# 1

# Demand for Intelligent Search Tools in Medicine and Health Care

Karl P. Pfeiffer, Georg Göbel, and Karin Leitner

## 1.1 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 [294]) or other coding systems; on the other hand indexing of scientific literature has been done using key words from MeSH (Medical Subject Headings [213]). Coding systems like ICD, classifications and medical thesauri have been available for years. Scientifically validated terminologies like SNOMED (Standardized Nomenclature in Medicine [291]) and standardised messaging standards like HL7 (Health Level 7 [155]) and DICOM (Digital Imaging and Communication in Medicine [99]) have been facilitating communications between computer systems and different modalities and have achieved a broad market acceptance within the healthcare industry. 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.

In March 2003 the BMJ (British Medical Journal) Publishing Group started the first excerpt from BestTreatments, a website built for patients and their doctors that looks at the effectiveness of treatments for chronic medical conditions, based officially on “Clinical Evidence”, which is recognised internationally as a gold standard for evidence-based information [226]. From

H. Blanken et al. (Eds.): Intelligent Search on XML Data, LNCS 2818, pp. 5–18, 2003.

**Table 1.1.**

| | Citizen, Patient | Medical Professional |
|---|---|---|
| General Medical Knowledge Facts & Figures | Health Information System Patient Information System | Health Professional Information Systems EBM - Sources |
| Personalised Medical Information | Personal Health Record | Electronic Health Care Record, Electronic Patient Record |

Table 1.1 it can easily be deduced that health professionals and patients need the same evidence-based information, served up in parallel, drawn from the same sources. Additionally, personalised medical and health information on patients – allergies, medications, health plans and emergency contacts should be accessible online from any location via the internet, always being aware of data safety and security. Sharing this information in case of emergency with doctors and processing it with powerful health tools (e.g. immunization planner, health risk appraisal, personal health plan) are fascinating challenges for scientists and industrial vendors.

## 1.2 Medical Knowledge Representation: Status Quo

Medical knowledge bases *contain all the knowledge and experience to be invoked by a reasoning program to produce advice that supports a decision* [301]. Generally, medical knowledge is retrievable from

- the medical literature (documented knowledge)
- experts in a specific domain (clinical experience).

Some authors distinguish between knowledge about terminology – conceptual knowledge – and more general (medical) knowledge – general inferential knowledge [246]. Neither current medical literature nor experiences from experts today are usually being processed in a comprehensive form which supports medical decisions. One mechanism to transfer experiences from one person to another is the creation of knowledge bases, which can potentially provide health care workers with access to large bodies of information. Knowledge, the next step of complexity, must then be expressed in the terminology and semantics of the knowledge base and several methods must be designed to acquire knowledge from experts. V. Bemmel [301] defines a medical knowledge base as *a systematically organised collection of medical knowledge that is accessible electronically and interpretable by the computer. ... Usually medical knowledge bases include a lexicon (vocabulary or allowed terms) and specific relationships between the terms.*

This definition does not specify a detailed formalism of how relationships can express expert knowledge. Thus long term research projects like UMLS

(Unified Medical Language Systems) [168] or GALEN (Generalised Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine) [130] showed the necessity of methodologies to map medical knowledge to machine-readable information sources in different languages.

### 1.2.1 UMLS

The UMLS project [168] is a long-term NLM (National Library of Medicine) research and development effort to facilitate the retrieval and integration of information from multiple machine-readable biomedical information sources. The goal is to make it easier to develop systems that link information from patient record systems, bibliographic databases, factual databases, expert systems etc. The UMLS can also facilitate the development of data creation and indexing applications. It consists of three parts:

- The Metathesaurus contains semantic information about biomedical concepts, their various names, and the relationships among them. It is built from thesauri, classifications, coding systems and lists of controlled terms that are developed and maintained by many different organisations. It includes many concepts from a broad multilingual range of vocabularies and classifications (e.g. SNOMED, ICD, MeSH).
- The Semantic Network is a network of general categories or semantic types to which all concepts in the Metathesaurus have been assigned.
- The specialist lexicon contains syntactic information about biomedical terms and may cover the majority of component terms in the concept names presented in the meta-thesaurus.

### 1.2.2 GALEN

In Europe a consortium of universities, agencies and vendors has formed the GALEN project to develop standards for representing coded patient information from 1992 to 1995 [130]. This initiative is based on a conceptual knowledge formalism, which categorises conceptual knowledge (knowledge about terminology) into three subtypes (conventional knowledge, descriptive knowledge, formal knowledge). To address the problems of clinical terminologies, GALEN is constructing a semantically sound model of clinical terminology - the GALEN Coding reference (GALEN CORE) model. This model comprises

- elementary clinical concepts such as 'fracture', 'bone', 'left', and 'humerus';
- relationships such as 'fractures can occur in bones', that control how these may be combined;
- complex concepts such as 'fracture of the left humerus' composed from simpler ones. This compositional approach allows for detailed descriptions while preserving the structure provided by the individual components.

## 1.3 Challenges for Information and Knowledge Management in Medicine

Hospital information (health care) systems are usually composed of a distributed set of heterogeneous applications. The efficient management of a patient may involve the exchange of information and the co-operation among diverse applications [324]. Most of the contents of medical databases today are designed for humans to read, not for computer programs to manipulate the data in a sensible way.

The electronic health care record (EHCR) will play a key role in the development of future health care systems. Patient-centred storage of data and communication between different health care providers using different information systems requires not only a well-defined communication standard but also precise definition of the data in order to further use it, for example in decision support or personalized search for medical knowledge.

Computers can adeptly parse information for layout and routine processing – here a header, there a link to other information – but in general, computers have no reliable way to process the semantics: this content deals with Morbus Parkinson (MP), and MP can be linked to several information sources, but also this content may constitute information contained in an EHCR of the MP patient called *John Smith*, who *has been visiting* the hospital *monthly* due to his condition, always on *Friday*. Instead, these semantics were encoded by clinical nurses or doctors. In future, intelligent agents will "know" how to perform this task and will also be able to find out the meaning of semantic data by following hyperlinks to definitions of key terms and rules for reasoning about them logically.

From the technical viewpoint, two important technologies are already in place: XML and RDF (Resource Description Framework). XML is used by RDF, which provides the technology for expressing the meaning of terms and concepts in a way that enables computers to process and "understand" this information. RDF uses XML for its syntax and URIs to specify entities, concepts, properties and relations.

If a program wants to compare or combine information across two records or databases, it has to know that two different terms may be used to designate the same concept (e.g. LARGE BOWEL – COLON), they may be synonyms. But in other situations the same word may have different meanings in different contexts (e.g. CERVICAL may refer to the cervix – an anatomical part of the womb, or to the neck). They are so-called homonyms. Ideally, the program must have a way to discover such common meanings for whichever database it encounters. For this purpose, ontologies within a domain field must be built. Usually these types of ontologies include a taxonomy and a set of inference rules. The taxonomy defines classes of objects and relations among them [30]. Inference rules in ontologies enable computers to manipulate the terms much more effectively for the human user: an ontology may express the rule: "If a patient is associated with a social insurance number of a certain insurance, and a prescription contains this insurance number, then the patient receives drugs based on guidelines of that specific insurance."

An example from the clinical environment shall be given that illustrates the potential use of search agents supporting physicians in routine patient care. The search agent uses heterogeneous databases, automatically makes inferences from the EHCR and advises the physician regarding specific medical issues. This scenario demonstrates the use of a search agent during a consultation with a woman experiencing typical symptoms of menopause who is concerned about hormone replacement therapy (HRT), one option to successfully treat these symptoms.

*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 [77]) or DARE (Database of Abstracts of Reviews of Effects [88]), search biomedical databases such as Medline or Embase [106], 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.*

Software Agents in medicine run without direct human control or constant supervision to accomplish goals provided by medical experts. Agents typically collect, filter and process information found in distributed heterogeneous data sources, sometimes with the assistance of other agents. It will be a big challenge in the future to train these agents to find the appropriate and very specific information for a patient with certain symptoms, diagnoses or treatments.

## 1.4 Data, Information, and Knowledge in Medicine

Data management in the medical field will be expanded by information processing techniques using different heterogeneous data sources like Hospital IS, Laboratory IS etc. Data will be compiled to information as well as prior and explicit knowledge. Knowledge acquisition, modelling, storage and processing methods will offer users new possibilities to access global medical knowledge bases [293]. Bibliographic databases like PubMed [240] will evolve to become global knowledge sources offering *up-to-date, relevant, high-quality, multimedia* and *multilingual* knowledge to consumers all over the world. From the viewpoint of medical experts and based on the experiences with decision support systems and expert systems, *content* and *data / information representations* are key challenges for knowledge engineers.

### 1.4.1 Medical Content

Relevance, context-based personalisation, links to related topics, integration of EPHR-data are viewed as challenging issues for architects of Medical IS in the future.

Structuring and 'semantic labelling' of medical topics constitute an increasingly important focus of interest [21]. These tasks must be performed in an efficient way at the point of information acquisition and storage even though context-based personalisation of contents is of interest to users only at medical content retrieval. By adding structure where there is none, users can query and re-use knowledge. The usefulness of existing resources can be extended if they are represented within a structured knowledge framework.

Information, knowledge and experience existing only in people's minds can be connected with data. This creates intelligent resources that can be stored, navigated, re-used and expanded. Term- or word based linkage of information will be extended by dynamic multidirectional semantic linkage of topics to powerful link engines, which improve their functionality and scope automatically and autonomously.

Medical knowledge bases face an increasingly important issue: considering the increase of information available and the rise of the users' demand for personalised content, why should a customer be satisfied with contents from knowledge bases if they are not suited to their particular requirements? Content managers need to be able to find flexible ways of structuring the content and delivering it to different users according to their individual requirements. This will even lead to the integration of personal clinical data gathered from the customers' individual electronic health care records.

In future, users' individual behaviour will affect not only representation of knowledge, but knowledge acquisition- and structuring strategies based on retrieval agents will be customised and influenced by user (inter-)actions as well. Within the scope of learning curves, domain-specific knowledge will be separated from process-oriented knowledge. In this context, customers' interactions will be recognised and processed as well as behavioural patterns of user groups.

XML technologies enable managing semi-structured documents and structured data in a uniform way. Hence, it should be possible to mix data and documents in order to build *virtual health related documents issued from several sources.*

### 1.4.2 Knowledge Representation

Future targets of Knowledge Representation and Visualisation are the "representation of complex information for discovery of hypothesis and illustration of results and coherences" [234].

### 1.4.3 Independence from Modalities

It is commonly recognized in work on multimodal information presentation that much of the true value of modality-independent representation lies in appropriate presentation of non-identical but overlapping information. Textual representations and graphical visualisation have different strengths and weaknesses and so their combination can achieve powerful synergies. Conversely, simply placing textual and graphical information together is no guarantee that one view is supportive of another: if the perspective on the information taken in a graphical and in a text representation have no relation (or worse, even clash), then the result is incoherence rather than synergy.

Despite these problems, independence from modalities is one of the key issues for users working within the medical field. Especially health professionals must have access to their knowledge bases via different modalities like TV or computer screens, table PCs, handheld devices, shutter glasses etc.

### 1.4.4 Interactivity

Interactive multimedia can be used to provide information, to train, educate, entertain, store collections of audiovisual material, as well as distribute multimedia and allow for user input. The range of tools, including the well-known PC-mouse or keyboard, will be extended by new interaction tools (e.g. haptic or vocal tools). Nonlinear media structures will challenge the development of powerful knowledge browsers controlled via voice interfaces.

### 1.4.5 Process Representation

One key issue in representing medical knowledge is the organisation of process-oriented knowledge ('How-to knowledge'). Formal methods like UML (Unified Modeling Language) extend production rules, semantic networks and frame-technology (terminology-based knowledge) in order to describe medical procedures, processes, treatment protocols etc. in a reusable, system-independent way.

### 1.4.6 Pro-activity

Electronic guides assist users during their work and try to detect and solve problems beforehand. Additionally, users may receive notifications on updates or new documents according to specified criteria. Already today there are several rule based systems which perform an automatic analysis of drug prescription to avoid drug interaction or just look for cheaper drugs or produce automatic warnings in the case of allergies.

## 1.5 Expectations from an Intelligent Search Engine

As medical environments become increasingly electronic, clinical databases are continually growing, accruing masses of patient information. This wealth of data is an invaluable source of information to researchers, serving as a test bed for the development of new information technologies and as a repository of real-world data for data mining and population-based studies or clinical epidemiology. However, the real value of this information is not fully taken advantage of, in part because of issues concerning security and patient confidentiality, but also due to the lack of an effective infrastructure to access the data. Users want to query and retrieve data from multiple clinical data sources, automatically de-identifying patient data so it can be used for research purposes. Key aspects may be XML-based querying of existing medical databases, easy inclusion of new information resources, minimal processing impact on existing clinical systems via a distributed cache, and flexible output representation (e.g. via XSL(T)).

Despite the fact that several authors distinguish between fact retrieval and information retrieval, many authors are noticing that all these methods are used to manage knowledge *between* people. Therefore, aspects of human knowledge processing, including problems and methods from cognitive psychology, psycholinguistics and psycho-mnemonics, must be considered. It is widely accepted that human information processing problems can not be fully explained or solved using well-known models like deductive logic or neural networks.

Concept systems enable the representation of complex clinical states or procedures through relational links between concepts, expressing e.g. negation, quantification, temporal relations, localisations and precise or fuzzy time spans. The concept ABDOMINAL PAIN, for example, could have the LOCALISATION 'RIGHT UPPER QUADRANT' and 'RADIATING TO RIGHT SHOULDER'. It could have the QUALITY 'EPISODIC' and 'SEVERE'. It could have its ONSET 'AFTER FATTY MEAL'. Further, it could have a DURATION of 'SEVERAL HOURS'.

ENV 1828 [107] provides a categorical structure and combinatorial rules to support the exchange of surgical procedure information between different national and international coding systems and languages within Europe.

It covers surgical procedures of all surgical disciplines and includes definitions and examples of concepts (e.g. ENDOSCOPE), different types of relations (e.g. INTERVENTIONAL EQUIPMENT is a super-ordinate concept to ENDOSCOPE in a generic relation;, HEAD, FACE, NECK, CHEST are sub-ordinate concepts to HUMAN BODY in a partitive relation), semantic links (e.g. BY MEANS OF; FROM), modifiers (e.g. RIGHT; LEFT; TWO; THREE) and combinatorial rules (e.g., removal of the kidney for cancer: cancer is not necessary, it is a removal of a kidney). In order to enable health information systems architects to use well-known IR techniques like (semantic) query expansion, text classification algorithms etc., this standard needs to be integrated into their system. Thus, new requirements may challenge traditional DB techniques (e.g. relational or object-oriented data modelling, SQL query formulation) which might, in fact, prove insufficient in this new context.

The complexity and diversity of medical information and knowledge representation goes far beyond the scope of the traditional information sources. A library, for example, does not provide quality rankings of the works in the collection. Because of the greater volume of networked information, Internet users want guidance about where to spend the limited amount of time they have to research a subject.

They may need to know the three "best" documents for a given purpose. They want this information without paying the costs of employing humans to critique the myriad of Web sites. One solution that again calls for human involvement is to share judgments about what is worthwhile. Software-based rating systems have to let users describe the quality of particular Web sites.

## 1.6 Search Functionalities from the Viewpoint of Users

Communication between users and the source applications is usually managed by mediator applications [324] 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.

## 1.7 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 [156], MedCertain [214] 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.

## 1.8 Search Functionalities 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 1.2): 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.

## 1.9 Integration of Semantic Intelligence

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 1.2), 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 [100].

**Table 1.2.**

| | 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 |

- 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.

## 1.10 Conclusions

The medical discipline is currently moving from data representation towards information and knowledge representation. Hospital and clinical information systems are changing from database to multimedia information- and knowledge-based systems and may evolve into pro-active, self-learning systems. Health care has a long-standing tradition of standardized documentation, using different catalogues for diagnoses and procedures. The main purpose of this type of documentation has been administration or financing rather than medical documentation. The complexity of medical contents and the need for representing time-dependent relations, detailed descriptions of symptoms, findings and therapies etc. make it necessary to have the main documents available in free text, in order to achieve the precision and granularity required for personal communication. In recent years a lot of work has been done to extend structured documentation to replace free text and set up concepts for the representation of complex medical situations and processes.

This also applies to the representation of health related knowledge for patients and health professionals. The ever-increasing amount of scientific medical literature available via the internet as well as the countless web pages with health related information for patients necessitate very special tools to achieve specific and precise search results. One precondition for a successful search is to have a certain extent of structure in the data, e.g. using object-attribute-value triples and standardized terminologies. Standardized, controlled medical terminologies like the ICD-code, SNOMED or MeSH may constitute a starting point for the identification of key words in free text, but are not sufficient to represent the dynamic relationships between different objects, making modifiers and semantic links necessary.

There is also a long-standing tradition of standardized communication of data in medicine using e.g. HL7. Nevertheless, the next generation of health care information systems must not only allow exchange of data between heterogeneous systems, but must enable the representation of complex medical contents as well. Intelligent search engines, virtual agents and very specific data analysis tools will process semi-structured data and will help make the latest, quality assured information available for health care professionals, financiers and patients. Health information generators will search for individual person-centred information in the web using person-specific information from the EHCR and will design health promotion programs based on the latest evidence-based, quality-assured knowledge. Structured routine data in a hospital or clinical information system will be analysed for clinical studies looking at e.g. long-term side effects, epidemiology, quality standards, cost effectiveness, outcomes etc.

The expected functionalities from intelligent search engines in medicine can be summarized by the following points:

- Extraction of the most important facts from an electronic health care record or personal health care record, which may be distributed across various types of databases at different locations, using a master patient index
- Expressing these facts as medical concepts and linking them using ontologies and specific medical grammar to represent the complex health status of a patient in a semi-structured way
- Searching in quality-assured information systems for the latest information, relevant specifically to the individual person
- Searching for the state-of-the-art treatment for a certain disease analysing the outcomes of the latest clinical studies and using evidence-based medicine databases

- Transforming patient queries expressed in a common, non-medical language into queries using standardized medical terminology
- Identifying synonyms or homonyms and interpreting them correctly via their medical context
- The mathematical models of the search are multi-dimensional similarity or distance functions, taking the uncertainty and imprecision of medical terms into account
- Pro-active generation of health information and creation of prevention, treatment, rehabilitation and wellness programs for the individual consumer, presented in lay terminology.

Data, information and knowledge stored in different information systems will form one lifelong, virtual electronic health care record and will help improve efficiency, effectiveness and safety of the health care system.