MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Use of Semantic Web Technologies in Medicine and Health Care

Georg Goebel, Karl-Peter Pfeiffer, Karin L. Leitner

Department for Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Austria

ABSTRACT

We present a scenario from a medical context, in which functionalities of intelligent search tools in health care are described. We conclude, that in addition to knowledge representation, independence from modalitites, interactivity, process representation and proactivity semantic intelligence based on Semantic Web Technologies are essential elements of retrieval tools in health care.

INTRODUCTION

The high demand for medical knowledge poses a big challenge for information technology to offer user-friendly systems which help healthy citizens, patients and health professionals to find proper data, information and knowledge.

Medicine has a long history in structured or semi-structured documentation. On the one hand medical documentation of diagnoses has been performed using the ICD-10 (International Classification of Diseases, 10th revision) or other coding systems; on the other hand indexing of scientific literature has been done using key words from MeSH (Medical Subject Headings). Coding systems like ICD, classifications and medical thesauri have been available for years. Scientifically validated terminologies like SNOMED (Standardized Nomenclature in Medicine) and standardised messaging standards like HL7 (Health Level 7) and DICOM (Digital Imaging and Communication in Medicine have been facilitating communications between computer systems and different modalities and have achieved a broad market acceptance within the healthcare industry.

INTEGRATION OF SEMANTIC INTELLIGENCE

Medical queries are among the most popular topics people are searching for in different databases and knowledge sources. Due to the early development of medical domain knowledge sources, most of the coding systems are only available in proprietary, non standardised structures or schemes.

Although there might be no specific field of domain knowledge which has been more penetrated with thesauri, classifications etc, it has taken a long time to accept XML technologies as a standard to meet challenges of medical content management, data communication and medical knowledge representation.

If retrieval systems are supposed to interact with their users in an 'intelligent way' the systems have to include semantic intelligence in addition to text repositories, lexica, hit lists and search indices. Besides technical issues on lexical and morphological levels (see Table), which concern user interfaces, four levels of interest can be identified which can be separated from other aspects like medical natural language processing, information modelling, and user interface design.

The *Syntactical Level* concerns medical grammars and syntactic rules.

The *Semantic Level* includes linking between different coding schemes and classifications. On this level, trans-lingual problems can be targeted.

On the *Discourse Level,* a medical topic is identified and may be linked to other knowledge sources. At this point, the integration of additional sources (e.g. figures about morbidity or resource statistics) may be appropriate. Applications working on a discourse level are able to identify relevant related data sources and search them automatically.

The *Common Sense Level* is a level of abstraction comparable to the 'general medical knowledge' of a medical specialist.

Intelligent functionalities are a key issue in medical information retrieval systems even if the availability of data in XML format enables conversion of human readable data to machine readable data. Challenges of intelligent semantic integration concern, for example, the description of the health status of patients whose data could be used for different queries according to various contexts.

SCENARIO FOR THE USE OF A MEDICAL AGENT BASED ON WEB SERVICES

Mary Jones is 49 years old. For the past 4 years she has been noticing variations in the length of her monthly cycle and pattern of bleeding. Moreover, she has been experiencing hot flushes, night sweats, vaginal dryness, as well as joint pains. She suffers from unpredictable mood swings. Mary has heard of HRT, but usually in the context of its associated risk of breast cancer. On the other hand Anne, her neighbour, has told her that she should start HRT, as it would greatly alleviate her suffering, prevent her bones from fracturing in her old age and protect her from cardiovascular disease. Moreover, Anne said that, according to the latest research, HRT reduces the risk of Alzheimer's disease. Anne feels quite confused and decides to see Dr. Eleanor Trevor, her local GP, about this issue. Dr. Trevor understands Mary's concerns, as HRT has been the subject of ongoing debate for many years. She knows that the fear of breast cancer is one of the prime reasons for rejection or discontinuation of HRT. And even though HRT had been promoted for many years in relation to several health issues such as prevention of osteoporosis and cardiovascular disease, Dr. Trevor is aware that recent research suggests not to use HRT for prevention of osteoporosis and that it may actually increase the risk of cardiovascular disease. She knows that in addition there are several other organs affected by the hormonal constituents used in HRT such as the endometrium, colon and central nervous system. Moreover, it depends very strongly on the individual person receiving HRT whether it is useful or may actually be harmful. She wonders about Mary's physical constitution (her age, body mass index, parity, her age when Tim, her first child was born, etc.) and risk factors (Mary is overweight and smokes). Dr. Trevor lets her search agent support her in this issue. She is glad she has this tool available because it is near-impossible to stay up to date with the latest results in medical research. Before she had her agent, she would have had to look for best evidence in databases such as CDSR (Cochrane Database of Systematic Reviews) or DARE (Database of Abstracts of Reviews of Effects), search biomedical databases such as Medline or Embase, search the internet or even hand search the literature. She was glad she had done a course in searching for evidence, as she knew from a colleague who didn't even try to treat according to best practice, as he didn't know how to find the evidence. After finding the evidence herself she would have had to apply it to the individual patient. She would have had to go through all the patient notes, call the hospital and other specialists for any additional information needed, and the decisions would have been based mainly on her own expertise and experience, weighing risks and benefits of a particular treatment. This whole process became much more convenient with her agent. Basically, the search agent performs all tasks of information retrieval, integration with patient information, and knowledge representation automatically, in a speedy, comprehensive, reliable and safe manner. Dr. Trevor feels that it provides her with many benefits such as saving her time, supporting her in her decisions, and ultimately enabling her to offer better patient care. When she lets the agent run over Mary's particular case, it automatically searches for the best evidence currently available in the field of HRT, retrieves Mary's online health record (a health record pulling information together from all medical facilities Mary had been visiting), detects that Mary also has high blood pressure and a positive family history of breast cancer, which Dr. Trevor hadn't been aware of, and independently determines the overall risks (breast cancer, blood clots, stroke and coronary heart disease) and benefits (fracture reduction and reduced risk of colorectal cancer) HRT would have in Mary's case. The agent presents its findings to Dr. Trevor who is very satisfied with the feedback, comments and helpful decision support. She tells Mary that firstly she should try to alter her lifestyle – eat healthy, exercise regularly and quit smoking. She also lets her know that there are several alternative therapies around that may or may not be helpful in relieving menopausal symptoms but that in general, there is more research needed in that area. She remarks that herbal therapies may have adverse side effects or exhibit harmful interactions with other medications. She tells Mary that HRT should be considered only a short-term option, as in the long run, according to the best evidence currently available and in consideration of Mary's status the risks do outweigh the benefits.

	Medical Application / Challenge	(XML) Application (Examples) / Current projects
Phonological level	Free Text Speech Recognition	VoiceML
Lexical level	Medical Term Recognition	HL7-XML, ICD-10, MeSH
Morphological level	Stemming, Compound analysis Linking of medical terms	
Syntactical level	Medical Grammars	SNOMED
Semantic level	Knowledge Representation, linking of medical concepts	UMLS, RDF(S), DAML-OIL, GALEN
Discourse level	Linking of medical topics	Topic Maps
Common Sense level	Health Agents	Intelligent Web Services

Lexical and morphological levels for medical applications

PRESENTATION OF SEARCH RESULTS

***Personalisation*:**
if a doctor is seeking an effective treatment for heart failure, the search results must be customised according to his needs and may depend on previous queries and interactions. Drill-down from top level overviews to specific topics may lead to different search results depending on several parameters like regional, cultural or temporal circumstances and the professional background.

Aggregation / Navigation:
'Search results' will imply a new meaning: because medical documents may be virtual and – as they contain structured information – can be considered databases, future search results will not be static links but link-traverse triggers (re)calling the search (web) service based on relevant feedback algorithms.

Ranking:
Due to its implications, medical information must be as reliable as possible. Quality assurance strategies like HON, MedCertain etc. are based on optional cooperation of content providers. Within this context, XML-based indexing and rating services offer technical access and rating controls. The majority of current retrieval engines are using term or link frequencies as ranking criteria, which doesn't include assessments on semantic or 'common sense' level.

SEARCH FUNCTIONALITY FROM THE VIEWPOINT OF USERS

Communication between users and the source applications is usually managed by mediator applications [11] which encapsulate the communication APIs of the various systems and are responsible for the interpretation of information (its structure, syntax and semantics). Thus, a 'medical search angel' may consist of a *communication manager* encapsulating the communication API of the system, a *syntax manager*, which may itself be associated with coding modules for the various types of syntax used in different systems, such as ICD-10, HL7, etc. and a *semantics manager*, which carries out "semantic mapping" between the terms of the different vocabularies used in the different systems and converts the data.

Virtual Case Managers (VCM) must be able to interact with users and determine their personal backgrounds. Up until now, websites have generally been developed separately for patients and doctors. A VCM derives information about the individual background from the vocabulary used by and interactions with the person or by posing additional questions.

SEARCH FUNCTIONALITY FROM THE VIEWPOINT OF A SEARCH SYSTEM

Medical search engines will include "Medical Knowledge" for better understanding and processing of user queries. Linking information sources on a supra-sentential level goes far beyond ordinary text-linking on a syntactic or morphological level. XML plays a key role in different layers of knowledge engineering (see Table 2): searching heterogeneous knowledge sources is a domain of multi-agent system applications as well as representing and linking semi-structured information. Within all layers, XML technology is used to increase independency from specific models and application systems.

Address for correspondence

Georg Goebel, Ph.D.
Department for Medical Statistics, Informatics and Health Economics
Schoepfstrasse 41
A-6020 Innsbruck, Austria
Email: georg.goebel@uibk.ac.at