Navigation and interaction in medical knowledge spaces using topic maps

Jürgen Beier*, Tom Tesche

hyperCIS AG, Am Köllnischen Park 1, 10179 Berlin, Germany

Abstract

The access to relevant, up-to-date and reliable information is a time critical, but nevertheless very important task in the daily work of physicians and nurses. hyperCIS developed an intelligent information retrieval system (IRS) with a knowledge-guided user front-end and an automatic generation of search engine queries. The medical knowledge of MeSH (Medical Subject Headings) classification was transferred into a Topic Map, the new XML-based knowledge representation standard (ISO 13250). A graphical user-interface allows the fast and associative browsing in networks of themes. Each theme or topic, as a node of the topic map, contains information on title, synonyms, translations, definition, scope, sub- and superclasses. This domain-specific knowledge of a topic is exploited by the hyperCIS IRS to automatically generate search queries. These queries are passed to a text search engine, which determines adequate documents from different configurable categories with customizable contents. The IRS was integrated into the hyperCIS hospital information system and allows the fast and comfortable access to relevant documents fitting to the demands of the current task. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Medical knowledge; Topic maps; Semantic networks; Retrieval tools; Information retrieval

1. Introduction

The medical occupation requires a widespread and up-to-date access to various information sources, for instance, text books, journals, guidelines, medical indexes (PubMed/Medline), selected internet sites, news groups, colleagues and medical experts. An efficient use of these information resources in clinical routine is hampered by the

---

* Corresponding author. Tel.: +49-30-223-370-35; fax: +49-30-223-370-01.
E-mail address: juergenbeier@hotmail.com (J. Beier).

heterogeneity and spatial distribution of the data, implying the necessity to utilize different retrieval techniques (libraries, dialogue, telephone, internet). Using these conventional techniques, a comprehensive research requires a considerable amount of time, which — under most circumstances — is not available.

The reasonable and successful usage of computer-based medical information retrieval tools (library OPAC, Medline, internet search engines, such as Altavista or Google) is hampered by several limitations.

• Internet search engines do not make use of a medical thesaurus but perform simple Boolean text pattern matching. Especially patients, who do not know the right and preferred terms or synonyms of the medical vocabulary to enter, will receive only a small subset of hits.

• Internet search engines collect each web page, disregarding its medical or non-medical content. Thus, the inquirer will receive numerous non-medical web pages, incidentally containing key terms of the query string.

• The search string remains in the language it was entered. A translation to other languages of interest is usually not performed.

• In most retrieval tools, the context of the search (scope, aims) is not regarded and has to be expressed by the user specifying additional search terms.

• The vector space model most search engines use, calculates a hit ranking order based on word frequencies at document level. Unfortunately, an individual weighting of the specified search terms themselves is not possible. This is of importance when taking into consideration the previous list item: the search string may consist of numerous terms. Here, the possibilities of the ranking procedure are only partially exploited, therefore relevant hits are hidden somewhere in the information glut.

In this paper, an information retrieval system (IRS) for medicine [1–5] is presented, which tries to overcome most of the above mentioned pitfalls. The developed IRS enables an interactive navigation through topics of the medical (or another) domain. The knowledge contained at each topic is utilized to control the search for documents or other information related to the query.

2. Materials and methods

For information retrieval, the realized IRS uses the new ISO standard 13250 of Topic Maps (TM) (see www.ornl.gov/sgml/sc34/document/0058.htm and www.topicmaps.org). Similar to semantic networks known from artificial intelligence, topic maps express graph nodes, their description, and interrelations. The basic elements of a TM are "Topics" (themes). TM "Associations" represent the relations between the topics themselves. Extending the basic principles of semantic nets, topic maps connect nodes to related documents, images or other data, referred to in the TM standard as "Occurrences" (Fig. 1).

Our solution transforms the MeSH classification of the National Library of Medicine (Medical Subject Headings, NIH, Version 2000, www.nlm.nih.gov/mesh) with its approximately 19,000 entries into a topic map, thus, allowing an appropriate retrieval of medical information. MeSH is the most used controlled vocabulary for document indexing, deployed by many libraries and by Medline (www.nlm.nih.gov/databases/freemedl.html) as well.
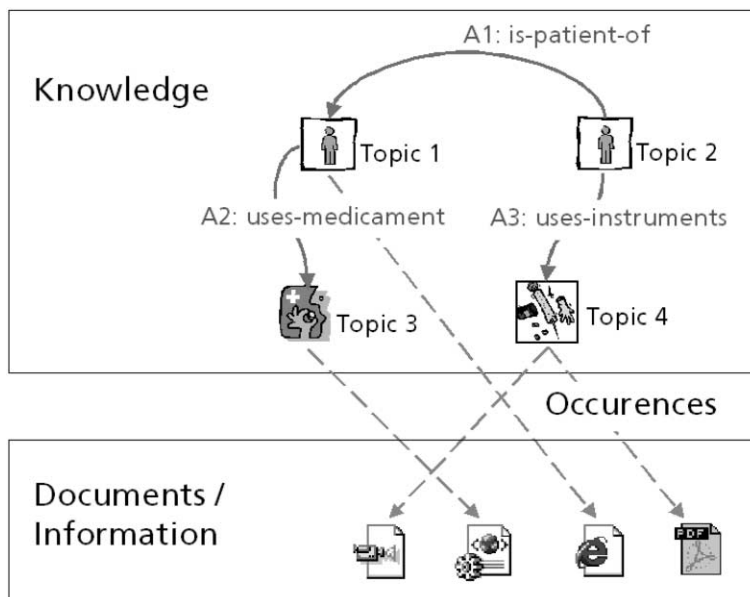
Fig. 1. Topic Maps combine a knowledge representation with information resources. TM consist of topics, which can own types such as "Person", "Drug" or "Examination Device". Topics are interconnected by associations (here, for instance: Topic 1 "is-patient-of" Topic 2). Using hyperlinks of the Xlink standard, topics point to one or more documents. These references are called occurrences.

For the developed IRS, special TM associations were chosen: is-subclass-of, is-superclass-of, has-synonyms, has-preferred-term, is-related-to, is-definition, is-scope-note.

3. Results

The IRS was realized in a two-level architecture: (1) knowledge organized by TM topics, associations and occurrences and (2) documents described by their metadata and full text. After entering a search string, the IRS determines a list of one or more topics that refer to this query concerning its title, synonyms and annotation in German as well as in English. Here, according to the desired scope, the user chooses a topic and navigation within the topic map is started (Fig. 2).

Ongoing from that starting point, the user interactively and graphically navigates through the network of themes, determining the topics that optimally fit to his or her demands. Here, only a small clear subset of the entire topic map is visible to the user. For the currently selected topic, additional information is displayed (MeSH code, definition and annotations, synonyms, translation).

Using these topic commentaries, the system on-the-fly automatically generates a search query, expanding the search string with topic name, synonyms, translations, and definition. Pre-defined weighting factors at search string level control the impact of these different word groups for ranking the hits.
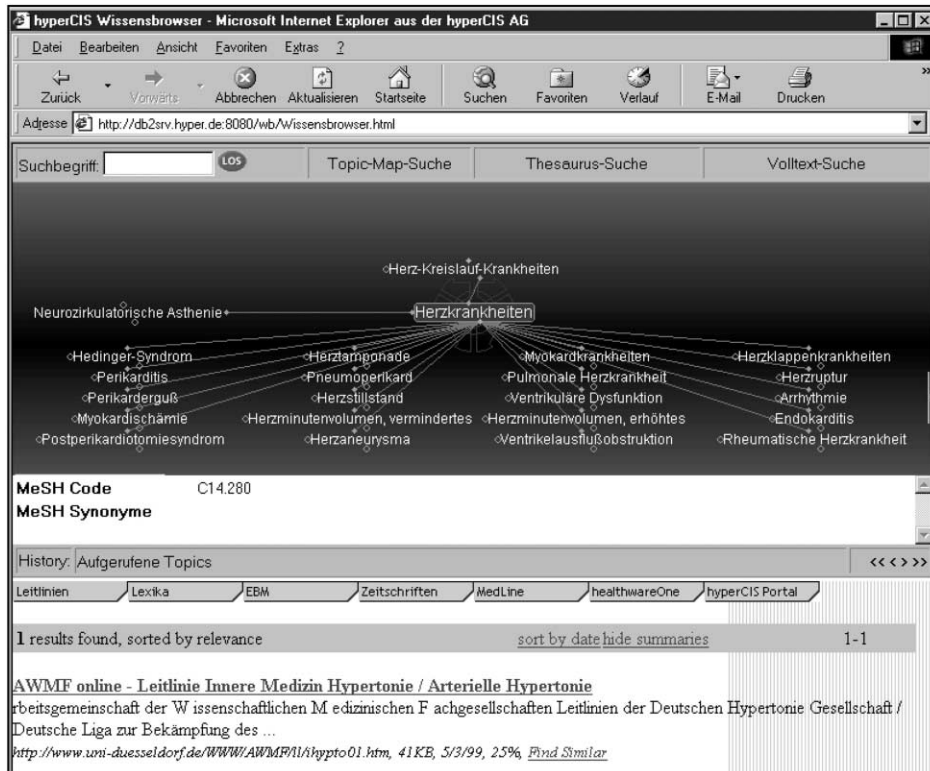
Fig. 2. User-interface of the information retrieval system (English version is in progress). From top to bottom: • Navigation area for topic maps, • Description area of the selected topic, • Cards for document categories: Guidelines, Lexica, EBM, Journals, etc., • Hit list of the category Guidelines.

The document search space was divided into several categories (e.g. AHCPR guidelines, journals, selected internet sites). The categories were defined, extended and customized by the system's administrator. The resulting hits are displayed in a manner known from internet search engines and are grouped according to their category. Fig. 2 demonstrates the features of this knowledge browser for TM navigation and document retrieval.

Besides the MeSH TM, other topic maps were realized (hospital organization structure, knowledge on certain diseases, individual views on the electronic patient record).

4. Discussion

Our approach offers several advantages when searching for medical information in distributed and heterogeneous data sources:

- Due to the implemented client/server architecture and the employment of standard web browsers, the system can be used platform independently.

- Users, unfamiliar with the terminology specific to medicine, are automatically directed to the correct preferred terms by the intelligent thesaurus, scope notes, and definition functions.
- Only one system is necessary to access different types of distributed information. This allows the integration of data on the local file system, in the internet, and in the local hospital intranet.
- Hit lists are no longer ordered only by creation date or alphabet but are classified into configurable categories relevant to individual physicians or clinical departments.

References

[1] K.F.C. Kaltenborn, Informations- und Wissenstransfer in der Medizin und im Gesundheitswesen, Verlag Vittorio Klostermann, Frankfurt am Main, 1999.
[2] D. Paparoditis, Ein Modell für Informationssuchsysteme im Medizinischen Umfeld, Verlag Dr. Kovac, Hamburg, 1997.
[3] D. Paparoditis, Towards a medical terminology infrastructure in an open clinical application environment, in: E. Fleck (Ed.), Open Systems in Medicine, IOS Press, 1995.
[4] W.R. Hersh, Information Retrieval—A Health Care Perspective, Springer, 1996.
[5] R. Ferber, Data Mining und Information Retrieval. Skript zur Vorlesung an der TU Darmstadt, www.darmstadt.gmd.de/~ferber, 1999.