with better support for informed decision-making. To enable proper service customisation, three main types of metadata have to be collected and processed: information resource descriptions, user profiles (including user location profile which directly affects user health), and user device descriptions.

Delivering real-time, location-enhanced and personalised information (i.e., information that is *immediately relevant* to users) can help consumers and providers accelerate and optimise their decision-making process in many medical/ health situations and problems. The integration of a carefully selected variety of medical/ health Internet information services and resources with users' tasks, needs, preferences and their device capabilities should enable users to focus more on informed decision-making.

The ideas presented in this chapter concerning HealthCyberMap's future possibilities and available technologies to implement them could be easily adapted to other similar services as well. A general introduction to location-based services, technologies for detecting user location (including tools for mapping user IP address to a city/ country), and their potential applications in health and healthcare was also provided.

Mapping real-world medical/ health problems remains the "native" application of GIS in health (compared to mapping cyberspace), and besides its classic benefits [47], it can also inform medical/ health information providers to develop any required resources or modify existing ones in light of any identified real-world needs that are unmet by current resources.

15 Discussion and Recommendations

HealthCyberMap (<http://healthcybermap.semanticweb.org>) is a Web-based service for healthcare professionals and librarians, patients and the public in general that aims at mapping selected parts of medical/ health information resources in cyberspace in novel ways to improve their retrieval and navigation. HealthCyberMap pilot service currently provides six different interfaces to its metadata base, which has over 1600 resource records in it. Some of these interfaces are visual (maps for browsing resources by clinical/ health topic, by provenance and by type), while others are textual (multilingual interfaces for browsing resources by language, and a directory of topical resource categories, besides HealthCyberMap Semantic Subject Search Engine). The visual maps are used to locate, launch medical/ health resources on the Web, and display their bibliographic metadata records.

HealthCyberMap relies on stand-alone metadata (in a central database) based on the Dublin Core (DC) metadata set [22] with HealthCyberMap's own extensions for resource quality and geographical provenance.

15.1 Resource Quality Benchmarking in HealthCyberMap

Health and medical Web resources are not all written by qualified, unbiased professionals, hence the need for rigorous quality benchmarking when selecting resources for HealthCyberMap [64]. HealthCyberMap also subscribes to the HONcode principles and has received HONcode active seal and certificate.

A quality element has been introduced in HealthCyberMap's extended DC metadata set to store a resource's level of evidence, and any other relevant information regarding its compliance to a recognised code of ethics or quality seal, and whether it has been published by a trusted publisher or listed in trusted directory However, we still need to define a consistent way for storing this quality information and all its facets in the slot that has been introduced to ensure reliable processing of slot values later on, especially if HealthCyberMap's metadata are to be made available for use by other external services as well. Emerging quality "standards" like HIDDEL (Health Information Disclosure, Description and Evaluation Language—<http://www.medcertain.org/english/about_us/overview.htm>) and European Commission Guidelines (eEurope 2002/ eHealth Quality Criteria for Health Related Websites

<http://europa.eu.int/information_society/eeurope/ehealth/quality/draft_guidelines/index_en.htm>) must be also taken into consideration.

15.2 Manual vs. Automatic Indexing of Resources

Many people have proposed using Natural Language Processing (NLP) to figure out Web resources. Unfortunately, NLP is still immature, and has not yet overcome many tremendous obstacles [3]. NLP tools still cannot index a wide range of multimedia Web resources, e.g., medical images in digital pathology and dermatology atlases, clinical diagrams, videos of operative procedures, and audio clips of heart murmurs. And even with purely textual resources, NLP cannot always pick up the correct context of identified concepts or establish their actual relevance (e.g., "penicillin" referred to as an antibiotic and prescription vs. allergy to "penicillin").

"Automatic" topic indexing tools like Nomindex (using MeSH[‡‡‡]—<http://www.med.univ-rennes1.fr/doc/nomindex/noomindex.html>) and Reuters Health version of Metatagger (using SNOMED[§§§]—<http://www.interwoven.com/products/content_intelligence/index.html>) are at best semi-automatic tools that could possibly assist but never totally replace the human cataloguer (much like the quality checklists and benchmarking tools also used by human cataloguers [64]). Indeed, Interwoven, the manufacturer of Metatagger, stresses this important fact [127].

Many high quality subject gateways and Web directory services like NMAP (Nursing, Midwifery and the Allied health Professions—<http://nmap.ac.uk/>), OMNI (<http://omni.ac.uk/>) and the Open Directory Project (<http://dmoz.org/>) still rely completely on human cataloguers. Manual indexing by trained humans as in HealthCyberMap (or semi-automatic indexing assisted by suitable software tools) remains the only reliable way today to guarantee the quality of selected resources and the precision of their topic indexing. However, services must make explicit their actual topical indexing process to ensure consistency between different cataloguing officers (if more than one person is in charge) and reduce human error [127].

Automatic free-text resource indexing by conventional Web spiders, although possibly providing much wider coverage in less time, cannot ensure the quality or

---

[‡‡‡] Medical Subject Headings (<http://www.nlm.nih.gov/mesh/meshhome.html>)

[§§§] Systematised Nomenclature of Medicine (<http://www.snomed.org>)

topic indexing precision of spidered resources, and cannot index non-textual, multimedia Web resources.

15.3 On the Use of Clinical Codes in HealthCyberMap

The cybermaps in HealthCyberMap can be considered as semantically-spatialised browsing views of the underlying resource metadata base. Mapping conceptual information spaces of Web resources based on their semantics has been demonstrated in several other systems, e.g., StarWalker [45] (Figure 4.8). However, HealthCyberMap adopts a unique clinical metadata framework that builds upon a clinical coding scheme. This is very much suited for the semantic categorisation, navigation and retrieval of medical/ health information resources on the Internet.

A clinical coding scheme (vocabulary or ontology) can fulfil the following tasks in relation to digital libraries [31]:

- indexing knowledge–both general medical/ health knowledge and information about individual patients; this can form the basis of clinical problem-to-knowledge linking as in HealthCyberMap (<http://healthcybermap.semanticweb.org/pk.htm>); and
- navigating and browsing through information (automatically caring for topic synonyms, and semantic relationships and classification of resources).

15.3.1 Which Clinical Coding Scheme?

Reuters Health (<http://www.reutershealth.com>) currently uses SNOMED, a clinical coding scheme, to categorise medical stories and provide information specific to clients' interests. Compared to MeSH, a bibliographic coding scheme used by many medical Web portals today, clinical coding schemes like SNOMED (and to a lesser extent ICD-9-CM used in the current pilot HealthCyberMap service) offer more precise coding, more specificity of medical conditions (narrower terms) and more sophisticated relationships [65].

ICD is not a true clinical terminology but a hierarchical classification with a much coarser granularity (compared to SNOMED) and no support for multiple parentage. However, HealthCyberMap pilot service uses a clinically modified version of ICD-9 (ICD-9-CM) that is much more suited for clinical applications than the original ICD-9, with generally equal or better clinical expressiveness than MeSH.

The main reason for choosing ICD-9-CM is that the BodyViewer maps in HealthCyberMap (<http://www.healthcybermap.semanticweb.org/bodyviewer/>) rely on a

GIS extension (BodyViewer) that only understands ICD-9-CM (Chapter 9). However, the author believes that in the future, HealthCyberMap should ideally use SNOMED-CT (Systematised Nomenclature of Medicine–Clinical Terms; released in February 2002 and chosen as the standard terminology for UK Electronic Patient Records (EPRs)—<http://www.snomed.org>). With SNOMED-CT, it would be possible to automatically cross map to ICD-9-CM (this is a feature of SNOMED-CT) to continue using the BodyViewer tool in HealthCyberMap to generate the maps, while benefiting from the much finer granularity of SNOMED-CT to provide much better semantic search capabilities.

SNOMED-CT is a compositional terminology that contains over 325,000 concepts linked to clinical knowledge (compared to only about 20,000 main headings in MeSH), 800,000 synonyms/ descriptions, and more than 950,000 links or semantic relationships that can be explored by terminology servers (see below). According to SNOMED International Authority, SNOMED can be used for tagging Web-based consumer health information and for medical literature search and retrieval [128].

Furthermore, using MeSH in HealthCyberMap for Problem to Knowledge Linking with a SNOMED-coded EPR would not be an optimal solution. SNOMED to MeSH mapping via UMLS (Unified Medical Language System) has not been very successful [129], and with only about 20,000 main headings in MeSH, such mapping will definitely result in a considerable loss of clinical details, and hence much less relevant and focused problem to knowledge links.

15.3.2 Using a Terminology Server

Terminologies and classifications of the past were used as mere phrase books that could be browsed with simple word processors and queried with simple relational databases [130]. These approaches rapidly become inadequate for more complex terminologies and uses. With the standardisation of terminologies, disparate clinical applications like HealthCyberMap and EPRs are expected to use them in similar and complementary ways (also for interacting with each other, e.g., EPR to HealthCyberMap clinical Problem to Knowledge Linking). According to Bechhofer and colleagues [131], such clinical terminologies represent ontologies that should not be embedded in client applications, but should rather be shared and reused as a distributed resource by implementing it as a service through a terminology server.

A terminology server is a special type of ontology servers that allows, given a terminology concept, the retrieval of synonyms and related broader/ narrower concepts (parent, cousin, uncle, sibling and child concepts) from the underlying clinical terminology. It should be also able to query and cross-map multiple terminologies/ classifications at the same time if needed. Ideally, a terminology server should support concept mapping, which involves processing free text queries to identify corresponding terms from a controlled vocabulary. This relieves users from any restrictions while ensuring accurate results and can also support spelling variants and, if necessary, multiple languages (SNOMED-CT concepts are language neutral and so can serve different natural languages). A terminology server can be implemented around a description logic classifier to exploit the description logics underpinning modern terminologies like SNOMED [132]. Chute et al [133] mention the following desiderata for a clinical terminology server: word normalisation, word completion, target terminology specification, spelling correction, lexical matching, term completion, semantic locality, term composition and decomposition.

The current early pilot HealthCyberMap Semantic Subject Search Engine (Chapter 8) does not support many of the above mentioned desiderata. For example, it only supports US spelling and has no tolerance for spelling mistakes. A suitable terminology server (fulfilling all of Chute's desiderata) would clearly improve the situation by processing raw user queries in superior ways and optimising the resultant DC subject queries passed to HealthCyberMap, while supporting all kinds of semantic relationships (anatomical, physiological, pathological, clinical, etc.) between concepts when retrieving resources (Figure 15.1). Examples of terminology servers include jTerm (<http://mycin.ucdavis.edu/jterm/index.html>), a Java-based open source terminology server, and Apelon solutions (<http://www.lexical.com/>).

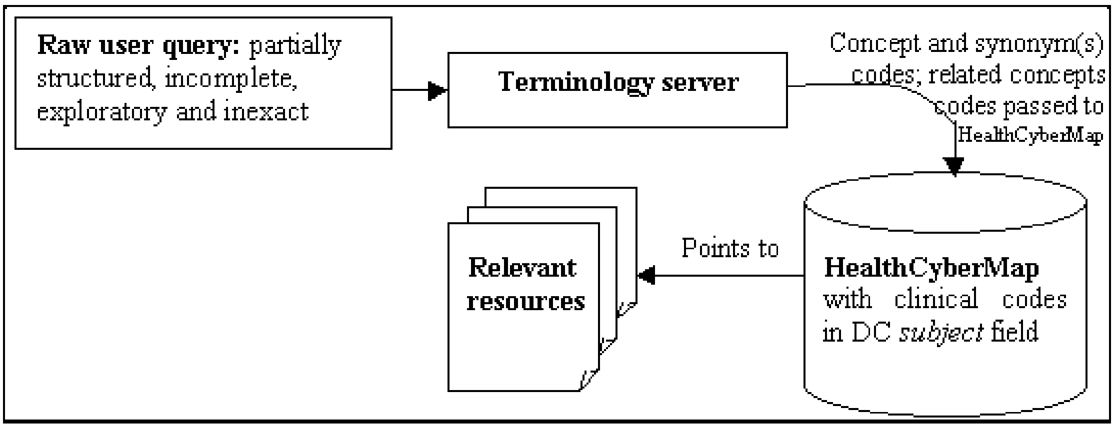


Figure 15.1. Using a terminology server can help formulating and optimising DC subject queries passed to HealthCyberMap. It can also assist in relating the resource pointers stored in HealthCyberMap to each other and classifying them based on their subject and clinical context as determined by their DC subject field.

15.4 Combining Resource Metadata with Related Ontologies (Knowledge Domains)

Semantic Web resources must be marked-up with metadata and/ or indexed in a central database using metadata. Explicit concepts in the metadata could then map onto an ontology, e.g., a clinical terminology or classification, or a collection of merged ontologies allowing a Semantic Web agent/ search engine to infer implicit meanings not directly mentioned in either the resource or its metadata (Figures 15.2 and 15.3) [62]. This is the approach which the author adopted in HealthCyberMap Medical Semantic Subject Search (Chapter 8).

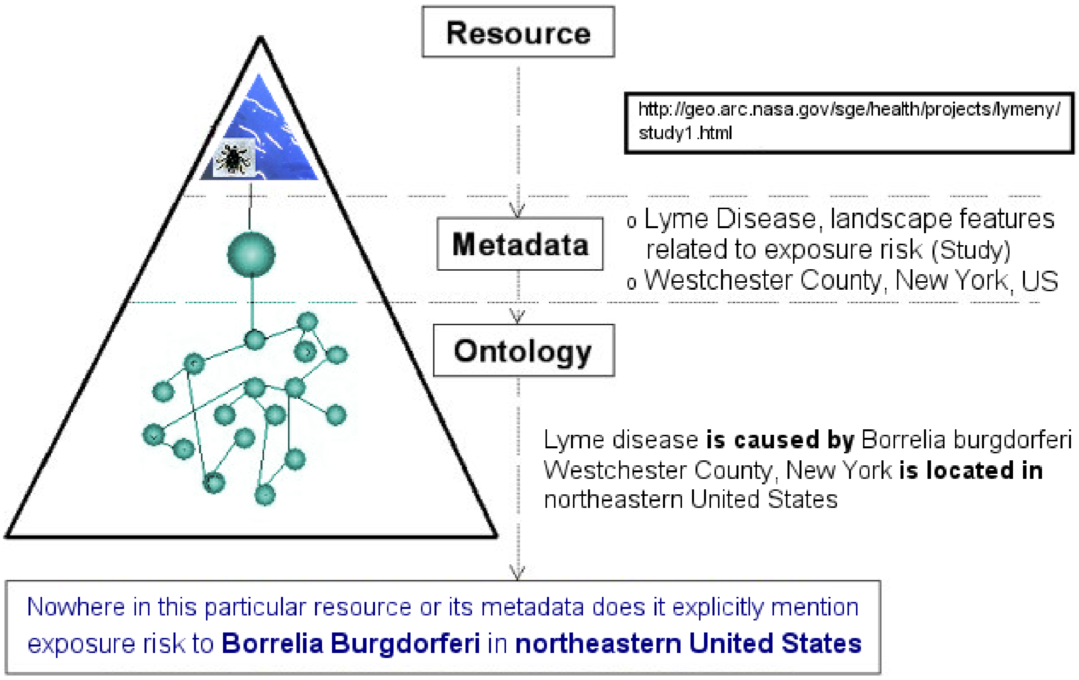


Figure 15.2 (Modified from [62]). Metadata alone is not enough for successful retrieval of this resource. In this example, even though the resource and its metadata do not mention 'exposure risk to Borrelia Burgdorferi in north-eastern United States', an assisted search for 'exposure risk to Borrelia Burgdorferi in north-eastern United States' would find <http://geo.arc.nasa.gov/sge/health/projects/lymeny/study1.html>.

Metadata ontologies are not limited to describing information resources. Two other groups of metadata are equally important, namely user profiles (including user's location profile which directly affects user's health) and device/ agent descriptions (Figure 15.3). An ideal system should be able to reason with all three types of metadata to personalise and optimise a Web user's experience [126] (see Chapter 14).

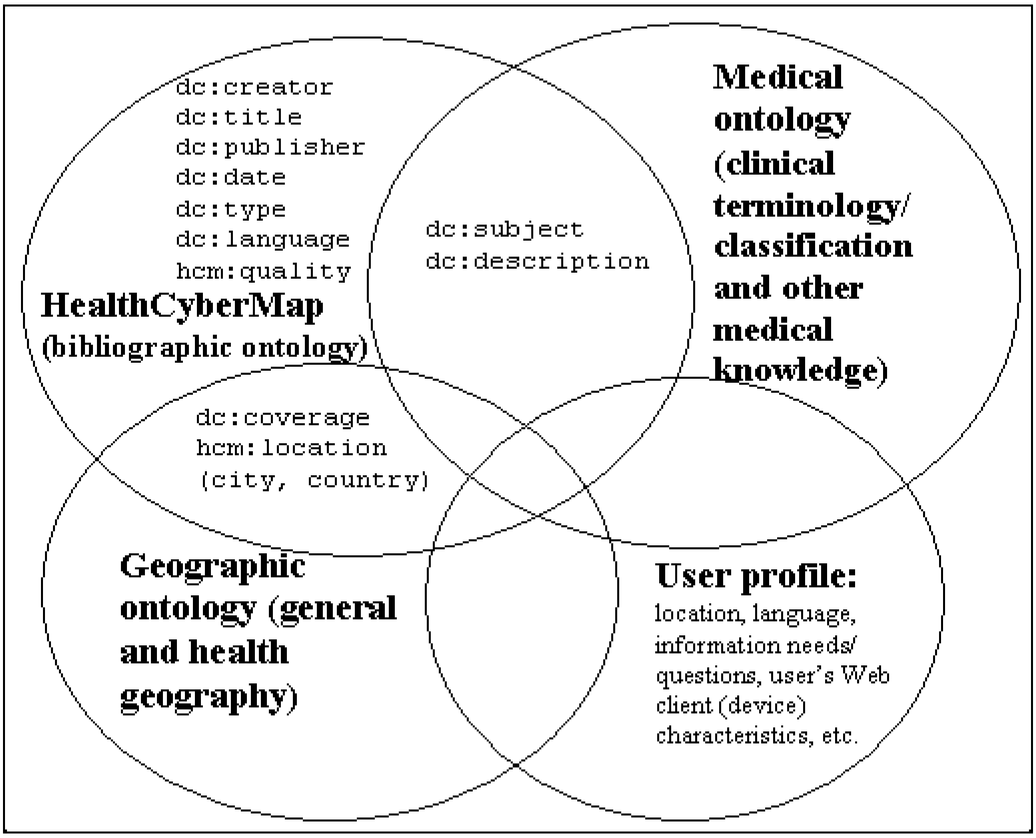


Figure 15.3. This diagram illustrates some of the possible ontologies that can be used and merged together in HealthCyberMap, and their intersections with each other and with HealthCyberMap's main Dublin Core metadata vocabulary.

15.5 HealthCyberMap Medical Semantic Subject Search Engine

HealthCyberMap Medical Semantic Subject Search Engine attempts to overcome the limitations of conventional free text search engines without the need to encode all synonyms, semantic relationships and other possibilities of related topics in a resource or a metadata record of it, which is not practical or computationally efficient. HealthCyberMap achieves this by mapping explicit concepts in resource metadata onto a brokering domain ontology (a clinical terminology or classification—ICD-9-CM in the current pilot service) allowing a Semantic Web search engine to infer implicit meanings (synonyms and semantic relationships) not directly mentioned in either the resource or its metadata. Similarly, user queries would map to the same ontology allowing the search engine to infer the implicit semantics of user queries and use them to optimise retrieval. The early pilot HealthCyberMap Semantic Subject Search Engine currently available online supports synonyms, disease variants and subtypes (<http://healthcybermap.semanticweb.org/icd.htm>).

15.6 A Simple Method for Serving Web Hypermaps with Dynamic Database Drill-down

HealthCyberMap is an ESRI ArcView GIS 3.1 project that uses GIS (Geographic Information System) spatialisation methods to generate interactive navigational cybermaps from an underlying resource metadata base. (The visual map used to browse resources by type is the only map created outside ArcView in Boutell's Mapedit v2.63.) WebView, the Internet extension to ArcView, publishes HealthCyberMap ArcView Views as Web client-side imagemaps. The basic WebView set-up does not support any GIS database connection, and published Web maps become disconnected from the original project. A dedicated Internet map server would be the best way to serve HealthCyberMap metadata-base-driven interactive Web maps, but is an expensive and complex solution to acquire, run and maintain. The author developed a simple, low-cost method for "patching" WebView to serve hypermaps with dynamic database drill-down functionality on the Web using dynamic ASP pages (Active Server Pages) to query the same metadata base used in ArcView and registered on HealthCyberMap Web server as an ODBC (Open DataBase Connectivity) data source. This solution is currently used for publishing HealthCyberMap GIS-generated navigational information maps on the Web while maintaining their links with the underlying resource metadata base.

The author believes his map serving approach as adopted in the current HealthCyberMap pilot service has been very successful, especially in cases when only map attribute data change without a corresponding effect on map appearance. It should be also possible to use the same solution to publish other interactive GIS-driven maps on the Web, e.g., maps of real world health problems.

15.7 Meaningful Maps Without Clutter

Using a clinical coding ontology as a metric for spatialisation ("semantic distance") to generate meaningful navigational cybermaps is unique to HealthCyberMap. Ontology-based information visualisation is a rapidly growing research field [134] to which HealthCyberMap proudly belongs by adopting a clinical ontology-based framework (DC and ICD-9-CM vocabularies) for the semantic lassification and visualisation (browsing and navigation) of medical/ health Internet resources.

The author believes that the use of familiar medical metaphors (pictorial/ associative icons) for visualising these resources is far more superior to using arbitrary, geometric map symbols to represent these resources on a map (like the stars and dots in StarWalker [45]—Figure 4.8 and Visual Net PubMed interface—Figures 4.2, 11.2 and 11.3).

Topical coverage gaps can be also easily identified thanks to HealthCyberMap's human body choropleth (shaded) maps of resource counts and addressed by HealthCyberMap's cataloguer(s) and/ or external information providers.

HealthCyberMap's map query results (resources) are listed in a separate text window (Figure 9.1) to provide comprehensive information about resources and avoid map clutter. The latter would have been unavoidable had we opted to represent each resource using a distinct point symbol on the map (cf. Visual Net PubMed interface—Figure 4.2). Moreover, returned resource record details in HealthCyberMap are collapsed by default to show only the few most essential metadata elements (the "Show/ Hide Details" option can be used to expand/ collapse a resource's bibliographic details—Figure 9.1). Map query results are paginated to display only ten resource records per page with handy navigation links at the bottom of each page to move back and forth between returned resource records. Query results will always reflect the latest updates carried on HealthCyberMap metadata base without the need to change any code.

15.8 Complementary Interfaces

The different forms of spatialisation and corresponding hypermaps in HealthCyberMap complement each other rather than being mutually exclusive. Though nobody interested in information about say "angina pectoris" will try to search and call up this information by looking for and clicking on a map with the geographical location of the servers carrying that information (they would go instead to the human body maps for this kind of queries), the geographical world maps can still prove useful when browsing for no specific reason (exploring), or doing some analytical research on the provenance of different resources [40, 41], or looking for location-specific health services, disease rates or guidelines (see Table 14.1).

HealthCyberMap's hypermaps should be also perceived as a complementary improvement over (rather than total replacement of) HealthCyberMap's textual interfaces. Depending on user's primary role (healthcare professionals and librarians, patients and the general public), his/ her prior knowledge and query, hypermaps could be more intuitive and faster than textual category lists and keyword searches in locating and selecting topics/ resources.

15.9 HealthCyberMap Problem to Knowledge Linking Service

HealthCyberMap Problem to Knowledge Linking (PKL) service makes use of metadata and clinical codes to contextually link disparate EPR clients to resources in an online medical/ health knowledge service (like HealthCyberMap). Clinical codes act as crisp knowledge hooks, providing a reliable common backbone language (grounding ontology) for communication between EPRs and HealthCyberMap (Chapter 10). A prototype Web-based Emergency Room EPR, InfoMED ER (<http://www.infomed-epr.com>), developed in 2002 by Dr. R.G. Rochin at City University, London, is currently the first remote client EPR to use HealthCyberMap pilot PKL service to provide contextual resource links within patient records.

15.9.1 HealthCyberMap as an XML Web Service for EPR clients

HealthCyberMap PKL service could be also implemented in the future as an XML (eXtensible Markup Language) Web Service (<http://www.w3.org/2002/ws/>), giving EPR clients more control over presentation of query results and user interface (since resource metadata records will be returned as "neutral" XML that can be further processed by clients). However, no final standards exist yet for Web Services (SOAP—Simple Object Access Protocol and WSDL—Web Service Description Language are still W3C (World Wide Web Consortium) notes and drafts, not recommendations, as of March 2002), and there are still some incompatibilities between different manufacturers' implementations of Web Services (e.g., IBM and Microsoft—see <http://ws-i.org/>).

15.9.2 Refining Problem to Knowledge Linking Searches

To maximise contextual relevance, indexed resources should ideally be qualified using a mechanism similar to MeSH qualifiers or subheadings, which are used to better define a topic, narrow retrieval, or express a certain aspect of a main heading ([135]—ICD has no similar support). Another idea could be to send the clinical heading (under which the code to be matched occurs [136]) as a query argument along with the clinical code itself (the resource metadata base will need some modification to answer queries that include a clinical heading). It is noteworthy that SNOMED features modifiers (string which when added to a term changes the meaning of the term in a clinical sense, e.g., clinical stage or severity of illness) and qualifiers (string which when added to a term changes the meaning of the term in a temporal or administrative sense, e.g., "history of" or "recurrent").

15.10.3 Dynamic and Flexible Problem to Knowledge Linking

Unlike other inflexible solutions with hard-coded knowledge like Dr. Weed's Problem Knowledge Couplers (<http://www.pkc.com/>), HealthCyberMap's solution is flexible and dynamic. We can keep adding/ deleting resources to HealthCyberMap and have all new changes instantly reflected on our output without modifying HealthCyberMap architecture or code (or the client's calling code). In a future implementation, it should be possible to choose to call certain resource categories we need most (according to user profile and specific needs), e.g., only official guidelines on a given topic, instead of all resource types on that topic, thanks to the HealthCyberMap's DC type field (see also <http://healthcybermap.semanticweb.org/multiaxial.htm>). Dr Weed's Problem Knowledge Couplers are somewhat inflexible and the knowledge in them is hard-coded and cannot be easily and dynamically changed or updated.

By minimising irrelevant leads (noise) and reducing the time needed to find relevant information (the right contextually relevant knowledge is linked to real patient data in the EPR), HealthCyberMap's proposed PKL service is potentially beneficial. Most respondents to HealthCyberMap's formative evaluation questionnaire, especially clinicians, welcomed the idea of PKL and thought it could be useful. The ultimate success of the service will depend on the quality and granularity of the metadata it uses, the topical coverage and quality of resources it points to, and the use of a suitable concept qualifier mechanism to maximise contextual relevance.

15.10 HealthCyberMap vs. Topic Maps

HealthCyberMap is clearly sharing most of ISO Topic Maps' pivotal concepts [137]. Thanks to its resource metadata base, HealthCyberMap can automatically and dynamically categorise (classify) the resources in its index in many different ways to generate different sets of visual and textual "topic maps". Although the acquisition of metadata in HealthCyberMap depends on a human cataloguer, the automated categorisation of these resources based on clinical codes and other metadata fields should save the cataloguer lots of effort and time. The underlying clinical coding scheme could also help the automatic generation of a "See also" list of topics/ resources related to a given resource subject code. (Human cataloguers at Web directory services perform a two-folded task of indexing and classifying Web resources. The first step involves gathering metadata about resources, while the second step deals with putting selected resources under suitable categories and possibly creating new categories if needed. We propose making the second task fully automated.) ISO Topic Maps on the other hand largely depend on manual categorisation.

The DC metadata set can be easily mapped to ISO Topic Maps. In HealthCyberMap, the actual topics (concepts) are the clinical codes, which are themselves extracted from a separate ontology: ICD-9-CM (to populate the DC subject field). The occurrences are the Web resources themselves [DC identifier field using URI (Uniform Resource Identifier) scheme]. Occurrence roles correspond to DC type field, e.g., an image of some skin condition vs. a fact-sheet on the same condition.

15.10.1 RDF vs. XML Topic Maps

Metadata serialised in XML-based languages are very much suited for data interchange. In the first part of Chapter 7, we proposed using RDF/RDFS (Resource Description Framework and its Schema—RDF is serialised in XML) to build a medical/ health metadata base and share it with other services following the Open Directory Project (<http://dmoz.org/>) metadata sharing model, which is also based on RDF [76].

It is also possible to share metadata saved in XTM (XML Topic Maps) format. Although Topic Maps are an ISO standard, XTM is not (ISO recommends serialising Topic Maps in a different way and in SGML—Standard Generalised Markup Language, a superset of XML). Moreover, ISO Topic Maps currently lack a schema/ standardised way of expressing constraints and do not fully support class hierarchies in a formal, standardised way. (A forthcoming Topic Map Constraint Language (TMCL) should provide a formal definition of class semantics as they apply to Topic Maps.) RDF on the other hand is superior in these respects and is an approved W3C Recommendation, though RDF schema (RDF Vocabulary Description Language 1.0) is still a Working Draft (as of November 2002—<http://www.w3.org/TR/rdf-schema>).

15.11 HealthCyberMap Evaluation

The author believes that evaluation should run throughout the whole lifetime of any service (not just for a limited time) to make sure the service is always delivering what it has promised, and to overcome designers' blindness (discover deficiencies overlooked by designers and only seen by users). Different evaluation methods have different strengths and weaknesses, and only a good approach to Web service evaluation involving more than one method can help putting together an adequate, more complete picture of how a Web service is being used and received by users [93, 94]. For this reason, the author conducted a formative evaluation of HealthCyberMap pilot service using two complementary methods: an online user questionnaire (<http://healthcybermap.semanticweb.org/questionnaire.asp>) and server log analysis.
Thirty-five subjects responded to HealthCyberMap online evaluation questionnaire during the 45-day period from 18 April 2002 to 1 June 2002. The sample of respondents was in general a good representative of the wider Internet/ HealthCyberMap audience. Most respondents hold a very positive attitude towards the Internet as a credible source of health information and towards visual maps as a navigational aid for medical and health-related Internet resources.

HealthCyberMap scored high on all general and specific questions covering user satisfaction and service usability, e.g., metaphor comprehension, ease of use and adequacy of online help, map loading speed and management of detail overload. More than 94% of respondents were successful in completing the usability task (part of the online evaluation questionnaire).

HealthCyberMap BodyViewer maps of Web resources classified according to medical/ health topics received the highest usability and usefulness rates among the evaluated pilot interfaces. The human body proved to be an excellent and familiar metaphor for the topical browsing of medical/ health Internet resources. Indeed, whenever medicine or health are mentioned we nearly always remember our bodies.

The big majority of respondents thinks HealthCyberMap resource pointers are of good quality and relevant to user queries.

All proposed HealthCyberMap future directions received favourable ratings by respondents.

Internet search engines and links from other sites proved to be more effective than targeted e-mail and postings to mailing lists in advertising the service.

The results of the comparative task (also part of the online evaluation questionnaire) were also very encouraging, with most respondents finding HealthCyberMap approach to be better than or as good as that of Visual Net PubMed interface.

Sawmill 6.3.8 was used to analyse HealthCyberMap server log files for the 45-day period beginning 18 April 2002 and ending 1 June 2002. The overall trend is that HealthCyberMap traffic and popularity have been growing steadily over the analysed period, with adequate server performance throughout the same period. Better visitor retention figures could be expected if the service is developed past the current pilot stage to offer a wider, more comprehensive coverage of health and clinical topics and more functional Web interfaces.

BodyViewer's human body topical maps received the highest number of unique visitors compared to other HealthCyberMap interfaces. HealthCyberMap visitors during the 45-day period of this study could be traced to more than 50 countries around the world. More sophisticated server log analysis tools are needed that utilise their own database to resolve more visitors' IP addresses into more recognisable and useful provenance data. Determining the most active countries accessing a site can help prioritising development plans of new interfaces and content in other languages. This in turn will ensure rapid and sustained leadership over any competitor services that might be also exploiting the potential of multilingual interfaces and content.

Based on the top Web browsers and operating systems accessing HealthCyberMap during the period of this analysis, if one is to focus future development efforts on the few most prevalent platforms/ browsers, these would be Windows 2000/XP and Windows 98/Me operating systems, and Internet Explorer browser version 5.x or later. Moreover, from the questionnaire analysis it was also learned that developing for a Web browser window size of a little less than 800 by 600 pixels remains the safest choice today to avoid the need for scrolling.

Google and Yahoo! (also powered by Google) ranked top among referring search engines/ sites.

Sawmill showed no particular recurring navigation path pattern taken by visitors through the site, though nearly half of all recorded visitor sessions started at HealthCyberMap main front page.

Counting page views and hits alone is not an accurate measure of user interest. The author believes that a time-based metric of user interest or content usefulness based on standard server logs is also not adequate because of browser/ proxy caching effects, network latency, and other reasons. A representative sample of service users have to be surveyed (as the author did in HealthCyberMap) if we are to clearly determine whether or not they found the information presented to them useful and were able to easily find all what they wanted.

15.12 Main Recommendations

The following recommendations for further developing and improving HealthCyberMap were guided by HealthCyberMap's formative evaluation results:

- Using a more comprehensive clinical coding scheme like SNOMED-CT combined with better human body maps and a true terminology server [131]. The latter would allow us to reason with clinical codes (resource subjects) in more sophisticated semantic ways when retrieving resources.
- Organising information resources by intended primary audience as patients, health professionals, or basic researchers.
- Designing better icons and maps with enhanced details for better metaphor comprehension, caring for different user roles, backgrounds and needs.
- Further developing HealthCyberMap textual directory of topical resource categories to allow users to browse and select hierarchical sub-levels (subcategories) of the main (top-level) ICD-9-CM categories to focus queries on smaller groups of resources/ narrower topics (Figure 15.4).
- Introducing additional resource groupings and visual metaphors based on the same underlying resource metadata, e.g., small images of the different blood cells linked to resources on blood diseases classified according to the major blood cell type affected in each disease. For skin conditions, a regional (affected body region/ sub-region) and morphological (type of skin lesion, e.g., macule, papule, plaque, nodule, etc.) grouping of resources could prove very useful.
- Using HealthCyberMap world maps as an interface to location-specific health services, travellers' health information, disease rates and guidelines, in addition to the current use as an interface to resources classified by the geographic provenance of author(s)/ publisher(s).
- Having other sets of world maps (could be implemented as map layers or themes that users can turn on or off on demand), each mapping only resources covering certain topic(s), e.g., Tuberculosis, to their respective countries (vs. all resources/ topics in the current pilot world map of resources). This could help us know the

geographic provenance of resources on a particular topic or topics, and a choropleth rendition could visually order the countries of the world according to their contribution to this specific subset of topics.

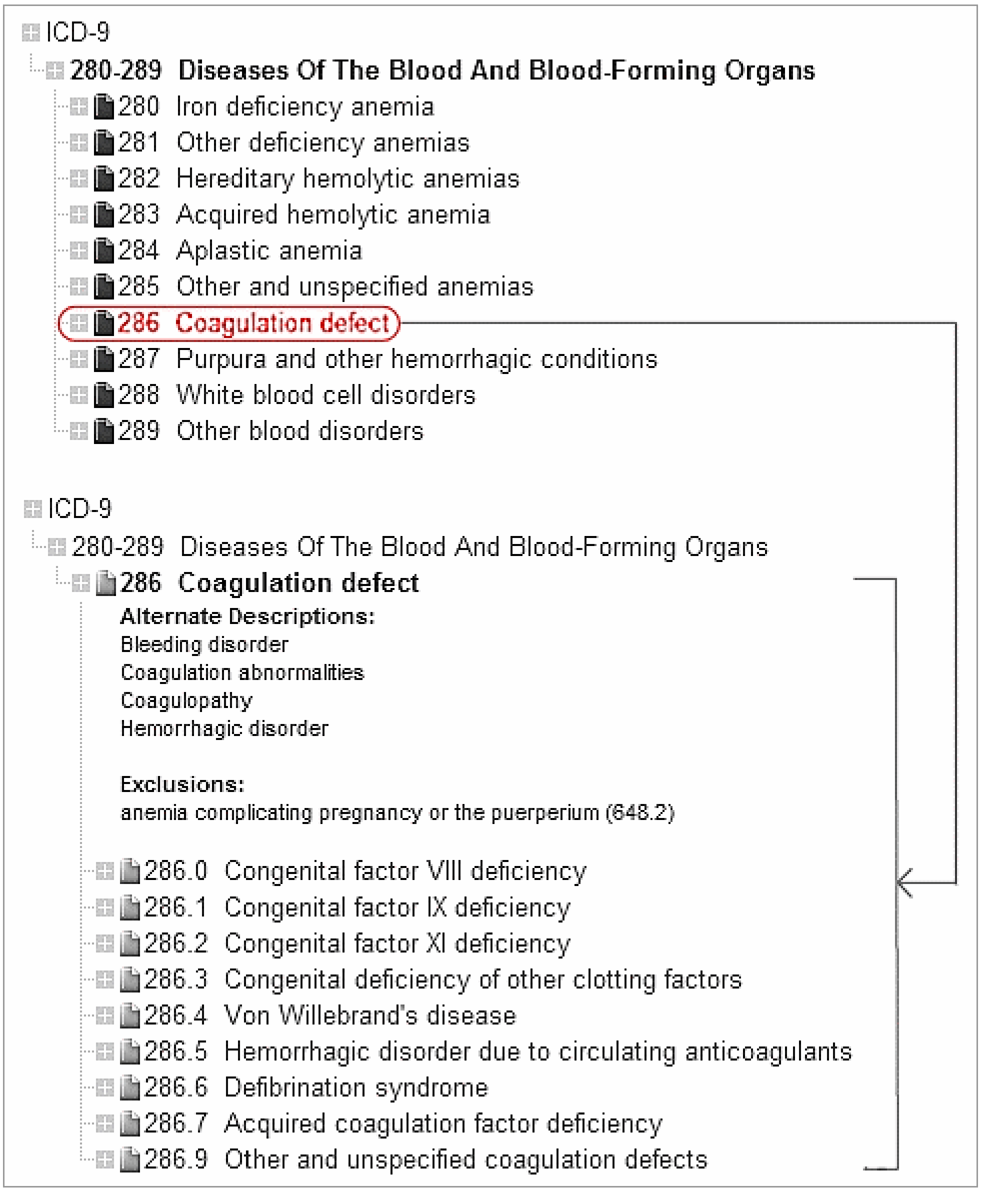


Figure 15.4. Mock-up screenshot of the proposed browsing of hierarchical sub-levels (subcategories) of the main (top-level) ICD-9-CM categories. In this example, “Coagulation defect” under “Diseases Of The Blood And Blood-Forming Organs” has been expanded.

- Transparently integrating the current “classification of resources by language” interface into other interfaces, so that, when the different types of HealthCyberMap maps are clicked, only Web resources that are in user’s preferred language are retrieved or displayed first before resources in other languages. The same applies to the multi-axial classification of resources based on two or more Dublin Core metadata elements, which should be also transparently integrated into other HealthCyberMap interfaces to help filtering and focusing query results to a much smaller, more relevant and more easily manageable set according to user needs. A good example could be retrieving only guidelines in a user's language on a particular condition, if this is all what the user is asking for, rather than returning all types of resources in all languages on that condition.

- Supporting customisation based on user's geographical location (IP geolocation—[123]) to automatically deal with language, as well as any specific health needs/ online resources related to user's location. HealthCyberMap as a customisable location-based medical/ health information service requires the collection and processing of metadata describing not only resources, but also user locations (from a health perspective), users themselves, and their devices.
- An opportunity also exists to map real-world health problems (using the same GIS/ imagemap-serving infrastructure), relate them to available Web resources, and identify and address Web knowledge gaps. The author thinks medical/ health information providers will much appreciate this kind of information (cf. urban planning when assisted by good real-world maps).
- Implementing HealthCyberMap's metadata base in RDF/RDFS based on the *qualified* DC metadata set, which offers many opportunities in the future, including the possibility of other services making use of HealthCyberMap's RDF metadata base to develop their own Web portals. HealthCyberMap Problem to Knowledge Linking Service must be also developed in accordance to W3C (World Wide Web Consortium) recommendations for Web Services. RDF/RDFS and related languages, and Web Services provide the "grammar" necessary for reliable information exchange across services and transparent service interoperability using shared vocabularies.

16 Conclusions

16.1 Aims and Objectives Revisited

16.1.1 Aims

HealthCyberMap's core aims "to map selected parts of medical/ health information resources in cyberspace in novel meaningful ways to improve their retrieval and navigation" has been largely fulfilled in the current Web-based pilot service (<http://healthcybermap.semanticweb.org/>). The service features separate, though tightly coupled, semantic (resource index) and navigation (visualisation) layers for maximum flexibility and maintainability (Chapter 6).

16.1.2 Semantic Arm Objectives

All semantic arm objectives have been achieved in the current pilot service (with only very few sub-items thoroughly investigated as possible future work). The author extended the generic Dublin Core (DC) metadata set to suit the description of medical/ health Internet resources. A new metadata element was introduced to describe resource quality. A clinical coding scheme was used to describe resource topics in a consistent way. This method of using clinical concept codes to represent resource subject is a great improvement over the more conventional way of using textual keywords for this purpose. A clinical coding scheme can automatically care for topic synonyms, semantic relationships between topics, and classification of resources under broader topic categories as demonstrated in HealthCyberMap.

A pilot medical/ health resource metadata base was first modelled in RDF/RDFS (Resource Description Framework and its Schema, which is rapidly becoming the lingua franca of the future Semantic Web) based on the qualified DC vocabulary. This approach based on RDF and the qualified DC metadata set offers many opportunities in the future, including the possibility of other services making use of HealthCyberMap's RDF metadata base to develop their own Web portals. However, for the purpose of the current HealthCyberMap pilot service, a simpler metadata base was next built in Microsoft® Access based on the non-qualified DC metadata set. At the time of writing, this metadata base has 1640 resource records in it. HealthCyberMap's editorial policy for resource selection and quality benchmarking is based on HONcode principles ([84]—HealthCyberMap pilot service received

HONcode active seal and certificate) and observes the spirit of the major quality checklists available today like DISCERN (<http://www.discern.org.uk/>), and Net Scoring (<http://www.chu-rouen.fr/netscoring/netscoringeng.html>) [64]. (See Chapter 7.)

A pilot Semantic Subject Search Engine has been developed and is currently available online (<http://healthcybermap.semanticweb.org/icd.htm>). It supports synonyms, disease variants, subtypes and some semantic relationships between medical/ health topics. It attempts to overcome the limitations of conventional free text search engines without the need to encode all synonyms, semantic relationships and other possibilities of related topics in a resource or a metadata record of it, which is not practical or computationally efficient. Explicit concepts in resource metadata map onto a brokering domain ontology (the same clinical coding scheme used for resource topic indexing) allowing the search engine to infer implicit meanings (synonyms and semantic relationships) not directly mentioned in either the resource or its metadata. Similarly, user queries map to the same ontology allowing the search engine to infer the implicit semantics of user queries and use them to optimise retrieval (Chapter 8).

A pilot clinical Problem to Knowledge Linking (PKL) service was also built (see demonstrator at: <http://healthcybermap.semanticweb.org/pk.htm>—Chapter 10). The service makes use of metadata and clinical codes to contextually link real patient data and problems in remote electronic patient record (EPR) clients to relevant information resource pointers in HealthCyberMap. Clinical codes act as crisp knowledge hooks, providing a reliable common backbone language for communication between EPRs and HealthCyberMap. It is noteworthy that a prototype Web-based Emergency Room EPR, InfoMED ER (<http://www.infomed-epr.com>), developed in 2002 at City University, London, is currently the first remote client EPR to use HealthCyberMap pilot PKL service to provide contextual resource links within patient records.

Combining resource metadata with other types of metadata and related ontologies (knowledge domains) can help reasoning with resource metadata in semantically superior ways and inferring implicit meanings not directly mentioned in either the resources or their metadata (as demonstrated in the above-mentioned pilot Semantic Subject Search Engine). Metadata are not limited to describing information resources. Two other groups of metadata are equally important, namely user profiles (including user's location profile which directly affects user's health) and device/ agent descriptions. An ideal system should be able to reason with all three types of metadata to personalise and optimise a Web user's experience and this has been thoroughly investigated as possible future work to further develop HealthCyberMap as a customisable, location-based medical/ health information service (Chapter 14).

16.1.3 Visualisation/ Navigation Arm Objectives

All visualisation/ navigation arm objectives were also fulfilled in the current pilot service (Chapter 9). A review/ critique of related cybermapping projects was also carried to inform HealthCyberMap design and development (Chapter 4).

In HealthCyberMap, the author used a "distance" metric based on the "semantic locations" of resource topics within a clinical coding scheme projected on a human body organs/ systems map. The clinical coding scheme acts as a semantic conceptual space with resources occupying different locations in this space based on their subject topics. The "semantic distance" between two resources will then depend on how close (or related) the two resources are from a semantic perspective. For example, a resource on "myocardial infarction" should be much closer to a resource on "angina pectoris" than to another resource on "psoriasis".

The current HealthCyberMap pilot service provides different (but complementary) graphical and textual interfaces to browse and navigate medical/ health Internet resources based on clinical codes and other metadata elements in HealthCyberMap's resource metadata base. It uses "conventional" geographical hypermaps to map health resources on the Web to the country of their corresponding providers (<http://healthcybermap.semanticweb.org/world_map/>). Another set of hierarchical human body topical maps allows to visually browse resources by body location/ system according to ICD-9-CM, which acts as HealthCyberMap's medical ontology in the current pilot service (<http://healthcybermap.semanticweb.org/bodyviewer/>). The human body is an excellent familiar metaphor for the topical browsing of medical/ health Internet resources. Indeed, whenever medicine or health are mentioned we nearly always remember our bodies. A third type of hypermaps in HealthCyberMap categorises resources by type, e.g., electronic journal articles, digital atlases, etc. (<http://healthcybermap.semanticweb.org/type.htm>).

HealthCyberMap has been developed as an ArcView GIS 3.1 project. It uses GIS (Geographic Information Systems) spatialisation methods to generate its interactive navigational cybermaps from the underlying resource metadata base.

HealthCyberMap also introduces a useful form of cyberspatial analysis in the current pilot service for the detection of topical coverage gaps in the resource metadata base using its human body topical choropleth (shaded) maps of resource counts (body organs with darker red tints on these maps have more resources associated with them than organs with lighter red shades; a grey colour denotes no resources).

A dedicated Internet map server would have been the best way to serve HealthCyberMap's database-driven interactive Web maps, but is an expensive and complex solution to acquire, run and maintain. WebView (an Internet extension to ArcView) was instead used to publish HealthCyberMap's ArcView project as Web client-side imagemaps. The basic WebView set-up does not support any GIS database connection, and published Web maps become disconnected from the original project. To overcome this major limitation, the author developed a simple, low-cost method for "patching" WebView to serve hypermaps with dynamic database drill-down functionality on the Web using dynamic ASP pages (Active Server Pages) to query the same metadata base used in ArcView and registered on HealthCyberMap Web server as an ODBC (Open DataBase Connectivity) data source. This solution is currently used in HealthCyberMap pilot service on the Web and can be also used to publish other interactive GIS-driven maps on the Web, e.g., maps of real-world health problems.

A strategy for maintaining the currency of the generated hypermaps and dealing with problems related to Web resource link persistence is described in Chapter 9, Section 9.6. HealthCyberMap also uses a link checker to identify any broken resource links in its metadata base (<http://healthcybermap.semanticweb.org/linkchecker.htm>).

16.1.4 Evaluation Objectives

All evaluation objectives were met successfully (Chapters 11, 12, 13 and Appendix 2). A two-method formative evaluation study of HealthCyberMap pilot service using an online user evaluation questionnaire (<http://healthcybermap.semanticweb.org/questionnaire.asp>), in addition to analysis of HealthCyberMap server transaction log, has been conducted during the period from 18 April 2002 to 1 June 2002 (45 days) with very encouraging results. The study aimed at evaluating the concepts behind the current pilot service and learning how it is being used and received by users.

Thirty-five subjects responded to the evaluation questionnaire during the 45-day period of the study. The questionnaire covered issues related to service usability, usefulness, and user satisfaction. It included a usability task and a comparative task, comparing HealthCyberMap to another approach (Visual Net). Both tasks were assessed qualitatively.

Sawmill 6.3.8 server log analysis tool was selected from among several other tools to perform the analysis of HealthCyberMap server transaction log for the same 45-day period of the formative evaluation study. HealthCyberMap server log analysis offered valuable quantitative information on server activity and traffic, the geographic provenance of visitors, and referring search engines and sites during this period, among other things.

The two chosen evaluation methods (questionnaire and server log) acted in a complementary synergistic way to provide a more complete picture about the evaluated pilot service. The results from this study helped shaping the recommendations for future work presented in Chapter 15, Section 15.12.

16.2 Novel Contribution to the Research Area

As the title of this research suggests, HealthCyberMap features a novel and unique methodology that brings together a Geographic Information System (GIS) and a clinical coding scheme for the first time for the purpose of classifying and mapping conceptual spaces occupied by collections of medical/ health Internet information resources based on their topics and other metadata attributes.

Geographic Information Systems are robust and reliable tools, optimised for handling, cross-linking and visualising data with spatial and/ or spatialised components. In GIS, data semantics and visualisation are separated for maximum flexibility, but remain tightly coupled, which corresponds to HealthCyberMap's core architecture model, making GIS a perfect choice as a classification and visualisation engine.

The clinical coding scheme provides a conceptual semantic space that GIS projects on graphical maps of the human body and its organs. GIS then maps selected medical/ health Internet resources to different semantic locations in this conceptual space (and corresponding human body maps) according to the semantics of resource topics.

The conventional geographic map metric of distance translates well into a new "semantic distance" metric on HealthCyberMap's human body maps. The "semantic distance" between two resources on these maps depends on how close (or related) the two resources are from a semantic perspective (i.e., their semantic proximity based on their medical/ health topics as determined by the underlying clinical coding scheme which preserves the semantic relations between topics). For example, a resource on “myocardial infarction” will be much closer to a resource on “angina pectoris” than to another resource on “psoriasis”.

The resultant graphical hypermaps build on humans’ spatio-cognitive abilities and the familiar human body metaphor to provide a highly meaningful way for browsing and visually querying large collections of medical/ health Internet resources.

GIS also classifies information resource counts per body region into ranges and associates each range with a colour shade or tint on HealthCyberMap’s human body maps (i.e., a choropleth rendition). This allows map users to visually spot “infogaps” (topical coverage gaps to be addressed by information providers).

HealthCyberMap’s proposed future direction to become a customisable, location-based medical/ health information service is also novel, and will provide information that is immediately relevant to users by allowing better presentation and automatic linking of the distribution of users’ location-specific health and healthcare needs and Internet resources answering them across different geographical areas.

16.3 Conclusion

The Semantic Web initiative (<http://www.w3.org/2001/sw/> and <http://semanticweb.org>), to which HealthCyberMap belongs, aims at creating a Web where information semantics (meaning and context) are represented in a form “understandable” by machines as well as by humans through proper use of metadata and ontologies (vocabularies, and resource, user and device descriptions). This will pave the way for more “intelligent” machine-to-machine communication, e.g., between electronic patient record clients and online information services like HealthCyberMap, and ultimately empower humans.

HealthCyberMap features a novel and unconventional use of GIS to map conceptual spaces occupied by collections of medical/ health information resources. Besides mapping the semantic and non-geographical aspects of these resources using suitable spatial metaphors, HealthCyberMap also collects and maps some geographical aspects of these resources like provenance.

Metadata-driven information classification and retrieval is usually associated with better precision and recall rates (compared to automated spider indexing/ free text search). Using clinical codes to reliably describe the subjects of medical/ health Web resources can further enhance metadata quality. Clinical codes can establish the semantic relationships (as defined by the underlying coding vocabulary) between related resources, automate the topical categorisation of resources (e.g., automatically classify a resource on “diabetes mellitus” under “endocrine disorders” and other relevant categories), and ensure highly relevant information retrieval.

Unlike other approaches like SHOE (Simple HTML Ontology Extension—[26, 27]), HealthCyberMap does not impose any structural changes like embedded metadata or ontology instance mark-up on the peripheral resources or their hosting servers. However, Web authors should be encouraged in the future to annotate their resources with appropriate (peripheral) metadata tags using some standard vocabulary like the qualified Dublin Core set (in a standard way to be agreed upon). As more and more Web authors begin to embed suitable metadata tags in peripheral Web resources, central cataloguing might one day become straightforward and nearly automated, except for the process of ascribing quality rating information by a human assessor.

The Web hypermaps in HealthCyberMap are client-side imagemaps with dynamic metadata base links. HealthCyberMap human body maps with their “semantic zooming” feature allow the navigation of Internet health resources by body location/ system according to ICD-9-CM codes, which act as HealthCyberMap medical/ health ontology and are used to describe resource subjects in the metadata base. The “semantic distance” between two resources on these maps depends on how close (or related) the two resources are from a semantic perspective based on the “semantic locations” of their topics within ICD-9-CM. The maps are used to locate, launch health resources on the Web, and display their bibliographic metadata records.

HealthCyberMap addresses many cyber-knowledge needs of Internet medical/ health information providers and consumers. The author believes that the visual categorisation of Internet health resources using familiar spatial metaphors for image-word association could give users a broad overview and understanding of what is available in this complex conceptual space and help them navigate it more efficiently and effectively. Topical coverage gaps can be also easily identified using the human body choropleth (shaded) maps of resource counts and addressed by information providers.

16.4 Coda

HealthCyberMap is the *first medical/ health Semantic Web project in the world* to develop a comprehensive methodology for realising a medical/ health Semantic Web information service. HealthCyberMap's collective methodology as proposed in this thesis, with its cross-fertilisation of different technologies (e.g., GIS and clinical vocabularies) is unique, and lays the foundation for a robust and effective next-generation medical/ health information service. Besides the novel and unconventional use of GIS in HealthCyberMap to map conceptual spaces occupied by medical and health information resources (based on a clinical metadata framework), GIS is also indispensable for further developing HealthCyberMap into a fully location-aware/ location-based service. The following four ingredients have been found to be pivotal to high service usefulness as evidenced in HealthCyberMap pilot service (*HealthCyberMap URIQ formula for maximising service utility*—could be also applied to any other similar services):

- **U**sability (by developing easy-to-understand, visual interfaces based on resource semantics and tailored to suit different user levels/ needs and to complement conventional textual interfaces);
- **R**elevance—always serving the right content in suitable form and format (this requires the acquisition of high quality metadata about not only information resources, but also service users, their location and their devices, coupled with "intelligent" semantic reasoning to infer information not explicitly mentioned in the metadata or user queries);
- **I**ntegrability into users' workflow/ everyday clinical practice (problem-to-knowledge linking); and
- **Q**uality, currency and maintainability of service content.

17 References

N.B.: All Web links below have been checked working in September 2002.

1. Fensel D (editor). The Semantic Web and its languages. *IEEE Intelligent Systems*. 2000;15(6):67-73. Available from: <http://www.cs.man.ac.uk/~horrocks/Publications/download/2000/faqs-on-oil.pdf>
2. Berners-Lee T, Hendler J. Nature Debates: Scientific publishing on the 'semantic web.' *Nature*. 26 April 2001. Available from: <http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm>
3. Parallel Understanding Systems Group, University of Maryland, US. The SHOE FAQ. Available from: <http://www.cs.umd.edu/projects/plus/SHOE/faq.html>
4. Berners-Lee T, Hendler J, Lassila O. Feature Article: The Semantic Web. *Scientific American*. 2001 May. Available from: <http://www.sciam.com/2001/0501issue/0501berners-lee.html>
5. Kamel Boulos MN. *A Preparatory Study on the Dermatology Virtual Branch Library of the NeLH (MSc Thesis in Medical Informatics)*. GKT School of Medicine, KCL, University of London, 2000. Available from: <http://healthcybermap.org/publications/KamelBoulos_MSc_Med_Inform_Thesis_Kings_2000.zip>
6. PubMed (National Library of Medicine). Computation of Related Articles. Bethesda, MD: National Institutes of Health. Available from: <http://www.ncbi.nlm.nih.gov/PubMed/computation.html>
7. Liu X, Altman RB. Updating a bibliography using the related articles function within PubMed. *Proc AMIA Symp*. 1998 Nov;750-4. Available from: <http://www-smi.stanford.edu/pubs/SMI_Reports/SMI-98-0725.pdf>
8. Risden K. Toward usable browse hierarchies for the Web. In H. Bullinger and J. Ziegler (editors). *Human Computer Interaction: Ergonomics and User Interfaces*. 1999;1:1098-1102. Available from: <http://www.microsoft.com/usability/UEPostings/HCI-kirstenrisden.doc>
9. Maedche A. ECML-PKDD 2001 Tutorial: Development and Applications of Ontologies. In: 12th European Conference on Machine Learning (ECML'01)/ 5th European Conference on Principles and Practice of Knowledge Discovery in

Databases (PKDD'01). Available from: <http://www.afia.polytechnique.fr/CAFE/ECML01/ontology.html>

10. Gruber T. What is an Ontology? Available from: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
11. SemanticWeb.org. Markup Languages and Ontologies. Available from: <http://www.semanticweb.org/knowmarkup.html>
12. Stoffel K, Taylor M, Hendler J. Efficient Management of Very Large Ontologies. *Proceedings of 14th National Conference on Artificial Intelligence (AAAI-97), July 1997, Providence, Rhode Island*. American Association for Artificial Intelligence (AAAI)/MIT Press, 1997. Available from: <http://www.cs.umd.edu/projects/plus/Parka/aaai97.ps>
13. Abasolo JM, Gómez M. MELISA: An ontology-based agent for information retrieval in medicine. *Proceedings of ECDL 2000 Workshop on the Semantic Web, Lisbon, Portugal (Session 3; 21 September 2000)*. Available from: <http://www.ics.forth.gr/proj/isst/SemWeb/proceedings/session3-1/html_version/index.html>
14. Berners-Lee T. Information Management: A Proposal (The Original Proposal of the WWW; March 1989, May 1990). Available from: <http://www.w3.org/History/1989/proposal.rtf>
15. Hjelm J. *Creating the Semantic Web with RDF*. New York: John Wiley & Sons, 2001
16. Staab S, Erdmann M, Maedche A, Decker S. An Extensible Approach for Modeling Ontologies in RDF(S). *Proceedings of ECDL 2000 Workshop on the Semantic Web, Lisbon, Portugal (Session 2; 21 September 2000)*. Available from: <http://www.ics.forth.gr/proj/isst/SemWeb/proceedings/session2-1/html_version/index.html>
17. Ontology Inference Layer (OIL) Home Page. Available from: <http://www.ontoknowledge.org/oil>
18. Horrocks I, van Harmelen F, Patel-Schneider P, Berners-Lee T, Brickley D, Connolly D, Dean M, Decker S, Fensel D, Hayes P, Heflin J, Hendler J, Lassila O, McGuinness D, Andrea Stein L. DAML+OIL (March 2001). Available from: <http://www.daml.org/2001/03/daml+oil-index>

19. Broekstra J, Klein M, Decker S, Fensel D, van Harmelen F, Horrocks I. Enabling knowledge representation on the Web by extending RDF Schema (10 November 2000). Available from: <http://www.ontoknowledge.org/oil/extending2.pdf>
20. Fikes R, McGuinness DL. An Axiomatic Semantics for RDF, RDF-S, and DAML+OIL. Knowledge Systems Laboratory, Stanford University (11 January 2001). Available from: <http://www.daml.org/2000/12/axiomatic-semantics.html>
21. World Wide Web Consortium (W3C). Composite Capabilities-Preferences Profile Working Group Public Home Page. Available from: <http://www.w3.org/Mobile/CCPP/>
22. Dublin Core Metadata Initiative (DCMI). Available from: <http://www.dublincore.org/>
23. Hillmann D. Using Dublin Core. 12 April 2001. Available from: <http://www.dublincore.org/documents/usageguide/>
24. Reggie – The Metadata Editor. Available from: <http://metadata.net/dstc/>
25. The UK Office for Library and Information Networking (UKOLN). DC-dot Dublin Core metadata editor. Available from: <http://www.ukoln.ac.uk/cgi-bin/dcdot.pl>
26. Hendler JA. SHOE (Simple HTML Ontology Extension) Description. Parallel Understanding Systems Group, University of Maryland, US. Available from: <http://www.cs.umd.edu/users/hendler/shoe.html> and <http://www.cs.umd.edu/projects/plus/SHOE/>
27. Heflin J, Hendler J. Dynamic Ontologies on the Web. *Proceedings of 17th National Conference on Artificial Intelligence (AAAI-2000), July-August 2000, Austin, Texas*. American Association for Artificial Intelligence (AAAI), 2000. Available from: <http://www.cs.umd.edu/projects/plus/SHOE/pubs/aaai2000.pdf>
28. Bechhofer S, Goble C. STARCH Project (Structured Terminologies for ARCHives – Managing Semi-Structured Information). The Information Management Group (IMG) at the University of Manchester, UK. Available from: <http://potato.cs.man.ac.uk/starch/starch.ppt>
29. Appleyard RJ, Malet G. A Proposal for using Metadata Encoding Techniques for Health Care Information Indexing on the WWW. *Proc AMIA Annu Fall Symp*. 1997 Oct. Available from: <http://medir.ohsu.edu/~appleyar/d004245.html>
30. Detmer WM, Barnett GO, Hersh WR. MedWeaver: Integrating Decision Support, Literature Searching, and Web Exploration using the UMLS Metathesaurus. *Proc*

AMIA Annu Fall Symp. 1997 Oct; 490. Available from: <http://www.people.virginia.edu/~wmd4n/AMIA97F.pdf>

31. Rector AL. Clinical Terminology: Why Is it so Hard? *Methods Inf Med*. 1999;38(4-5):239-52
32. Network Inference Limited, UK. Technology: Description Logics and Inference. Available from: <http://networkinference.semanticweb.org/technology.asp?id=44&menu=2>
33. McIlraith SA, Son TC, Zeng H. Semantic Web Services. *IEEE Intelligent Systems*. 2001;16(2):46-53. Available from: <http://www.computer.org/intelligent/ex2001/pdf/x2046.pdf>
34. Hendler J. Agents and the Semantic Web. *IEEE Intelligent Systems*. 2001;16(2):30-37. Available from: <http://www.cs.umd.edu/users/hendler/AgentWeb.html>
35. Gibson W. *Neuromancer*. London: HarperCollins, 1984, p. 67
36. Dodge M, Kitchin R. *Mapping Cyberspace*. London: Routledge, 2001
37. Skupin A. From Metaphor to Method: Cartographic Perspectives on Information Visualization. *Proceedings of IEEE Symposium on Information Vizualization (InfoVis), October 2000, Salt Lake City, Utah*. Available from: <http://www.geog.ucsb.edu/~sara/teaching/geo234/papers/skupin.pdf> and <http://www.geog.uno.edu/~askupin/research/infovis2000/figures/>
38. Kraak MJ, Brown A. *Web Cartography: Developments and Prospects*. London: Taylor & Francis, 2001
39. Dodge M, Kitchin R. Atlas of Cyberspace. Harlow, England: Addison Wesley, 2001. Related Web site: <http://www.cybergeography.org/atlas/atlas.html>
40. Ding J, Gravano L, Shivakumar N. Computing Geographical Scopes of Web Resources. *Proceedings of the 26th Very-Large Database (VLDB) Conference, September 2000, Cairo, Egypt*. Available from: <http://www.cs.columbia.edu/%7Egravano/Papers/2000/vldb00.pdf>
41. Tutarel O. Geographical distribution of publications in the field of medical education. *BMC Med Educ*. 2002 Apr 30;2(1):3. Available from: <http://www.biomedcentral.com/1472-6920/2/3/>
42. Staple GC. Notes on Mapping the Net: From Tribal Space to Corporate Space. *TeleGeography 1995: Global Traffic Statistics & Commentary*. Washington, DC: TeleGeography, Inc., 1995 Oct, pp. 66-73

43. Girardin L. Mapping the virtual geography of the World-Wide Web. *Proceedings of the 5th International World Wide Web Conference, May 1996, Paris, France*. Available from: <http://www.girardin.org/luc//cgv/www5/index.html>
44. Dodge M. Indexing the Web with Geography. *Mappa.Mundi Magazine – Map of the Month*. 1 November 2000. Available from: <http://mappa.mundi.net/maps/maps_017/>
45. Chen C, Thomas L, Cole J, Chennawasin C. Representing the Semantics of Virtual Spaces. *IEEE Multimedia.* 1999;6(2):54-63
46. Andrews W. Commentary: 3D visualization may succeed in some systems. *Tech News – CNET.com*. 14 November 2000. Available from: <http://news.cnet.com/news/0-1005-200-3681261.html>
47. Kamel Boulos MN, Roudsari AV, Carson ER. Health Geomatics: An Enabling Suite of Technologies in Health and Healthcare (Methodolical Review). *J Biomed Inform*. 2001;34(3):195-219. Available from: <http://www.idealibrary.com/links/doi/10.1006/jbin.2001.1015>
48. Fabrikant SI. Spatialized Browsing in Large Data Archives. *Transactions in GIS*. 2000;4(1),65-78. Available from: <http://www.geog.ucsb.edu/~sara/teaching/geo234/papers/tig99.pdf>
49. Old LJ. Web Access to Non-spatial Data using Internet Map Servers. *Proceedings of the 21st Annual ESRI International User Conference, July 2001, San Diego, California*. Available from: <http://gis.esri.com/library/userconf/proc01/professional/papers/pap237/p237.htm>
50. Terpstra P. Mapping Cyberspace with GIS. *Proceedings of the 18th Annual ESRI International User Conference, July 1998, San Diego, California*. Available from: <http://gis.esri.com/library/userconf/proc98/PROCEED/TO650/PAP615/P615.HTM>
51. Fabrikant SI. Spatial Metaphors for Browsing Large Data Archives (Ph.D. Dissertation). University of Colorado-Boulder, Department of Geography, Boulder, CO, 2000. Available from: <http://www.geog.ucsb.edu/%7Esara/html/research/diss/spatialization.html> and <http://www.geog.ucsb.edu/%7Esara/html/research/diss/sf_diss.zip>
52. The Visual Read company, UK. Graphical Read Codes Browser v6.0 Example screenshot. Available from: <http://www.visualread.com/visread/page04a_GCB_image.htm>

53. Lie HW. Navigation. In: *The Electronic Broadsheet: All the news that fits the display (MSc Thesis)*. The Media Arts and Sciences Section, School of Architecture and Planning, Massachusetts Institute of Technology, 1991. Available from: <http://www.w3.org/People/howcome/TEB/www/hwl_th_6.html>
54. Kahn P, Lenk K. Mapping Web Sites. Switzerland: RotoVision, 2001
55. Dodge M. Accessibility to Information within the Internet: How can it be Measured and Mapped? In: Janelle DG, Hodge DC (editors). *Information, Place, and Cyberspace.* Berlin: Springer Verlag, 2000, pp. 187-204. Available from: <http://www.casa.ucl.ac.uk/martin/varenius_accessibility.pdf>
56. Graham I. Introduction to HTML: ISMAP server-side programs (January 1998). Available from: <http://www.utoronto.ca/webdocs/HTMLdocs/NewHTML/serv-ismap.html>
57. McCauley JD, Navulur KCS, Engel BA, Srinivasan R. Serving GIS Data through the World Wide Web. *Proceedings of the Third International Conference Integrating GIS and Environmental Modeling, January 1996, Santa Fe, New Mexico, US*. Available from: <http://abe.www.ecn.purdue.edu/~engelb/ncgia96/engel.html>
58. Crossley D, Boston T. A Generic Map Interface to Query Geographic Information Using the World Wide Web. *Proceedings of the 4th International World Wide Web Conference, December 1995, Boston, Massachusetts, US*. Available from: <http://www.w3.org/Conferences/WWW4/Papers/australia/>
59. Federal IT Accessibility Initiative Home Page (Section 508: The Road to Accessibility). Available from: <http://www.section508.gov/>
60. Frappier J, Williams D. An overview of the National Atlas of Canada. In: Keller CP (editor). *Proceedings of the 19th International Cartographic Conference ICC99, Ottawa, Canadian Institute of Geomatics*. 1999;261-267
61. Environmental Systems Research Institute, Inc., US. ArcView Internet Map Server Frequently Asked Questions. Available from: <http://www.esri.com/software/arcview/extensions/imsfaq.html>
62. Network Inference Limited, UK. Technology: The Semantic Web. Available from: <http://networkinference.semanticweb.org/technology.asp?id=46&menu=2>

63. Leroy G, Tolle KM, Chen H. Customizable and Ontology-Enhanced Medical Information Retrieval Interfaces. *Proceedings of IMIA WG6 Triennial Conference on Natural language and Medical Concept Representation, December 1999, Phoenix, Arizona*. Available from: <http://uaeller.eller.arizona.edu/~gleroy/papers/imia-wg6.pdf>
64. Kamel Boulos MN, Roudsari AV, Gordon C, Muir Gray JA. The Use of Quality Benchmarking in Assessing Web Resources for the Dermatology Virtual Branch Library of the National electronic Library for Health (NeLH). *J Med Internet Res*. 2001;3(1):e5. Available from: <http://www.jmir.org/2001/1/e5/>
65. McKillen D. News Report: SNOMED RT® Enables Reuters Health to Categorize Medical Stories and Provide Information Specific to Clients' Interests. Available from: <http://www.snomed.org/reuters.pdf> and <http://www.snomed.org/prod/tepr_reuters00.pdf>
66. Bechhofer S, Drummond N, Goble CA. Supporting public browsing of an art gallery collections database. *Proceedings of the 11th International Conference and Workshop on Database and Expert Systems Applications (DEXA), September 2000, London, UK/Springer-Verlag LNCS*. 2000;1873. Slides/ video available from: <http://potato.cs.man.ac.uk/slides/dexa.zip>
67. Nielsen J. *Designing Web Usability: The Practice of Simplicity*. Indianapolis: New Riders, 2000
68. Kokkelink S, Schwänzl R—Dublin Core Metadata Initiative (DCMI) Architecture Working Group. Expressing Qualified Dublin Core in RDF. 29 August 2001. Available from: <http://www.mathematik.uni-osnabrueck.de/projects/dcqual/qual21.3.1/> and <http://dublincore.org/documents/2001/08/29/dcq-rdf-xml/>
69. World Wide Web Consortium (W3C) Date and Time Formats. Available from: <http://www.w3.org/TR/NOTE-datetime>
70. Beckett D, Miller E, Brickley D—Dublin Core Metadata Initiative (DCMI). An XML Encoding of Simple Dublin Core Metadata. 1 December 2000. Available from: <http://dublincore.org/documents/2000/11/dcmes-xml/>
71. Dublin Core Metadata Initiative (DCMI). DC type element schema. Available from: <http://dublincore.org/2000/03/13/dctype#>

72. World Wide Web Consortium (W3C). Describing and retrieving photos using RDF and HTTP. 19 April 2002. Available from: <http://www.w3.org/TR/photo-rdf/>
73. What is Protégé-2000? Available from: <http://protege.stanford.edu/whatis.html>
74. Li Q, Shilane P, Noy NF, Musen MA. Ontology Acquisition from On-line Knowledge Sources. *Proc AMIA Symp*. 2000;497-501. Available from: <http://www.amia.org/pubs/symposia/D200524.PDF>
75. Noy NF, Sintek M, Decker S, Crubézy M, Fergerson RW, Musen MA. Creating Semantic Web Contents with Protégé-2000. *IEEE Intelligent Systems* (computer.org/intelligent). 2001;16(2):60-71 (March/April 2001). Available from: <http://protege.stanford.edu/publications/ProtegeForSemanticWeb.pdf>
76. Alexaki S, Christophides V, Karvounarakis G, Plexousakis D. On Storing Voluminous RDF Descriptions: The case of Web Portal Catalogs. In: 4th International Workshop on the the Web and Databases (WebDB2001 - in conjunction with ACM SIGMOD/ PODS) Available from: <http://139.91.183.30:9090/RDF/publications/webdb2001.pdf>
77. Karvounarakis G. RDF Query Languages: A state-of-the-art. Available from: <http://139.91.183.30:9090/RDF/publications/state.html>
78. Berners-Lee T. What the Semantic Web can represent (The Semantic Web and Relational Databases). 17 September 1998. Available from: <http://www.w3.org/DesignIssues/RDFnot.html>
79. Marchiori M, Saarela J. Query + Metadata + Logic = Metalog. In: *W3C QL'98 - The Query Languages Workshop, December 1998, Boston, Massachusetts*. Available from: <http://www.w3.org/TandS/QL/QL98/pp/metalog>
80. Decker S, Brickley D, Saarela J, Angele J. A Query and Inference Service for RDF. In: *W3C QL'98 - The Query Languages Workshop, December 1998, Boston, Massachusetts*. Available from: <http://www.w3.org/TandS/QL/QL98/pp/queryservice.html>
81. Magkanaraki A, Karvounarakis G, Anh TT, Christophides V, Plexousakis D. Ontology Storage and Querying. Foundation for Research and Technology Hellas, Institute of Computer Science Information Systems Laboratory. Technical Report No 308. April 2002. Available from: <http://139.91.183.30:9090/RDF/publications/tr308.pdf>

82. UK Department of Health. The Expert Patient: A New Approach to Chronic Disease Management for the 21st Century. September 2001. Available from: <http://www.doh.gov.uk/healthinequalities/ep_report.pdf>
83. Centers for Disease Control and Prevention (CDC) - National Center for Health Statistics, US. Classification of Diseases (ICD-9-CM). Available from: <http://www.cdc.gov/nchs/icd9.htm>
84. Health On the Net Foundation. HON Code of Conduct (HONcode) for medical and health Web sites: Principles. Version 1.6 April 1997. Available from: <http://www.hon.ch/HONcode/Conduct.html>
85. ZEBRIS Geoinformation Systems and Consulting, München, Germany. WebView - The Internet Extension for ArcView GIS. Available from: <http://www.zebris.com/english/main_webview.htm>
86. GeoHealth, Inc., US. BodyViewer Extension for ArcView GIS. Available from: <http://www.geohealth.com/bodyviewer.html>
87. Spence R. *Information Visualization*. Essex: ACM Press, 2001
88. National Library of Australia. Managing Web Resources for Persistent Access. Available from: <http://www.nla.gov.au/guidelines/2000/persistence.html> and <http://www.nla.gov.au/guidelines/2000/webresources.html>
89. Gorman P. Information Needs in Primary Care: A Survey of Rural and Nonrural Primary Care Physicians. In: Patel V, Rogers R, Haux R (editors). *Proceedings of the 10th World Congress on Medical Informatics, September 2001, London, UK (Medinfo 2001—CD-ROM Edition)*. Amsterdam: IOS Press, 2001
90. Trochim WMK, Hovers D. Evaluating Websites/ Mapping Student Views of the Benefits of a Course Web Site. 1996. Available from: <http://trochim.human.cornell.edu/webeval/webeval.htm>
91. Crawford J. *Evaluation of Library and Information services (Second edition).* London: Aslib/ imi, 2000
92. Felton D, Heathfield H, Clamp S, Gowing E. Evaluation of the South Staffordshire pan-community EHR-demonstrator programme. In: J. Bryant (editor). *Proceedings of Healthcare Computing 2002, March 2002, Harrogate, UK (HC2002-CD-ROM Edition).* Guilford: The British Computer Society, Health Informatics Committee. 2002
93. Gillham B. Developing A Questionnaire. London: Continuum, 2000

94. QRC Division of Macro International Inc. Evaluation Design/Planning and Methodology for the NIH Web Site—Phase I (Evaluation and Performance Measures Toolkit). Available from: <http://irm.cit.nih.gov/itmra/weptest/acknow.htm>
95. Tec-Ed, Inc, US. Assessing Web Site Usability from Server Log Files. White Paper. December 1999. Available from: <http://www.teced.com/PDFs/whitepap.pdf>
96. Urquhart C, Yeoman A, Cooper J, Wailoo A. NeLH Pilot Evaluation Project - Final Report. University of Wales Aberystwyth and University of Sheffield for NHSIA, National electronic Library for Health. September 2001. Available from: <http://www.nhsia.nhs.uk/nelh/background/aber.doc>
97. Mazzi CP, Kidd M. A Framework for the Evaluation of Internet-based Diabetes Management. *J Med Internet Res.* 2002;4(1):e1. Available from: <http://www.jmir.org/2002/1/e1/>
98. Bohmann K. Useful Features are Not Always Usable (online article). 3 July 2000. Available from: <http://www.bohmann.dk/articles/useful_features_are_not_always_usable.html>
99. Nielsen J. Heuristic Evaluation. In Nielsen J, Mack RL (editors). *Usability Inspection Methods*. New York: John Wiley & Sons, Inc., 1994, pp. 25-62
100. Batagelj Z, Simšic J, Cesen T. Web Site Evaluation (PowerPoint presentation). Presented at the International Conference on Methodology and Statistics, September 1999, Preddvor. Available from: <http://surveys.over.net/wsep99.ppt>
101. Barna A. Comparative Task Analysis of Two E-Commerce Sites. Available from: <http://badco.ws/187/companal/index.html>
102. MacEachren AM. *How Maps Work: Representation, Visualization, and Design*. New York: The Guilford Press, 1995, pp. 256-266
103. Muir Gray JA. The Resourceful Patient. Oxford: eRosetta Press, 2002. Available from: <http://www.resourcefulpatient.org>
104. Goldberg J. Why web usage statistics are (worse than) meaningless. 1995. Available from: <http://www.goldmark.org/netrants/webstats/>
105. McDunn RA. Web Server Log File Analysis - Basics. Stanford Linear Accelerator Centre, US. 23 February 2001. Available from: <http://www-group.slac.stanford.edu/techpubs/logfiles/info.html>

106. Ferguson K. Web traffic analysis could save your site. *ZDNet Tech Update*. 22 August 2001. Available from: <http://www.zdnet.com/filters/printerfriendly/0,6061,2806111-92,00.html>

107. Bannan KJ. Analyze This! *PC Magazine*. 27 November 2001. Available from: <http://www.pcmag.com/print_article/0,3048,a=17510,00.asp>

108. Kagolovsky Y, Mohr JR. A new approach to the concept of "relevance" in information retrieval (IR). In: Patel V, Rogers R and Haux R (editors). *Proceedings of the 10th World Congress on Medical Informatics (Medinfo 2001)*. Amsterdam, The Netherlands: IOS Press, 2001 Sep;10(Pt 1):348-52

109. Glander-Höbel C, Voigt K, Welzl G. Comparing search engines with respect to environmental chemistry using statistical and mathematical evaluation methods. *Proceedings of Online Information 2001 Conference, December 2001, London, UK*. Available from: <http://www.online-information.co.uk/proceedings/online/2001/glander_hobel_c.asp>

110. Marchionini G. Evaluating digital libraries: a longitudinal and multifaceted view. *Library Trends*. 2000 Fall;49(2):304-333. Available from: <http://www.ils.unc.edu/~march/perseus/lib-trends-final.pdf>

111. Harter SP, Hert CA. Evaluation of information retrieval systems: approaches, issues, and methods. In: Williams ME (editor). *Annual Review of Information Science and Technology (ARIST)*. 1997;32:3–79

112. Cleverdon CW, Keen EM. *Factors determining the performance of indexing systems, volume 1: design, volume 2: test results*. Cranfield, UK: Aslib Cranfield Research Project, 1966

113. Wu G, Li J. Comparing web search engine performance in searching consumer health information: evaluation and recommendations. *Bulletin of the Medical Libraries Association*. 1999 Oct;87(4):456–61

114. Green D. The evolution of web searching. *Online Information Review*. 2000 Mar-Apr;24(2):124–37. Available from: <http://general.rau.ac.za/infosci/information/studyguide/First/unit_5/green.htm>

115. Fuller R, de Graaff JJ. Measuring User Motivation from Server Log Files. Microsoft Usability Conference—Designing for the Web: Empirical Studies, October 1996, Microsoft Campus. Available from: <http://www.microsoft.com/usability/webconf/fuller/fuller.htm>

116. Haigh S, Megarity J. Measuring Web Site Usage: Log File Analysis. Network Notes #57. Information Technology Services, National Library of Canada. August 1998. Available from: <http://www.nlc-bnc.ca/publications/1/p1-256-e.html>

117. Tillett LS. 'A 24-Hour Focus Group' - Sites dig into search queries to learn customer preferences. *Internetweek.com*. 7 April 2000. Available from: <http://www.internetwk.com/lead/lead040700-1.htm>

118. Scholtz J, Laskowski S, Downey L. Developing Usability Tools and Techniques for Designing and Testing Web Sites. *Proceedings of the 4th Conference on Human Factors & the Web, June 1998, Basking Ridge, New Jersey, US*. Available from: <http://www.research.att.com/conf/hfweb/proceedings/scholtz/index.html>

119. Brennan PF, Friede A. Public Health and Consumer Uses of Health Information. In: Shortliffe EH, Perreault LE (editors). *Medical Informatics - Computer Applications in health Care and Biomedicine*. New York: Springer-Verlag, 2001, pp. 397-420

120. Geobytes, Inc., US/ Australia. Geobytes' Technology Applications. 15 April 2002. Available from: <http://www.geobytes.com/applications.htm>

121. Löytönen M. Mobile devices and GIS in health – new opportunities and threats. *Proceedings of the First European Conference for Geographic Information Sciences in Public Health, September 2001, Sheffield, UK*. Available from: <http://gis.sheffield.ac.uk/newconferencepages/abstracts.htm>

122. Berkowitz T, Lopez X. Enhancing Mobile Applications with Location-Based Services. An Oracle Business White Paper. 2001. Available from: <http://www.oracle.com/ip/deploy/database/oracle9i/collateral/cm_lbs_bwp.pdf>

123. van Leeuwen A. Geo-targeting on IP Address - Pinpointing Geolocation of Internet Users. *GeoInformatics*. July/August 2001. Available from: <http://www.geoinformatics.com/issueonline/issues/2001/07_2001/pdf_07_2001/28_31_iptar.pdf>

124. Hogeweg M. Relocation Based Services. *GeoInformatics*. September 2001. Available from: <http://www.geoinformatics.com/issueonline/issues/2001/09_2001/pdf_09_2001/13_hogeweg.pdf>

125. Moore C. Location-based services called key to unplugged Internet. *InfoWorld*. 12 February 2001. Available from: <http://www.itworld.com/Tech/2424/IW010212hnmobilecomm/>

126. Kamel Boulos MN. HealthCyberMap Pilot Web Site: Customisation - Detecting User's Location. Available from: <http://healthcybermap.semanticweb.org/ip.htm>

127. Hall K, Rodgers M. Content tagging strategies (technical report). Cambridge, MA: Giga Information Group, Inc., 2001. Available from: <http://www.gigaweb.com/search/1,2197,results,00.html?iContentID=253191>

128. SNOMED International Authority. SNOMED uses. Northfield, IL: College of American Pathologists, SNOMED International, 2000. Available from: <http://www.snomed.org/uses_txt.html> and <http://www.snomed.org/press/SNOMED%20CT%20First%20Release.pdf>

129. Bodenreider O, Nelson SJ, Hole WT, Chang HF. Beyond Synonymy: Exploiting the UMLS Semantics in Mapping Vocabularies. Bethesda, MD: National Institutes of Health (National Library of Medicine, MeSH Web Site), November 2001. Available from: <http://www.nlm.nih.gov/mesh/beyond.html>

130. Cimino JJ. Terminology Tools: State of the Art and Practical Lessons. *Methods Inf Med*. 2001 Aug;40(4):298-306

131. Bechhofer SK, Goble CA, Rector AL, Solomon WD, Nowlan WA. Terminologies and terminology servers for information environments. *Proceedings of the IEEE 8th International Conference on Software Technology and Engineering Practice (STEP '97), London, UK*. 1997 Jul;35-42. Available from: <http://citeseer.nj.nec.com/354766.html>

132. Spackman KA. Normal forms for description logic expressions of clinical concepts in SNOMED RT. *Proc AMIA Symp*. 2001 Nov;627-31. Available from: <http://www.amia.org/online_html/S07.HTM>

133. Chute CG, Elkin PL, Sheretz DD, Tuttle MS. Desiderata for a clinical terminology server. *Proc AMIA Symp*. 1999 Nov;42-6. Available from: <http://www.amia.org/pubs/symposia/D005782.PDF>

134. van Harmelen F, Broekstra J, Fluit C, ter Horst H, Kampman A, van der Meer J, Sabou M. Ontology-based Information Visualisation. Presented at the workshop on Visualisation of the Semantic Web (VSW '01), September 2001, London, UK, in conjunction with the 5th International Conference on Information Visualisation.

Available from: <http://www.aidministrator.nl/users/development/files/VSW01.pdf>

135. Lambrou A. Understanding MeSH: Medical Subject Headings. Montreal, Canada: McGill University, Health Sciences Library. 29 November 2001. Available from: <http://www.health.library.mcgill.ca/eguides/mesh.htm>

136. NHS Information Authority, UK. NHS Headings Project. Available from: <http://www.nhsia.nhs.uk/headings/>

137. ISO/ IEC 13250 Topic Maps. 3 December 1999. Available from: <http://www.y12.doe.gov/sgml/sc34/document/0129.pdf>

Appendix 1: Example ASP Page from HealthCyberMap

The SQL query is highlighted in red in the code listing below. Note how all three DC subject fields in each resource record are searched for ICD-9-CM codes (E codes in this example—<http://healthcybermap.semanticweb.org/bodyviewer/e-codes.asp>):

`(((hcm.[`**`dc:Subject:1`**`]) Like 'E%')) OR (((hcm.[`**`dc:Subject:2`**`]) Like 'E%')) OR (((hcm.[`**`dc:Subject:3`**`]) Like 'E%')).`

(N.B.: in ASP, the % sign is the equivalent of the * wildcard used in standard SQL.)

Also note the code for paginating query results (highlighted in green in the code listing below).

Code listing

```
<%@Language=VBScript%>
<%Option Explicit%>
<%Response.Buffer = True%>
<!--#include virtual="adovbs.inc"-->
<html>
<TITLE>HealthCyberMap: E Codes</TITLE>
<link rel="stylesheet" type="text/css" href="/style.css" /><BODY
BGCOLOR="#FFFFFF">
<script>
function toggle(e) {
if (e.style.display == "none") {
e.style.display = "";
} else {
e.style.display = "none";
}
}
</script>
<body>
<a href="javascript:history.go(-1)">< Back</a>
<hr color="#0071C0">
<%
Dim Page_Size
Dim Current_Page
Dim Page_Count
Dim conn, RS, SQL
Page_Size = 10 'this is where you set the # of records displayed per page
If Request("Page") = "" Then
Current_Page = 1
Else
Current_Page = CInt(Request("Page"))
End If
Set conn = Server.CreateObject("ADODB.Connection")
Set RS=Server.CreateObject("ADODB.RecordSet")
conn.Open "hcm","",""
RS.CursorLocation = adUseClient
RS.PageSize = Page_Size
SQL = "SELECT hcm.[ID], hcm.[dc:Creator], hcm.[dc:Title],
hcm.[dc:Subject:1], hcm.[dc:Subject:2], hcm.[dc:Subject:3],
hcm.[dc:Description], hcm.[dc:Publisher], hcm.[dc:Date], hcm.[dc:Type],
hcm.[dc:Identifier], hcm.[dc:Language], hcm.[dc:Coverage],
hcm.[hcm:Location:city], hcm.[hcm:Location:country], hcm.[hcm:Quality],
hcm.[hcm:Comment] FROM hcm WHERE (((hcm.[dc:Subject:1]) Like 'E%')) OR
(((hcm.[dc:Subject:2]) Like 'E%')) OR (((hcm.[dc:Subject:3]) Like 'E%')) "
RS.Open SQL, conn, adOpenStatic, adLockReadOnly, adCmdText
On Error Resume Next
Page_Count = RS.PageCount
If 1 > Current_Page Then Current_Page = 1
If Current_Page > Page_Count Then Current_Page = Page_Count
RS.AbsolutePage = Current_Page
```

```
Do While RS.AbsolutePage = Current_Page And Not RS.EOF
%>
<b>Author(s)/ Creator(s):</b> <%=RS("dc:Creator")%><BR>
<b>Resource Title:</b> <%=RS("dc:Title")%><BR>
<b>Type/ Category:</b> <%=RS("dc:Type")%><BR>
<b>Language:</b> <%=RS("dc:Language")%><BR>
<b><font color="red">URI:</font></b> <A HREF="<%=RS("dc:Identifier")%>"
Title="<%=RS("dc:Identifier")%>" Target="_blank">Launch in new
window</A><BR>
<div style="cursor: hand" onclick="toggle(document.all.a<%=RS("ID")%>);">
<FONT COLOR=#00008b><B>Show/ Hide Details</B></FONT></div>
<span style="display: None" id=a<%=RS("ID")%>>
<b>ICD-9-CM Subject Code 1:</b> <%=RS("dc:Subject:1")%><BR>
<b>ICD-9-CM Subject Code 2:</b> <%=RS("dc:Subject:2")%><BR>
<b>ICD-9-CM Subject Code 3:</b> <%=RS("dc:Subject:3")%><BR>
<b>ICD-9-CM Subject Description(s):</b> <%=RS("dc:Description")%><BR>
<b>Publisher(s):</b> <%=RS("dc:Publisher")%><BR>
<b>Author(s) and/or Publisher(s) City:</b> <%=RS("hcm:Location:city")%><BR>
<b>Author(s) and/or Publisher(s) Country:</b>
<%=RS("hcm:Location:country")%><BR>
<b>Coverage:</b> <%=RS("dc:Coverage")%><BR>
<b>Date of Last Update/ Review:</b> <%=RS("dc:Date")%><BR>
<b>Resource Quality Information:</b> <%=RS("hcm:Quality")%><BR>
<b>Comment(s):</b> <%=RS("hcm:Comment")%><BR>
</span>
<a
href="http://healthcybermap.semanticweb.org/icd.asp?SearchText=<%=RS("dc:Sub
ject:1")%>">Find all resources having the same primary subject as this
one</a><BR>
<hr color="#0071C0">
<%
RS.MoveNext
Loop
RS.Close
Set RS = Nothing
conn.Close
Set conn = Nothing
Response.Write "<td colspan=""4"" align=""center"">"
If Current_Page = 1 Then
Response.Write "<font color=""silver"" & size=""2"">" & "First </font><font
size=""2""> |</font> "
End If
If Current_Page >= 2 Then
Response.Write "<a href=""e-codes.asp?Page=1"
Response.Write """><font size=""2""><< First</font></a><font size=""2"">
|</font> " & vbCrLf
End If
If Current_Page >= Page_Count Then
Response.Write "<font color=""silver"" Size=""2"">Next >></font><font
size=""2""> | </font>"
End If
If Current_Page < Page_Count Then
Response.Write "<a href=""e-codes.asp?Page="
Response.Write Current_Page + 1
Response.Write """><font size=""2"">Next ></font></a>" & " <font
size=""2"">|</font>" & vbCrLf
End IF
If Current_Page <> 1 Then
Response.Write "<a href=""e-codes.asp?Page="
Response.Write Current_Page - 1
Response.Write """><font size=""2"">< Previous </font></a><font
size=""2"">  |</font> " & vbCrLf
Response.Write " " & vbCrLf
End If
If Current_Page = 1 Then
Response.Write "<font color=""silver"" & size=""2"">" & "< Previous
</font><font size="""">|</font> "
End If
If Current_Page <> Page_Count Then
Response.Write "<a href=""e-codes.asp?Page="
Response.Write Page_Count
Response.Write """><font size=""2"">Last >></font></a>" & vbCrLf
End If
```

```
If Current_Page >= Page_Count Then
Response.Write "<font size=""2"" color=""silver"">Last</font>" & "</font>"
End If
%>
<br>
Page <%=Current_Page%> of <%=Page_Count%>
<br>
&copy; 2001, 2002 HealthCyberMap.org. All Rights Reserved.
</body>
</html>
```

Appendix 2: Formative Evaluation Questionnaire of HealthCyberMap Pilot Service

Questions and Web Layout

Formative Evaluation Questionnaire of HealthCyberMap Pilot Implementation

This is a questionnaire to evaluate HealthCyberMap pilot implementation (this is a formative or initial evaluation of concepts in their infancy rather than of a full-blown service). We request anonymous information about your background and ask some evaluation questions. Responding to the questionnaire should take about 30 minutes to complete. Thank you very much for your time and input.

About HealthCyberMap

HealthCyberMap (http://healthcybermap.semanticweb.org) is a Semantic Web project that aims at mapping parts of medical/health information resources in cyberspace in novel semantic ways to improve their retrieval and navigation. The Semantic Web (http://www.w3.org/2001/sw/ and http://www.semanticweb.org) aims to be the next-generation World Wide Web by giving machine-readable semantics and context to the currently presentation-based Web pages so that, for example, a search for resources on Mr. Wood does not return results on tree wood. Follow this link to learn more about HealthCyberMap.

HON CODE We subscribe to the HON code principles. Verify here.

You may also wish to read our privacy statement.

Questions (You may leave blank any question you don't want to answer)

Question 1. Gender:

- ○ Male
- ○ Female

Question 2. Which of the following categories includes your age:

- ○ Under 18
- ○ 18 - 25
- ○ 26 - 35
- ○ 36 - 45
- ○ 46 - 65
- ○ Over 65

Question 3. What is the highest level of education you completed:

- ○ School
- ○ University/College
- ○ Postgraduate
- ○ Vocational training

Question 4. Which of the following best describes you:

- ○ Member of the general public seeking health-related information
- ○ Member of the general public touched by disease
- ○ General practitioner
- ○ Hospital doctor
- ○ Nurse
- ○ PAM/AHP (Allied Health Professional)/other healthcare professional
- ○ Scientist/researcher
- ○ Medical librarian
- ○ Other

Question 5. How many hours per week, including work and home, do you currently use the Internet:

- ○ Less than 5 hours
- ○ More than 5 hours

Question 6. Regularly used a computer:

- ○ Not at all
- ○ Few weeks
- ○ 2-6 months
- ○ 6-24 months
- ○ >2 years

Question 7. Regularly used the Internet using a browser (e.g., Netscape, Internet Explorer):

- ○ Not at all
- ○ Few weeks
- ○ 2-6 months
- ○ 6-24 months
- ○ >2 years

Question 8. Do you have a working personal computer (PC) in your home:

- ○ Yes
- ○ No

Question 9. From which location are you most likely to use the Web:

- ○ Home
- ○ Work
- ○ School
- ○ Library
- ○ Other

Question 10. What type of Internet connection do you have currently (if you use more than one type of connection in different places, select the one you use most):

- ○ Modem
- ○ ISDN/ADSL/Broadband/Always connected

Question 11. What web browser do you use most frequently:

- ○ Microsoft Internet Explorer
- ○ Netscape Navigator or Communicator
- ○ Other

Question 12. Which version of the browser do you use most frequently:

- ○ Version 3.x
- ○ Version 4.x
- ○ Version 5.x
- ○ Version 6.x
- ○ Other

Question 13. The Desktop Area/Screen Resolution you *most commonly* use is:

- ○ 640 by 480 pixels
- ○ 800 by 600 pixels
- ○ 1024 by 768 pixels or more

N.B. We have detected that your current screen resolution is set to 1024x768.

Question 14. Which operating system are you currently running on your primary (most frequently used) PC:

- ○ Windows
- ○ Macintosh
- ○ Unix/Linux
- ○ WebTV
- ○ Other

Question 15. Do you consider the Internet an important source of reliable medical/health-related information:

- ○ Not at all
- ○ To some extent
- ○ Definitely

Question 16. How often have you accessed HealthCyberMap Web site over the past 3 months:

- ○ Daily
- ○ Weekly
- ○ Monthly
- ○ This is the first time I have accessed the Web site

Question 17. How did you hear about HealthCyberMap (Choose the one that best applies to the first way you learned about HealthCyberMap):

- ○ I listened to a presentation/read a paper on the project
- ○ A colleague told me about it
- ○ I received a notice through e-mail/e-mail list
- ○ I came across HealthCyberMap on the Internet (search engine, link from other site)
- ○ I am a personal friend or relative of researcher involved in this project
- ○ Other

Question 18. As a pilot system, HealthCyberMap does not include an exhaustive coverage of all medical/health conditions. Based on what you have seen so far, if HealthCyberMap was fully developed to provide a more complete coverage, would you use it as an information portal and one-stop access point to a variety of medical and health-related resources:

- ○ Yes
- ○ No

Question 19. HealthCyberMap concepts and interfaces meet my information retrieval and navigation needs (not topical coverage) better than other medical/health information portals/gateways:

- ○ Strongly disagree
- ○ Disagree
- ○ Neither disagree nor agree
- ○ Agree
- ○ Strongly agree

Question 20. (Only if you did not agree with the above statement) At this time, what other medical/health portals/gateway are better than HealthCyberMap regarding information retrieval and navigation (not topical coverage)

Question 21. I prefer to use a conventional search engine (e.g., Google) to gather information or to have a librarian, staff member, or family member gather information for me:

- ○ True
- ○ False

Question 22. How well does each of the six information interfaces featured in the current HealthCyberMap pilot perform its intended purpose? (Please rate each of the information interfaces from "not at all well" to "extremely well" considering aspects like ease of access, use and navigation, and logical arrangement of information on the pages of concerned interfaces)

Question 22.1. HealthCyberMap world maps of Web resources classified according to geographic provenance:

- ○ Not at all well
- ○ Slightly well
- ○ Somewhat well
- ○ Very well
- ○ Extremely well

Question 22.2. HealthCyberMap BodyViewer maps of Web resources classified according to medical/health topics:

- ○ Not at all well
- ○ Slightly well
- ○ Somewhat well
- ○ Very well
- ○ Extremely well

Question 22.3. HealthCyberMap textual resource index using ICD-9-CM top-level categories of medical/health topics:

- ○ Not at all well
- ○ Slightly well
- ○ Somewhat well
- ○ Very well
- ○ Extremely well

Question 22.4. HealthCyberMap map of Web resources classified according to resource type, e.g., audio-visual material, interactive resources, digital atlases, fact sheets, guidelines, etc.:

- ○ Not at all well
- ○ Slightly well
- ○ Somewhat well
- ○ Very well
- ○ Extremely well

Question 22.5. HealthCyberMap Web resources classified according to resource language:

- ○ Not at all well
- ○ Slightly well
- ○ Somewhat well
- ○ Very well
- ○ Extremely well

Question 22.6. HealthCyberMap semantic subject search engine (about this engine):

- ○ Not at all well
- ○ Slightly well
- ○ Somewhat well
- ○ Very well
- ○ Extremely well

Question 23. Did you find the meanings of the human body icons on HealthCyberMap BodyViewer maps familiar to you:

- ○ Yes
- ○ Somewhat (not always)
- ○ No

Question 24. Selecting the right words for search, and finding the right areas/icons to click on the maps was:

- ○ Very difficult
- ○ Difficult
- ○ Of moderate (acceptable) difficulty
- ○ Easy
- ○ Very easy

Question 25. How important/useful to you is each of the six information interfaces featured in the current HealthCyberMap pilot (Please rate each of the information interfaces from "not at all important" to "extremely important")

Question 25.1. HealthCyberMap world maps of Web resources classified according to geographic provenance:

- ○ Not at all important
- ○ Slightly important
- ○ Somewhat important
- ○ Very important
- ○ Extremely important

Question 25.2. HealthCyberMap BodyViewer maps of Web resources classified according to medical/health topics:

- ○ Not at all important
- ○ Slightly important
- ○ Somewhat important
- ○ Very important
- ○ Extremely important

Question 25.3. HealthCyberMap textual resource index using ICD-9-CM top-level categories of medical/health topics:

- ○ Not at all important
- ○ Slightly important
- ○ Somewhat important
- ○ Very important
- ○ Extremely important

Question 25.4. HealthCyberMap map of Web resources classified according to resource type, e.g., audio-visual material, interactive resources, digital atlases, fact sheets, guidelines, etc.:

- ○ Not at all important
- ○ Slightly important
- ○ Somewhat important
- ○ Very important
- ○ Extremely important

Question 25.5. HealthCyberMap Web resources classified according to resource language:

- ○ Not at all important
- ○ Slightly important
- ○ Somewhat important
- ○ Very important
- ○ Extremely important

Question 25.6. HealthCyberMap semantic subject search engine (about this engine):

- ○ Not at all important
- ○ Slightly important
- ○ Somewhat important
- ○ Very important
- ○ Extremely important

Question 26. Do you find HealthCyberMap map interface help (and other instructions/help hints around the site) adequate:

- ○ Yes, HealthCyberMap is easy to use once you understand how it works; the online help and instructions provided were useful
- ○ HealthCyberMap is very easy to use; I didn't need any help (or made very little use of it)
- ○ No, I need more help to locate the information I need on HealthCyberMap Web site

Question 27. The speed at which HealthCyberMap maps load on my Internet connection is:

- ○ Reasonably fast
- ○ Slow

Question 28. Now, your task is to find information resources about specific diseases using HealthCyberMap. Choose at least one disease from the following list: *prostate cancer*, *Crohn's disease*, *Guillain-Barre syndrome*, *amyotrophic lateral sclerosis*, *diabetes mellitus*, *Mitral valve prolapse*, *ectopic pregnancy* (or any other medical/health topic(s) of your choice). On the average (if you tried finding resources on more than one disease), how successful were you in completing this task:

- ○ Successful from the first attempt
- ○ Successful after one or more failed attempts
- ○ Not successful

Question 29. What interface did you use most of the time to accomplish the above task:

- ○ BodyViewer maps of Web resources classified according to medical/health topics
- ○ HealthCyberMap textual resource index using ICD-9-CM top-level categories of medical/health topics
- ○ HealthCyberMap semantic subject search engine (advanced keyword search)
- ○ All three interfaces were equally useful

Question 30. The pointers to information resources returned by HealthCyberMap were *relevant* to my queries (i.e., to search words I entered or map areas I have clicked):

- ○ Strongly agree
- ○ Agree
- ○ Neither agree nor disagree
- ○ Disagree
- ○ Strongly disagree

Question 31. HealthCyberMap pointers to information resources are of good *quality*, accurate, up-to-date and useful:

- ○ Strongly agree
- ○ Agree
- ○ Neither agree nor disagree
- ○ Disagree
- ○ Strongly disagree

Question 32. When I look for information in HealthCyberMap I get overloaded quickly with too much detail (presentation format):

- ○ Not at all a problem
- ○ A slight problem
- ○ A moderate problem
- ○ A significant problem
- ○ An extreme problem

Question 33. Do you think visual maps are a useful addition to/complementary improvement over conventional text-based Web portal interfaces:

- ○ Yes, definitely
- ○ Could be an improvement (but not always)
- ○ No, visual interfaces are useless

Question 34. Compared to the mapping approach of map.net, do you think (e.g., regarding the use of familiar metaphors and presentation format of resource details and links):

- ○ HealthCyberMap approach is superior
- ○ HealthCyberMap approach is somewhat better
- ○ HealthCyberMap approach is about the same
- ○ HealthCyberMap approach is somewhat worse
- ○ HealthCyberMap approach is inferior
- ○ Do not have an opinion

Question 35. How important/ useful do you think are the proposed HealthCyberMap future interfaces/possibilities/concepts

Question 35.1. Multi-axial classification of resources based on two or more Dublin Core metadata elements:

- ○ Not at all important
- ○ Slightly important
- ○ Somewhat important
- ○ Very important
- ○ Extremely important

Question 35.2. HealthCyberMap customisation/detecting user location for location-based customisation:

- ○ Not at all important
- ○ Slightly important
- ○ Somewhat important
- ○ Very important
- ○ Extremely important

Question 35.3. Problem to knowledge linking (clinical codes as knowledge hooks):

- ○ Not at all important
- ○ Slightly important
- ○ Somewhat important
- ○ Very important
- ○ Extremely important

Question 35.4. Mapping health problems in HealthCyberMap and identifying information needs and gaps:

- ○ Not at all important
- ○ Slightly important
- ○ Somewhat important
- ○ Very important
- ○ Extremely important

Question 36. Do you think it would be useful in a future implementation of HealthCyberMap to also organise information resources by user needs/intended primary audience as patients, health professionals, or basic researchers:

- ○ Yes
- ○ No

Question 37. Were there any parts of the service that you found especially helpful? What do you like most about HealthCyberMap and why?

Question 38. Were there any parts of the service that you found especially difficult to use or understand? What do you dislike most about HealthCyberMap and why?

Question 39. Should resources (if available) be invested to continue developing and implementing HealthCyberMap? Why or why not?

Question 40. What are your suggestions or comments about what would make the service better? (e.g., I would like the following added to, changed in, or deleted from HealthCyberMap)

Question 41. Finally we would like to hear your thoughts about this questionnaire. For example, are we asking the right questions? Are we asking the questions in the right way? Your feedback will help us design better questionnaires in the future:

Please submit your responses by clicking the "Save" button below. Again, thank you very much for your time and input.

Save Clear

Back to Home Page

Underlying Code Listing

The database `INSERT INTO` SQL statement is highlighted in red in the code listing of <http://healthcybermap.semanticweb.org/questionnaire.asp> below. The database name is `questionnaire`(.mdb) and the table name is `Results`. To help respondents answer question 13, the online questionnaire form has a built-in JavaScript that displays a respondent's current screen resolution. The code for this script is highlighted in green in the code listing below.

For string database fields in the code listing below (e.g., fields accepting free-text answers) where an apostroph (`'`) could be one of the letters making the value of such fields (e.g., a respondent typing "I didn't have any problems with HealthCyberMap"), an error will occur on submitting the form and attempting to save it to the database on HealthCyberMap server, unless the apostroph is replaced by a double apostroph (`''`). The author used the `Replace` VBScript function for this purpose, for example,

`strsuggest_comments = Replace(strsuggest_comments, "'", "''")`.

```
<!--#include virtual="adovbs.inc"-->
<html>
<head>
<meta HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=windows-1252">
<meta HTTP-EQUIV="Content-Language" CONTENT="en-uk">
<title>Formative Evaluation Questionnaire of HealthCyberMap Pilot Service</title>
<link rel="stylesheet" type="text/css" href="http://www.healthcybermap.org/style.css"
/>
<script language="JavaScript">
<!-- Begin
function Start(page)
{
OpenWin = this.open(page, "CtrlWindow",
"toolbar=no,menubar=no,location=no,scrollbars=yes,resizable=no,width=500,height=440");
}
// End -->
</script>
</head>
<body>
<h3>Formative Evaluation Questionnaire of HealthCyberMap Pilot Implementation</h3>
<p ALIGN="left">This is a questionnaire to evaluate HealthCyberMap pilot
implementation (this is a formative or initial evaluation of concepts in their infancy
rather than of a full-blown service). We request anonymous information about your
background and ask some evaluation questions. Responding to the questionnaire should
take about 30 minutes to complete. Thank you very much for your time and input.</p>
<h4><font SIZE="3"><img border="0" src="semweb.gif" align="right" hspace="2" alt="I
know what you mean..." width="140" height="129"></font>About HealthCyberMap</h4>
<p ALIGN="left">HealthCyberMap (<a HREF="http://healthcybermap.semanticweb.org/"
target="_blank">http://healthcybermap.semanticweb.org</a>) is a Semantic Web project
that aims at mapping parts of medical/health information resources in cyberspace in
novel semantic ways to improve their retrieval and navigation. The Semantic Web (<a
HREF="http://www.w3.org/2001/sw/" target="_blank">http://www.w3.org/2001/sw/</a> and
<a HREF="http://www.semanticweb.org/" target="_blank">http://www.semanticweb.org</a>)
aims to be the next-generation World Wide Web by giving machine-readable semantics and
context to the currently presentation-based Web pages so that, for example, a search
for resources on Mr. Wood does not return results on tree wood. <a
href="http://healthcybermap.semanticweb.org/#underhood" target="_blank">Follow this
link to learn more about HealthCyberMap</a>.</p>
```

```
<p><a href="http://www.hon.ch/HONcode/Conduct.html?HONConduct874664"
target="_blank"><img alt="We subscribe to the HONcode principles of the HON
Foundation. Click to verify." src="http://www.hon.ch/docs/HONcode/coc_velv_s.gif"
border="0" align="left" hspace="1" width="45" height="50"></a>We subscribe to the HON
code principles. <a href="http://www.hon.ch/HONcode/Conduct.html?HONConduct874664"
target="_blank">Verify here</a>.<br>
<br>
You may also wish to read our <a href="privacy.htm" target="_blank"> privacy
statement</a>.</p>
<hr color="#0071C0">
<%
Dim cnnFormToDB
Dim strSQL
Dim strgender
Dim strage
Dim streducation
Dim strrole
Dim strInternet_hrs_wk
Dim strcompuse
Dim strbrowseruse
Dim strownhomePC
Dim strWebaccessLoc
Dim strInternetConnection
Dim strbrowsertype
Dim strbrowserversion
Dim strscreenres
Dim strOS
Dim struseInternetforhealth
Dim strhcm_visit_frequency
Dim strhow_user_knew_hcm
Dim struse_if_fully_developed
Dim strmeets_my_ret_nav_needs
Dim strother_portals
Dim strother_means_or_google
Dim strworldmaps_well
Dim strBodyViewermaps_well
Dim strtextualcategories_well
Dim strtypemap_well
Dim strlanguageclassif_well
Dim strsemsearcheng_well
Dim strBodyViewer_icons_familiar
Dim strselecting_right_words_icons
Dim strworldmaps_important
Dim strBodyViewermaps_important
Dim strtextualcategories_important
Dim strtypemap_important
Dim strlanguageclassif_important
Dim strsemsearcheng_important
Dim strhelp_adequate
Dim strspeed
Dim strtask
Dim strtask_interface_used
Dim strrelevance
Dim strquality
Dim strpresentation_format
Dim strvisualmaps_vs_text
Dim strmap_net_compare
Dim strfuture_multiaxial
Dim strfuture_customisation
Dim strfuture_pkl
Dim strfuture_mappinghealthprob
Dim strorganise_by_audience
Dim strlike_found_helpful
Dim strdislike_found_difficult
Dim strcontinue_develop
Dim strsuggest_comments
Dim strquestionnaire_thoughts
If Request.Form("action") <> "Save Form Data" Then
' Show the form
%>
<h4>Questions (You may leave blank any question you don't want to answer)</h4>
<form action="<%= Request.ServerVariables("SCRIPT_NAME") %>" method="post">
<input type="hidden" name="action" value="Save Form Data" />
<p><b>Question 1.</b> Gender:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="gender" VALUE="Male"> Male<br>
```

```
<input TYPE="RADIO" NAME="gender" VALUE="Female"> Female<br>
<br>
</p>
</blockquote>
<p><b>Question 2.</b> Which of the following categories includes your age:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="age" VALUE="Under 18"> Under 18<br>
<input TYPE="RADIO" NAME="age" VALUE="18 - 25"> 18 - 25<br>
<input TYPE="RADIO" NAME="age" VALUE="26 - 35"> 26 - 35<br>
<input TYPE="RADIO" NAME="age" VALUE="36 - 45"> 36 - 45<br>
<input TYPE="RADIO" NAME="age" VALUE="46 - 65"> 46 - 65<br>
<input TYPE="RADIO" NAME="age" VALUE="Over 65"> Over 65<br>
<br>
</p>
</blockquote>
<p><b>Question 3.</b> What is the highest level of education you completed:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="education" VALUE="School"> School<br>
<input TYPE="RADIO" NAME="education" VALUE="University/College">
University/College<br>
<input TYPE="RADIO" NAME="education" VALUE="Postgraduate"> Postgraduate<br>
<input TYPE="RADIO" NAME="education" VALUE="Vocational training"> Vocational
training<br>
<br>
</p>
</blockquote>
<p><b>Question 4.</b> Which of the following best describes you:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="role" VALUE="Member of the general public seeking health-
related information"> Member of the general public seeking health-related
information<br>
<input TYPE="RADIO" NAME="role" VALUE="Member of the general public touched by disease
"> Member of the general public touched by disease<br>
<input TYPE="RADIO" NAME="role" VALUE="General practitioner"> General practitioner<br>
<input TYPE="RADIO" NAME="role" VALUE="Hospital doctor"> Hospital doctor<br>
<input TYPE="RADIO" NAME="role" VALUE="Nurse"> Nurse<br>
<input TYPE="RADIO" NAME="role" VALUE="PAM/AHP (Allied Health Professional)/other
healthcare professional">
PAM/AHP (Allied Health Professional)/other healthcare professional<br>
<input TYPE="RADIO" NAME="role" VALUE="Scientist/researcher"> Scientist/researcher<br>
<input TYPE="RADIO" NAME="role" VALUE="Medical librarian"> Medical librarian<br>
<input TYPE="RADIO" NAME="role" VALUE="Other"> Other<br>
<br>
</p>
</blockquote>
<p><b>Question 5.</b> How many hours per week, including work and home, do you
currently use the
Internet:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="Internet-hrs-wk" VALUE="Less than 5 hours"> Less than 5
hours<br>
<input TYPE="RADIO" NAME="Internet-hrs-wk" VALUE="More than 5 hours"> More than 5
hours<br>
<br>
</p>
</blockquote>
<p><b>Question 6.</b> Regularly used a computer:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="compuse" VALUE="Not at all"> Not at all<br>
<input TYPE="RADIO" NAME="compuse" VALUE="Few weeks"> Few weeks<br>
<input TYPE="RADIO" NAME="compuse" VALUE="2-6 months"> 2-6 months<br>
<input TYPE="RADIO" NAME="compuse" VALUE="6-24 months"> 6-24 months<br>
<input TYPE="RADIO" NAME="compuse" VALUE="&gt;2 years"> &gt;2 years<br>
<br>
</p>
</blockquote>
<p><b>Question 7.</b> Regularly used the Internet using a browser (e.g., Netscape,
Internet Explorer):</p>
<blockquote>
<p><input TYPE="RADIO" NAME="browseruse" VALUE="Not at all"> Not at all<br>
<input TYPE="RADIO" NAME="browseruse" VALUE="Few weeks"> Few weeks<br>
<input TYPE="RADIO" NAME="browseruse" VALUE="2-6 months"> 2-6 months<br>
<input TYPE="RADIO" NAME="browseruse" VALUE="6-24 months"> 6-24 months<br>
<input TYPE="RADIO" NAME="browseruse" VALUE="&gt;2 years"> &gt;2 years<br>
<br>
</p>
```

```
</blockquote>
<p><b>Question 8.</b> Do you have a working personal computer (PC) in your home:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="ownhomePC" VALUE="Yes"> Yes<br>
<input TYPE="RADIO" NAME="ownhomePC" VALUE="No"> No<br>
<br>
</p>
</blockquote>
<p><b>Question 9.</b> From which location are you most likely to use the Web:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="WebaccessLoc" VALUE="Home"> Home<br>
<input TYPE="RADIO" NAME="WebaccessLoc" VALUE="Work"> Work<br>
<input TYPE="RADIO" NAME="WebaccessLoc" VALUE="School"> School<br>
<input TYPE="RADIO" NAME="WebaccessLoc" VALUE="Library"> Library<br>
<input TYPE="RADIO" NAME="WebaccessLoc" VALUE="Other"> Other<br>
<br>
</p>
</blockquote>
<p><b>Question 10.</b> What type of Internet connection do you have currently (if you
use more than one type of connection in different places, select the one you use
most):</p>
<blockquote>
<p><input TYPE="RADIO" NAME="InternetConnection" VALUE="Modem"> Modem<br>
<input TYPE="RADIO" NAME="InternetConnection" VALUE="ISDN/ADSL/Broadband/Always
connected"> ISDN/ADSL/Broadband/Always connected<br>
<br>
</p>
</blockquote>
<p><b>Question 11.</b> What web browser do you use most frequently:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="browsertype" VALUE="Microsoft Internet Explorer">
Microsoft Internet Explorer<br>
<input TYPE="RADIO" NAME="browsertype" VALUE="Netscape Navigator or Communicator">
Netscape Navigator or Communicator<br>
<input TYPE="RADIO" NAME="browsertype" VALUE="Other"> Other<br>
<br>
</p>
</blockquote>
<p><b>Question 12.</b> Which version of the browser do you use most frequently:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="browserversion" VALUE="Version 3.x"> Version 3.x<br>
<input TYPE="RADIO" NAME="browserversion" VALUE="Version 4.x"> Version 4.x<br>
<input TYPE="RADIO" NAME="browserversion" VALUE="Version 5.x"> Version 5.x<br>
<input TYPE="RADIO" NAME="browserversion" VALUE="Version 6.x"> Version 6.x<br>
<input TYPE="RADIO" NAME="browserversion" VALUE="Other"> Other<br>
<br>
</p>
</blockquote>
<p><b>Question 13.</b> The Desktop Area/Screen Resolution you <I>most commonly</I> use
is:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="screenres" VALUE="640 by 480 pixels"> 640 by 480
pixels<br>
<input TYPE="RADIO" NAME="screenres" VALUE="800 by 600 pixels"> 800 by 600 pixels<br>
<input TYPE="RADIO" NAME="screenres" VALUE="1024 by 768 pixels or more"> 1024 by 768
pixels or more<br>
<br>
</p>
<script language="JavaScript">
var width = screen.width
var height = screen.height
document.write("N.B. We have detected that your current screen resolution is set to
"+width+ "x" +height+".");
</script>
</blockquote>
<p><b>Question 14.</b> Which operating system are you currently running on your
primary (most frequently used) PC:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="OS" VALUE="Windows"> Windows<br>
<input TYPE="RADIO" NAME="OS" VALUE="Macintosh"> Macintosh<br>
<input TYPE="RADIO" NAME="OS" VALUE="Unix/Linux"> Unix/Linux<br>
<input TYPE="RADIO" NAME="OS" VALUE="WebTV"> WebTV<br>
<input TYPE="RADIO" NAME="OS" VALUE="Other"> Other<br>
<br>
</p>
</blockquote>
```

```
<p><b>Question 15.</b> Do you consider the Internet an important source of reliable
medical/health-related information:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="useInternetforhealth" VALUE="Not at all"> Not at all<br>
<input TYPE="RADIO" NAME="useInternetforhealth" VALUE="To some extent"> To some
extent<br>
<input TYPE="RADIO" NAME="useInternetforhealth" VALUE="Definitely"> Definitely<br>
<br>
</p>
</blockquote>
<p><b>Question 16.</b> How often have you accessed HealthCyberMap Web site over the
past 3 months:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="hcm-visit-frequency" VALUE="Daily"> Daily<br>
<input TYPE="RADIO" NAME="hcm-visit-frequency" VALUE="Weekly"> Weekly<br>
<input TYPE="RADIO" NAME="hcm-visit-frequency" VALUE="Monthly"> Monthly<br>
<input TYPE="RADIO" NAME="hcm-visit-frequency" VALUE="This is the first time I have
accessed the Web site"> This is the first time I have accessed the Web site<br>
<br>
</p>
</blockquote>
<p><b>Question 17.</b> How did you hear about HealthCyberMap (Choose the one that best
applies to the first way you learned about HealthCyberMap):</p>
<blockquote>
<p><input TYPE="RADIO" NAME="how-user-knew-hcm" VALUE="I listened to a
presentation/read a paper on the project"> I listened to a presentation/read a paper
on the project<br>
<input TYPE="RADIO" NAME="how-user-knew-hcm" VALUE="A colleague told me about it"> A
colleague told me about it<br>
<input TYPE="RADIO" NAME="how-user-knew-hcm" VALUE="I received a notice through e-
mail/e-mail list"> I received a notice through e-mail/e-mail list<br>
<input TYPE="RADIO" NAME="how-user-knew-hcm" VALUE="I came across HealthCyberMap on
the Internet (search engine, link from other site)"> I came across HealthCyberMap on
the Internet (search engine, link from other site)<br>
<input TYPE="RADIO" NAME="how-user-knew-hcm" VALUE="I am a personal friend or relative
of researcher involved in this project"> I am a personal friend or relative of
researcher involved in this project<br>
<input TYPE="RADIO" NAME="how-user-knew-hcm" VALUE="Other"> Other<br>
<br>
</p>
</blockquote>
<p><b>Question 18.</b> As a pilot system, HealthCyberMap does not include an
exhaustive coverage of all medical/health conditions. Based on what you have seen so
far, if HealthCyberMap was fully developed to provide a more complete coverage, would
you use it as an information portal and one-stop access point to a variety of medical
and health-related resources:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="use-if-fully-developed" VALUE="Yes"> Yes<br>
<input TYPE="RADIO" NAME="use-if-fully-developed" VALUE="No"> No<br>
<br>
</p>
</blockquote>
<p><b>Question 19.</b> HealthCyberMap concepts and interfaces meet my information
retrieval and navigation needs (not topical coverage) better than other medical/health
information portals/gateways:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="meets-my-ret-nav-needs" VALUE="Strongly disagree">
Strongly disagree<br>
<input TYPE="RADIO" NAME="meets-my-ret-nav-needs" VALUE="Disagree"> Disagree<br>
<input TYPE="RADIO" NAME="meets-my-ret-nav-needs" VALUE="Neither disagree nor agree">
Neither disagree nor agree<br>
<input TYPE="RADIO" NAME="meets-my-ret-nav-needs" VALUE="Agree"> Agree<br>
<input TYPE="RADIO" NAME="meets-my-ret-nav-needs" VALUE="Strongly agree"> Strongly
agree<br>
<br>
</p>
</blockquote>
<p><b>Question 20.</b> (Only if you did not agree with the above statement) At this
time, what other medical/health portals/gateway are better than HealthCyberMap
regarding information retrieval and navigation (not topical coverage)</p>
<blockquote>
<p><textarea NAME="other-portals" ROWS="5" COLS="35"></textarea><br>
</p>
</blockquote>
```

```
<p><b>Question 21.</b> I prefer to use a conventional search engine (e.g., <a
href="http://www.google.com" target="_blank"> Google</a>) to gather information or to
have a librarian, staff member, or family member gather information for me:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="other-means-or-google" VALUE="True"> True<br>
<input TYPE="RADIO" NAME="other-means-or-google" VALUE="False"> False<br>
<br>
</p>
</blockquote>
<p><b>Question 22.</b> How well does each of the six information interfaces featured
in the current HealthCyberMap pilot perform its intended purpose? (Please rate each of
the information interfaces from &quot;not at all well&quot; to &quot;extremely
well&quot; considering aspects like ease of access, use and navigation, and logical
arrangement of information on the pages of concerned interfaces)</p>
<blockquote>
<p><b>Question 22.1.</b> <a href="world_map/" target="_blank">HealthCyberMap world
maps of Web resources classified according to geographic provenance</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="worldmaps-well" VALUE="Not at all well"> Not at all
well<br>
<input TYPE="RADIO" NAME="worldmaps-well" VALUE="Slightly well"> Slightly well<br>
<input TYPE="RADIO" NAME="worldmaps-well" VALUE="Somewhat well"> Somewhat well<br>
<input TYPE="RADIO" NAME="worldmaps-well" VALUE="Very well"> Very well<br>
<input TYPE="RADIO" NAME="worldmaps-well" VALUE="Extremely well"> Extremely well<br>
<br>
</p>
</blockquote>
<p><b>Question 22.2.</b> <a href="bodyviewer/" target="_blank">HealthCyberMap
BodyViewer maps of Web resources classified according to medical/health
topics</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="BodyViewermaps-well" VALUE="Not at all well"> Not at all
well<br>
<input TYPE="RADIO" NAME="BodyViewermaps-well" VALUE="Slightly well"> Slightly
well<br>
<input TYPE="RADIO" NAME="BodyViewermaps-well" VALUE="Somewhat well"> Somewhat
well<br>
<input TYPE="RADIO" NAME="BodyViewermaps-well" VALUE="Very well"> Very well<br>
<input TYPE="RADIO" NAME="BodyViewermaps-well" VALUE="Extremely well"> Extremely
well<br>
<br>
</p>
</blockquote>
<p><b>Question 22.3.</b> <a
href="javascript:Start('http://healthcybermap.semanticweb.org/bodyviewer/icd9cm.htm')"
>HealthCyberMap textual resource index using ICD-9-CM top-level categories of
medical/health topics</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="textualcategories-well" VALUE="Not at all well"> Not at
all well<br>
<input TYPE="RADIO" NAME="textualcategories-well" VALUE="Slightly well"> Slightly
well<br>
<input TYPE="RADIO" NAME="textualcategories-well" VALUE="Somewhat well"> Somewhat
well<br>
<input TYPE="RADIO" NAME="textualcategories-well" VALUE="Very well"> Very well<br>
<input TYPE="RADIO" NAME="textualcategories-well" VALUE="Extremely well"> Extremely
well<br>
<br>
</p>
</blockquote>
<p><b>Question 22.4.</b> <a href="type.htm" target="_blank">HealthCyberMap map of Web
resources classified according to resource type, e.g., audio-visual material,
interactive resources, digital atlases, fact sheets, guidelines, etc.</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="typemap-well" VALUE="Not at all well"> Not at all
well<br>
<input TYPE="RADIO" NAME="typemap-well" VALUE="Slightly well"> Slightly well<br>
<input TYPE="RADIO" NAME="typemap-well" VALUE="Somewhat well"> Somewhat well<br>
<input TYPE="RADIO" NAME="typemap-well" VALUE="Very well"> Very well<br>
<input TYPE="RADIO" NAME="typemap-well" VALUE="Extremely well"> Extremely well<br>
<br>
</p>
</blockquote>
<p><b>Question 22.5.</b> <a href="language.htm" target="_blank">HealthCyberMap Web
resources classified according to resource language</a>:</p>
<blockquote>
```

```
<p><input TYPE="RADIO" NAME="languageclassif-well" VALUE="Not at all well"> Not at all
well<br>
<input TYPE="RADIO" NAME="languageclassif-well" VALUE="Slightly well"> Slightly
well<br>
<input TYPE="RADIO" NAME="languageclassif-well" VALUE="Somewhat well"> Somewhat
well<br>
<input TYPE="RADIO" NAME="languageclassif-well" VALUE="Very well"> Very well<br>
<input TYPE="RADIO" NAME="languageclassif-well" VALUE="Extremely well"> Extremely
well<br>
<br>
</p>
</blockquote>
<p><b>Question 22.6.</b><a
href="javascript:Start('http://healthcybermap.semanticweb.org/icd.asp')">HealthCyberMa
p semantic subject search engine</a> (<a
href="http://healthcybermap.semanticweb.org/icd.htm" target="_blank">about this
engine</a>):</p>
<blockquote>
<p><input TYPE="RADIO" NAME="semsearcheng-well" VALUE="Not at all well"> Not at all
well<br>
<input TYPE="RADIO" NAME="semsearcheng-well" VALUE="Slightly well"> Slightly well<br>
<input TYPE="RADIO" NAME="semsearcheng-well" VALUE="Somewhat well"> Somewhat well<br>
<input TYPE="RADIO" NAME="semsearcheng-well" VALUE="Very well"> Very well<br>
<input TYPE="RADIO" NAME="semsearcheng-well" VALUE="Extremely well"> Extremely
well<br>
<br>
</p>
</blockquote>
</blockquote>
<p><b>Question 23.</b> Did you find the meanings of <a href="bodyviewer/"
target="_blank"> the human body icons on HealthCyberMap BodyViewer maps</a> familiar
to you:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="BodyViewer-icons-familiar" VALUE="Yes"> Yes<br>
<input TYPE="RADIO" NAME="BodyViewer-icons-familiar" VALUE="Somewhat (not always)">
Somewhat (not always)<br>
<input TYPE="RADIO" NAME="BodyViewer-icons-familiar" VALUE="No"> No<br>
<br>
</p>
</blockquote>
<p><b>Question 24.</b> Selecting the right words for search, and finding the right
areas/icons to click on the maps was:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="selecting-right-words-icons" VALUE="Very difficult"> Very
difficult<br>
<input TYPE="RADIO" NAME="selecting-right-words-icons" VALUE="Difficult">
Difficult<br>
<input TYPE="RADIO" NAME="selecting-right-words-icons" VALUE="Of moderate (acceptable)
difficulty"> Of moderate (acceptable) difficulty<br>
<input TYPE="RADIO" NAME="selecting-right-words-icons" VALUE="Easy"> Easy<br>
<input TYPE="RADIO" NAME="selecting-right-words-icons" VALUE="Very easy"> Very
easy<br>
<br>
</p>
</blockquote>
<p><b>Question 25.</b> How important/useful to you is each of the six information
interfaces featured in the current HealthCyberMap pilot (Please rate each of the
information interfaces from &quot;not at all important&quot; to &quot;extremely
important&quot;)</p>
<blockquote>
<p><b>Question 25.1.</b> <a href="world_map/" target="_blank">HealthCyberMap world
maps of Web resources classified according to geographic provenance</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="worldmaps-important" VALUE="Not at all important"> Not at
all important<br>
<input TYPE="RADIO" NAME="worldmaps-important" VALUE="Slightly important"> Slightly
important<br>
<input TYPE="RADIO" NAME="worldmaps-important" VALUE="Somewhat important"> Somewhat
important<br>
<input TYPE="RADIO" NAME="worldmaps-important" VALUE="Very important"> Very
important<br>
<input TYPE="RADIO" NAME="worldmaps-important" VALUE="Extremely important"> Extremely
important<br>
<br>
</p>
</blockquote>
```

```
<p><b>Question 25.2.</b> <a href="bodyviewer/" target="_blank">HealthCyberMap
BodyViewer maps of Web resources classified according to medical/health
topics</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="BodyViewermaps-important" VALUE="Not at all important">
Not at all important<br>
<input TYPE="RADIO" NAME="BodyViewermaps-important" VALUE="Slightly important">
Slightly important<br>
<input TYPE="RADIO" NAME="BodyViewermaps-important" VALUE="Somewhat important">
Somewhat important<br>
<input TYPE="RADIO" NAME="BodyViewermaps-important" VALUE="Very important"> Very
important<br>
<input TYPE="RADIO" NAME="BodyViewermaps-important" VALUE="Extremely important">
Extremely important<br>
<br>
</p>
</blockquote>
<p><b>Question 25.3.</b> <a
href="javascript:Start('http://healthcybermap.semanticweb.org/bodyviewer/icd9cm.htm')"
>HealthCyberMap textual resource index using ICD-9-CM top-level categories of
medical/health topics</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="textualcategories-important" VALUE="Not at all
important"> Not at all important<br>
<input TYPE="RADIO" NAME="textualcategories-important" VALUE="Slightly important">
Slightly important<br>
<input TYPE="RADIO" NAME="textualcategories-important" VALUE="Somewhat important">
Somewhat important<br>
<input TYPE="RADIO" NAME="textualcategories-important" VALUE="Very important"> Very
important<br>
<input TYPE="RADIO" NAME="textualcategories-important" VALUE="Extremely important">
Extremely important<br>
<br>
</p>
</blockquote>
<p><b>Question 25.4.</b> <a href="type.htm" target="_blank">HealthCyberMap map of Web
resources classified according to resource type, e.g., audio-visual material,
interactive resources, digital atlases, fact sheets, guidelines, etc.</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="typemap-important" VALUE="Not at all important"> Not at
all important<br>
<input TYPE="RADIO" NAME="typemap-important" VALUE="Slightly important"> Slightly
important<br>
<input TYPE="RADIO" NAME="typemap-important" VALUE="Somewhat important"> Somewhat
important<br>
<input TYPE="RADIO" NAME="typemap-important" VALUE="Very important"> Very
important<br>
<input TYPE="RADIO" NAME="typemap-important" VALUE="Extremely important"> Extremely
important<br>
<br>
</p>
</blockquote>
<p><b>Question 25.5.</b> <a href="language.htm" target="_blank">HealthCyberMap Web
resources classified according to resource language</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="languageclassif-important" VALUE="Not at all important">
Not at all important<br>
<input TYPE="RADIO" NAME="languageclassif-important" VALUE="Slightly important">
Slightly important<br>
<input TYPE="RADIO" NAME="languageclassif-important" VALUE="Somewhat important">
Somewhat important<br>
<input TYPE="RADIO" NAME="languageclassif-important" VALUE="Very important"> Very
important<br>
<input TYPE="RADIO" NAME="languageclassif-important" VALUE="Extremely important">
Extremely important<br>
<br>
</p>
</blockquote>
<p><b>Question 25.6.</b><a
href="javascript:Start('http://healthcybermap.semanticweb.org/icd.asp')">HealthCyberMa
p semantic subject search engine</a> (<a
href="http://healthcybermap.semanticweb.org/icd.htm" target="_blank">about this
engine</a>):</p>
<blockquote>
<p><input TYPE="RADIO" NAME="semsearcheng-important" VALUE="Not at all important"> Not
at all important<br>
```

```
<input TYPE="RADIO" NAME="semsearcheng-important" VALUE="Slightly important"> Slightly
important<br>
<input TYPE="RADIO" NAME="semsearcheng-important" VALUE="Somewhat important"> Somewhat
important<br>
<input TYPE="RADIO" NAME="semsearcheng-important" VALUE="Very important"> Very
important<br>
<input TYPE="RADIO" NAME="semsearcheng-important" VALUE="Extremely important">
Extremely important<br>
<br>
</p>
</blockquote>
</blockquote>
<p><b>Question 26.</b> Do you find <a
href="javascript:Start('http://healthcybermap.semanticweb.org/world_map/help.htm')">
HealthCyberMap map interface help</a> (and other instructions/help
hints around the site) adequate:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="help-adequate" VALUE="Yes, HealthCyberMap is easy to use
once you understand how it works; the online help and instructions provided were
useful"> Yes, HealthCyberMap is easy to use once you understand how it works; the
online help and instructions provided were useful<br>
<input TYPE="RADIO" NAME="help-adequate" VALUE="HealthCyberMap is very easy to use; I
didnt need any help (or made very little use of it)"> HealthCyberMap is very easy to
use; I didn’t need any help (or made very little use of it)<br>
<input TYPE="RADIO" NAME="help-adequate" VALUE="No, I need more help to locate the
information I need on HealthCyberMap Web site"> No, I need more help to locate the
information I need on HealthCyberMap Web site<br>
<br>
</p>
</blockquote>
<p><b>Question 27.</b> The speed at which HealthCyberMap maps load on my Internet
connection is:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="speed" VALUE="Reasonably fast"> Reasonably fast<br>
<input TYPE="RADIO" NAME="speed" VALUE="Slow"> Slow<br>
<br>
</p>
</blockquote>
<p><b>Question 28.</b> Now, your task is to find information resources about specific
diseases using HealthCyberMap. Choose at least one disease from the following list:
<i>prostate cancer</i>, <i> Crohn's disease</i>, <i> Guillain-Barre syndrome</i>,<i>
amyotrophic lateral sclerosis</i>, <i> diabetes mellitus</i>, <i> Mitral valve
prolapse</i>,<i> ectopic pregnancy</i> (or any other medical/health topic(s) of your
choice). On the average (if you tried finding resources on more than one disease), how
successful were you in completing this task:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="task" VALUE="Successful from the first attempt">
Successful from the first attempt<br>
<input TYPE="RADIO" NAME="task" VALUE="Successful after one or more failed attempts">
Successful after one or more failed attempts<br>
<input TYPE="RADIO" NAME="task" VALUE="Not successful"> Not successful<br>
<br>
</p>
</blockquote>
<p><b>Question 29.</b> What interface did you use most of the time to
accomplish the above task:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="task-interface-used" VALUE="BodyViewer maps of Web
resources classified according to medical/health topics"> <a href="bodyviewer/"
target="_blank">BodyViewer maps of Web resources classified according to
medical/health topics</a><br>
<input TYPE="RADIO" NAME="task-interface-used" VALUE="HealthCyberMap textual resource
index using ICD-9-CM top-level categories of medical/health topics"> <a
href="javascript:Start('http://healthcybermap.semanticweb.org/bodyviewer/icd9cm.htm')"
>HealthCyberMap textual resource index using ICD-9-CM top-level categories of
medical/health topics</a><br>
<input TYPE="RADIO" NAME="task-interface-used" VALUE="HealthCyberMap semantic subject
search engine (advanced keyword search)"> <a
href="javascript:Start('http://healthcybermap.semanticweb.org/icd.asp')">HealthCyberMa
p semantic subject search engine (advanced keyword search)</a><br>
<input TYPE="RADIO" NAME="task-interface-used" VALUE="All three interfaces were
equally useful"> All three interfaces were equally useful<br>
<br>
</p>
</blockquote>
```

```
<p><b>Question 30.</b> The pointers to information resources returned by
HealthCyberMap were <i> relevant</i> to my queries (i.e., to search words I entered or
map areas I have clicked):</p>
<blockquote>
<p><input TYPE="RADIO" NAME="relevance" VALUE="Strongly agree"> Strongly agree<br>
<input TYPE="RADIO" NAME="relevance" VALUE="Agree"> Agree<br>
<input TYPE="RADIO" NAME="relevance" VALUE="Neither agree nor disagree"> Neither agree
nor disagree<br>
<input TYPE="RADIO" NAME="relevance" VALUE="Disagree"> Disagree<br>
<input TYPE="RADIO" NAME="relevance" VALUE="Strongly disagree"> Strongly disagree<br>
<br>
</p>
</blockquote>
<p><b>Question 31.</b> HealthCyberMap pointers to information resources are of good
<i>quality</i>, accurate, up-to-date and useful:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="quality" VALUE="Strongly agree"> Strongly agree<br>
<input TYPE="RADIO" NAME="quality" VALUE="Agree"> Agree<br>
<input TYPE="RADIO" NAME="quality" VALUE="Neither agree nor disagree"> Neither agree
nor disagree<br>
<input TYPE="RADIO" NAME="quality" VALUE="Disagree"> Disagree<br>
<input TYPE="RADIO" NAME="quality" VALUE="Strongly disagree"> Strongly disagree<br>
<br>
</p>
</blockquote>
<p><b>Question 32.</b> When I look for information in HealthCyberMap I get overloaded
quickly with too much detail (presentation format):</p>
<blockquote>
<p><input TYPE="RADIO" NAME="presentation-format" VALUE="Not at all a problem"> Not at
all a problem<br>
<input TYPE="RADIO" NAME="presentation-format" VALUE="A slight problem"> A slight
problem<br>
<input TYPE="RADIO" NAME="presentation-format" VALUE="A moderate problem"> A moderate
problem<br>
<input TYPE="RADIO" NAME="presentation-format" VALUE="A significant problem"> A
significant problem<br>
<input TYPE="RADIO" NAME="presentation-format" VALUE="An extreme problem"> An extreme
problem<br>
<br>
</p>
</blockquote>
<p><b>Question 33.</b> Do you think <a href="bodyviewer/" target="_blank"> visual
maps</a> are a useful addition to/complementary improvement over <a
href="javascript:Start('http://healthcybermap.semanticweb.org/bodyviewer/icd9cm.htm')"
>conventional text-based Web portal interfaces</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="visualmaps-vs-text" VALUE="Yes, definitely"> Yes,
definitely<br>
<input TYPE="RADIO" NAME="visualmaps-vs-text" VALUE="Could be an improvement (but not
always)"> Could be an improvement (but not always)<br>
<input TYPE="RADIO" NAME="visualmaps-vs-text" VALUE="No, visual interfaces are
useless"> No, visual interfaces are useless<br>
<br>
</p>
</blockquote>
<p><b>Question 34.</b> Compared to the mapping approach of <a
href="http://pubmed.antarcti.ca/start" target="_blank">map.net</a>, do you think
(e.g., regarding the use of familiar metaphors and presentation format of resource
details and links):</p>
<blockquote>
<p><input TYPE="RADIO" NAME="map.net-compare" VALUE="HealthCyberMap approach is
superior"> HealthCyberMap approach is superior<br>
<input TYPE="RADIO" NAME="map.net-compare" VALUE="HealthCyberMap approach is somewhat
better"> HealthCyberMap approach is somewhat better<br>
<input TYPE="RADIO" NAME="map.net-compare" VALUE="HealthCyberMap approach is about the
same"> HealthCyberMap approach is about the same<br>
<input TYPE="RADIO" NAME="map.net-compare" VALUE="HealthCyberMap approach is somewhat
worse"> HealthCyberMap approach is somewhat worse<br>
<input TYPE="RADIO" NAME="map.net-compare" VALUE="HealthCyberMap approach is
inferior"> HealthCyberMap approach is inferior<br>
<input TYPE="RADIO" NAME="map.net-compare" VALUE="Do not have an opinion"> Do not have
an opinion<br>
<br>
</p>
</blockquote>
```

```
<p><b>Question 35.</b> How important/ useful do you think are the proposed
HealthCyberMap future interfaces/possibilities/concepts</p>
<blockquote>
<p><b>Question 35.1.</b> <a href="multiaxial.htm" target="_blank">Multi-axial
classification of resources based on two or more Dublin Core metadata
elements</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="future-multiaxial" VALUE="Not at all important"> Not at
all important<br>
<input TYPE="RADIO" NAME="future-multiaxial" VALUE="Slightly important"> Slightly
important<br>
<input TYPE="RADIO" NAME="future-multiaxial" VALUE="Somewhat important"> Somewhat
important<br>
<input TYPE="RADIO" NAME="future-multiaxial" VALUE="Very important"> Very
important<br>
<input TYPE="RADIO" NAME="future-multiaxial" VALUE="Extremely important"> Extremely
important<br>
<br>
</p>
</blockquote>
<p><b>Question 35.2.</b> <a href="ip.htm" target="_blank">HealthCyberMap
customisation/detecting user location for location-based customisation</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="future-customisation" VALUE="Not at all important"> Not
at all important<br>
<input TYPE="RADIO" NAME="future-customisation" VALUE="Slightly important"> Slightly
important<br>
<input TYPE="RADIO" NAME="future-customisation" VALUE="Somewhat important"> Somewhat
important<br>
<input TYPE="RADIO" NAME="future-customisation" VALUE="Very important"> Very
important<br>
<input TYPE="RADIO" NAME="future-customisation" VALUE="Extremely important"> Extremely
important<br>
<br>
</p>
</blockquote>
<p><b>Question 35.3.</b> <a href="pk.htm" target="_blank">Problem to knowledge linking
(clinical codes as knowledge hooks)</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="future-pkl" VALUE="Not at all important"> Not at all
important<br>
<input TYPE="RADIO" NAME="future-pkl" VALUE="Slightly important"> Slightly
important<br>
<input TYPE="RADIO" NAME="future-pkl" VALUE="Somewhat important"> Somewhat
important<br>
<input TYPE="RADIO" NAME="future-pkl" VALUE="Very important"> Very important<br>
<input TYPE="RADIO" NAME="future-pkl" VALUE="Extremely important"> Extremely
important<br>
<br>
</p>
</blockquote>
<p><b>Question 35.4.</b> <a href="infogaps.htm" target="_blank">Mapping health
problems in HealthCyberMap and identifying information needs and gaps</a>:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="future-mappinghealthprob" VALUE="Not at all important">
Not at all important<br>
<input TYPE="RADIO" NAME="future-mappinghealthprob" VALUE="Slightly important">
Slightly important<br>
<input TYPE="RADIO" NAME="future-mappinghealthprob" VALUE="Somewhat important">
Somewhat important<br>
<input TYPE="RADIO" NAME="future-mappinghealthprob" VALUE="Very important"> Very
important<br>
<input TYPE="RADIO" NAME="future-mappinghealthprob" VALUE="Extremely important">
Extremely important<br>
<br>
</p>
</blockquote>
</blockquote>
<p><b>Question 36.</b> Do you think it would be useful in a future implementation of
HealthCyberMap to also organise information resources by user needs/intended primary
audience as patients, health professionals, or basic researchers:</p>
<blockquote>
<p><input TYPE="RADIO" NAME="organise-by-audience" VALUE="Yes"> Yes<br>
<input TYPE="RADIO" NAME="organise-by-audience" VALUE="No"> No<br>
<br>
</p>
```

```
</blockquote>
<p><b>Question 37.</b> Were there any parts of the service that you found especially
helpful? What do you like most about HealthCyberMap and why?</p>
<blockquote>
<p><textarea NAME="like-found-helpful" ROWS="5" COLS="35"></textarea><br>
</p>
</blockquote>
<p><b>Question 38.</b> Were there any parts of the service that you found especially
difficult to use or understand? What do you dislike most about HealthCyberMap and
why?</p>
<blockquote>
<p><textarea NAME="dislike-found-difficult" ROWS="5" COLS="35"></textarea><br>
</p>
</blockquote>
<p><b>Question 39.</b> Should resources (if available) be invested to continue
developing and implementing HealthCyberMap? Why or why not?</p>
<blockquote>
<p><textarea NAME="continue-develop" ROWS="5" COLS="35"></textarea><br>
</p>
</blockquote>
<p><b>Question 40.</b> What are your suggestions or comments about what would make the
service better? (e.g., I would like the following added to, changed in, or deleted
from HealthCyberMap)</p>
<blockquote>
<p><textarea NAME="suggest-comments" ROWS="5" COLS="35"></textarea><br>
</p>
</blockquote>
<p><b>Question 41.</b> Finally we would like to hear your thoughts about this
questionnaire. For example, are we asking the right questions? Are we asking the
questions in the right way? Your feedback will help us design better questionnaires in
the future:</p>
<blockquote>
<p><textarea NAME="questionnaire-thoughts" ROWS="5" COLS="35"></textarea><br>
</p>
</blockquote>
<b>Please submit your responses by clicking the &quot;Save&quot; button below. Again,
thank you very much for your time and input.</b>
<p><input TYPE="SUBMIT" VALUE="Save"> <input TYPE="RESET" VALUE="Clear"></p>
</form>
<%
Else
strgender = Request.Form("gender")
strage = Request.Form("age")
streducation = Request.Form("education")
strrole = Request.Form("role")
strInternet_hrs_wk = Request.Form("Internet-hrs-wk")
strcompuse = Request.Form("compuse")
strbrowseruse = Request.Form("browseruse")
strownhomePC = Request.Form("ownhomePC")
strWebaccessLoc = Request.Form("WebaccessLoc")
strInternetConnection = Request.Form("InternetConnection")
strbrowsertype = Request.Form("browsertype")
strbrowserversion = Request.Form("browserversion")
strscreenres = Request.Form("screenres")
strOS = Request.Form("OS")
struseInternetforhealth = Request.Form("useInternetforhealth")
strhcm_visit_frequency = Request.Form("hcm-visit-frequency")
strhow_user_knew_hcm = Request.Form("how-user-knew-hcm")
struse_if_fully_developed = Request.Form("use-if-fully-developed")
strmeets_my_ret_nav_needs = Request.Form("meets-my-ret-nav-needs")
strother_portals = Request.Form("other-portals")
strother_portals = Replace(strother_portals, "'", "''")
strother_means_or_google = Request.Form("other-means-or-google")
strworldmaps_well = Request.Form("worldmaps-well")
strBodyViewermaps_well = Request.Form("BodyViewermaps-well")
strtextualcategories_well = Request.Form("textualcategories-well")
strtypemap_well = Request.Form("typemap-well")
strlanguageclassif_well = Request.Form("languageclassif-well")
strsemsearcheng_well = Request.Form("semsearcheng-well")
strBodyViewer_icons_familiar = Request.Form("BodyViewer-icons-familiar")
strselecting_right_words_icons = Request.Form("selecting-right-words-icons")
strworldmaps_important = Request.Form("worldmaps-important")
strBodyViewermaps_important = Request.Form("BodyViewermaps-important")
strtextualcategories_important = Request.Form("textualcategories-important")
strtypemap_important = Request.Form("typemap-important")
strlanguageclassif_important = Request.Form("languageclassif-important")
```

```
strsemsearcheng_important = Request.Form("semsearcheng-important")
strhelp_adequate = Request.Form("help-adequate")
strspeed = Request.Form("speed")
strtask = Request.Form("task")
strtask_interface_used = Request.Form("task-interface-used")
strrelevance = Request.Form("relevance")
strquality = Request.Form("quality")
strpresentation_format = Request.Form("presentation-format")
strvisualmaps_vs_text = Request.Form("visualmaps-vs-text")
strmap_net_compare = Request.Form("map.net-compare")
strfuture_multiaxial = Request.Form("future-multiaxial")
strfuture_customisation = Request.Form("future-customisation")
strfuture_pkl = Request.Form("future-pkl")
strfuture_mappinghealthprob = Request.Form("future-mappinghealthprob")
strorganise_by_audience = Request.Form("organise-by-audience")
strlike_found_helpful = Request.Form("like-found-helpful")
strlike_found_helpful = Replace(strlike_found_helpful, "'", "''")
strdislike_found_difficult = Request.Form("dislike-found-difficult")
strdislike_found_difficult = Replace(strdislike_found_difficult, "'", "''")
strcontinue_develop = Request.Form("continue-develop")
strcontinue_develop = Replace(strcontinue_develop, "'", "''")
strsuggest_comments = Request.Form("suggest-comments")
strsuggest_comments = Replace(strsuggest_comments, "'", "''")
strquestionnaire_thoughts = Request.Form("questionnaire-thoughts")
strquestionnaire_thoughts = Replace(strquestionnaire_thoughts, "'", "''")
Set cnnFormToDB = Server.CreateObject("ADODB.Connection")
cnnFormToDB.Open "questionnaire","",""
strSQL = ""
strSQL = strSQL & "INSERT INTO Results "
strSQL = strSQL & "(gender, age, education, role, Internet_hrs_wk, compuse,
browseruse, ownhomePC, WebaccessLoc, InternetConnection, browsertype, browserversion,
screenres, OS, useInternetforhealth, hcm_visit_frequency, how_user_knew_hcm,
use_if_fully_developed, meets_my_ret_nav_needs, other_portals, other_means_or_google,
worldmaps_well, BodyViewermaps_well, textualcategories_well, typemap_well,
languageclassif_well, semsearcheng_well, BodyViewer_icons_familiar,
selecting_right_words_icons, worldmaps_important, BodyViewermaps_important,
textualcategories_important, typemap_important, languageclassif_important,
semsearcheng_important, help_adequate, speed, task, task_interface_used, relevance,
quality, presentation_format, visualmaps_vs_text, map_net_compare, future_multiaxial,
future_customisation, future_pkl, future_mappinghealthprob, organise_by_audience,
like_found_helpful, dislike_found_difficult, continue_develop, suggest_comments,
questionnaire_thoughts) " & vbCrLf
strSQL = strSQL & "VALUES ("
strSQL = strSQL & "'" & strgender & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strage & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & streducation & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strrole & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strInternet_hrs_wk & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strcompuse & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strbrowseruse & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strownhomePC & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strWebaccessLoc & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strInternetConnection & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strbrowsertype & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strbrowserversion & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strscreenres & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strOS & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & struseInternetforhealth & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strhcm_visit_frequency & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strhow_user_knew_hcm & "'"
```

```
strSQL = strSQL & ", "
strSQL = strSQL & "'" & struse_if_fully_developed & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strmeets_my_ret_nav_needs & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strother_portals & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strother_means_or_google & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strworldmaps_well & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strBodyViewermaps_well & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strtextualcategories_well & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strtypemap_well & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strlanguageclassif_well & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strsemsearcheng_well & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strBodyViewer_icons_familiar & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strselecting_right_words_icons & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strworldmaps_important & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strBodyViewermaps_important & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strtextualcategories_important & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strtypemap_important & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strlanguageclassif_important & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strsemsearcheng_important & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strhelp_adequate & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strspeed & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strtask & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strtask_interface_used & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strrelevance & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strquality & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strpresentation_format & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strvisualmaps_vs_text & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strmap_net_compare & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strfuture_multiaxial & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strfuture_customisation & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strfuture_pkl & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strfuture_mappinghealthprob & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strorganise_by_audience & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strlike_found_helpful & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strdislike_found_difficult & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strcontinue_develop & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strsuggest_comments & "'"
strSQL = strSQL & ", "
strSQL = strSQL & "'" & strquestionnaire_thoughts & "'"
strSQL = strSQL & ");"
cnnFormToDB.Execute strSQL, adCmdText Or adExecuteNoRecords
```

```
cnnFormToDB.Close
Set cnnFormToDB = Nothing
%>
<h3>Thank you!</h3>
<p>Your input has been successfully saved to HealthCyberMap evaluation database.</p>
<%
End If
%>
<hr color="#0071C0">
<a href="http://healthcybermap.semanticweb.org/" target="_top">Back to Home
Page</a><br><br>
<font size="2">© 2001, 2002 HealthCyberMap.org. All Rights Reserved. (Last revised 17
April 2002)</font>
</body>
</html>
```