An information retrieval system for computerized patient records in the context of a daily hospital practice: the example of the Léon Bérard Cancer Center (France)

P. Biron[1]; M.H. Metzger[2]; C. Pezet[1]; C. Sebban[1]; E. Barthuet[3]; T. Durand[1]

[1]Léon Bérard Cancer Center, Lyon, France; [2]Université Lyon I – CNRS-UMR 5558, Lyon, France; [3]SWORD, Saint Didier au Mont d'Or, France

Keywords

Electronic health records, data mining, information storage and retrieval, user-computer interface

Summary

Background: A full-text search tool was introduced into the daily practice of Léon Bérard Center (France), a health care facility devoted to treatment of cancer. This tool was integrated into the hospital information system by the IT department having been granted full autonomy to improve the system.

Objectives: To describe the development and various uses of a tool for full-text search of computerized patient records.

Methods: The technology is based on Solr, an open-source search engine. It is a web-based application that processes HTTP requests and returns HTTP responses. A data processing pipeline that retrieves data from different repositories, normalizes, cleans and publishes it to Solr, was integrated in the information system of the Leon Bérard center. The IT department developed also user interfaces to allow users to access the search engine within the computerized medical record of the patient.

Results: From January to May 2013, 500 queries were launched per month by an average of 140 different users. Several usages of the tool were described, as follows: medical management of patients, medical research, and improving the traceability of medical care in medical records. The sensitivity of the tool for detecting the medical records of patients diagnosed with both breast cancer and diabetes was 83.0%, and its positive predictive value was 48.7% (gold standard: manual screening by a clinical research assistant).

Conclusion: The project demonstrates that the introduction of full-text-search tools allowed practitioners to use unstructured medical information for various purposes.

Correspondence to:

MH Metzger
Unité d'Hygiène et d'Épidémiologie, Hôpital de la Croix-Rousse
103 Grande-Rue de la Croix-Rousse
69317 Lyon CEDEX 04, France

Appl Clin Inform 2014; 5: 191–205
DOI: 10.4338/ACI-2013-08-CR-0065
received: August 28, 2013
accepted: December 31, 2013
published : March 5, 2014
Citation: Biron P; Metzger MH; Pezet C; Sebban C; Barthuet E; Durand T. An information retrieval system for computerized patient records in the context of a daily hospital practice: the example of the Léon Bérard Cancer Center (France)Appl Clin Inf 2014; 5: 191–205
http://dx.doi.org/10.4338/ACI-2013-08-CR-0065

1. Introduction

In the last 20 years, the development of search engines has dramatically changed the techniques of information retrieval on the Internet, allowing users to instantly find relevant information by typing keywords into a search interface. However, the implementation of these techniques to search for medical information in computerized patient records is still very new and is essentially experimental, most often referred to research in university hospitals [1, 2]. Few studies have described the value and utility that full-text search tools might have in real time in healthcare environments [3]. Projects to develop such tools are underway, but these efforts are still mostly limited to research projects [4]. A team at the University of Michigan Comprehensive Cancer Center confirmed that the use of a search engine optimizes the recovery of medical data in the free text sections of computerized patient records, and may allow physicians to save time [5]. Such a tool can facilitate collaboration by allowing users to analyze and share different types of information between people and across disciplines [6].

Léon Bérard Cancer Center (France) wanted to provide practitioners with an IT solution that supported full-text search of patient medical records using natural language. The tool has two different functions: firstly, it allows users to search within a single electronic medical record; and secondly, it allows queries on all electronic medical records in the hospital database.

The objective of this paper is to describe the methods used to develop and deploy the solution, and explain how the solution is used in the daily hospital practice of a healthcare facility.

2. Methods

2.1 Computerized patient records at the Léon Bérard Cancer Center

2.1.1 Pioneer in computerization of patient records

The Léon Bérard Cancer Center (CLB), based in Lyon, is the regional center for diagnosis and care of cancer in the Rhône-Alpes region (France). The CLB is a referral center for oncology care for patients in the region, and is also involved in the research and development of innovative treatments at national and international level. In 1989, the CLB Management Board decided to computerize patient records. Computerization began in January 1993, with paper versions of medical records not used after July 2002.

This patient record system was developed by the CLB's internal IT and medical teams using web technology. Contrary to the recommendations of the French government, which at the time promoted specific IT packages, the CLB favored internal development to allow greater responsiveness to the changing requirements of the medical staff. Consequently, the IT department was able to design all aspects of the full-text search engine for computerized patient records, allowing this project to move forward.

2.1.2 Description of the computerized patient record system

The computerized patient record system is shown in ►Figure 1. It consists of a data acquisition module, a storage module, and a web form module for data query.

The data acquisition module is composed of different systems that correspond to diagnostic and care activity. Practitioners enter the required data, depending on the medical procedure that was performed (e.g., data from medical consultation, pathology data, chemotherapy, etc.), ensuring the traceability of these procedures in the database. In addition, documents produced outside of the institution (e.g., letters from treating physicians) are digitized and subjected to character recognition.

For reasons related to volume and performance, data entered into the medical record are stored in one of two instances of the database, according to their origin: one instance for biological data and the other instance for other data in computerized patient records. As of April 20, 2013, the SQL Server database contained 700,000 biological results.

Textual data

Two types of textual data are stored in the medical record system:

- Natural language data are stored on an SQL Server database (Windows Server 2008 R2; SQL Server 2008: 64-bit; .NET Framework 3.5 sp1). These reports are classified according to their type: consultation, pathology (including reports of pathology, cytology, and molecular biology), surgery, endoscopy, imaging, hospital discharge summary, and miscellaneous. As of April 20, 2013, the database contained 4.4 million documents.
- Letters written by physicians participating in the care of the patient outside the CLB are stored on a file server (network-attached storage). As of April 20, 2013, the database stored 1 million of these documents.

Structured data

A number of structured data describing diseases and their management are available in the medical record system.

- ADICAP codes: cyto-pathological-anatomical coding (French language), published by the Association for the Development of Information Technology in Anatomy and Pathological Cytology [7]. These structured data are entered into computerized medical records by pathologists.
- Codes of the International Classification of Diseases for Oncology, Third Edition (ICD-O-3): These structured data are entered into computerized patient records by clinicians.
- Chemotherapy codes: These codes are organized in a thesaurus of chemotherapy, developed by the CLB. These structured data are entered into computerized medical records by clinicians and nurses.

A web form allows the user to query computerized patient records using a full-text search engine. ►Figure 2 shows an example of a query for structured data:

"What is the number of patient records in the database of the cancer center with:

- a moderately differentiated squamous cell carcinoma, more or less keratinizing (code ADICAP „E7T3")
- located on the border of tongue (Code: C02.1)
- with the stage of T1 N0 M0 (TNM Classification of Malignant Tumors)?"

►Figure 3 shows an example of a query for keywords in the patient record database; a screenshot of the response is shown. Medical documents containing the keywords are selected. For each search result, an extract of the document (specifically, the portion of the document containing the queried keyword) is displayed, and the type of document is specified (hospital discharge report, letter to a practitioner, etc.). By clicking on the search result, the user can view the full corresponding document.

2.2 Development of a full-text search tool for computerized patient records

2.2.1 Project design

To make better use of data available in computerized patient records, the IT department proposed to work with clinicians to incorporate semantic analysis tools designed to extract and enrich medical data written in natural language. A multidisciplinary project team was established, consisting of a large panel of users, especially medical radiologists, physicians, clinical research assistants, computer specialists, the medical consultant for the development of computerized patient records, and the chairman of the medical committee of the institution.

The IT department solicited SWORD Group, an IT services company with experience of developing solutions in the field of full-text and semantic analysis, and with extensive experience of medical IT. These various solutions were presented to the medical team during brainstorming sessions aimed at selecting the most relevant functionalities for medical activities. The multidisciplinary nature of these brainstorming sessions was very effective. Indeed, although the proposed IT functionalities were based on complex semantic technologies, clinicians formulated much simpler needs, envisioning tools close to modern Internet search engines.

Two goals emerged from these brainstorming sessions. The first goal was to enable a search for medical information within a computerized patient record during the patient's stay in the hospital. The second goal was to allow a search for medical information on all (or a subset of) medical records in the CLB database, in order to retrieve data for evaluation analysis, research, financial analysis, etc.

The goal finally adopted was the development of an application to search for terms found in any text document, possibly in combination with structured data in the medical record.

2.2.2 Technical description

Overall architecture

The technology used for text mining in our healthcare facility is based on Solr, an open-source search engine which indexes content sources, processes query requests, and returns search results. Solr is a web-based application that processes HTTP requests and returns HTTP responses. The Solr search engine (http://lucene.apache.org/solr/) is provided by the Apache Software Foundation, and is based on the Apache Lucene library (http://lucene.apache.org/core/), which provides language analysis functionalities and core search features.

To integrate this tool into the information systems of the CLB, it was necessary to use a data-processing pipeline to retrieve data from different repositories, normalize and clean it, enrich it, and publish it to Solr. This pipeline is similar to an Extract, Transform, and Load (ETL) tool, which performs a data integration process for transferring raw data from a source server to a data warehouse on a target server. However, ETLs usually handle structured data only, whereas this pipeline also manipulates content from unstructured documents. This component, which was previously developed by SWORD Group, made the implementation of the search platform much easier and quicker. It was also necessary to develop user interfaces to allow users to access the search engine within the computerized medical record of a patient. This development was conducted by the CLB IT department.

Data indexing

Three data sources are indexed by the search engine, as follows: computerized medical records, stored in an SQL Server database; PDF documents generated inside the CLB; and PDF documents from practitioners outside the CLB. Some of the indexed PDF documents are PDF images, from which it is not possible to directly extract text for indexing in the search engine. In these cases, optical character recognition (OCR) is performed on the scanned documents when they are integrated into the medical record, and the textual output of OCR is indexed by the search engine. The search engine indexes changes every night and those records and documents that have been created or updated during the day are (re-)indexed at night so that the changes are available in the search engine the next day.

A full indexing is performed every weekend. During this process, the search engine's content is deleted and the contents of all data sources are indexed. Any documents and records that have been deleted since the previous indexing are deleted from the index.

Query workflow

►Figure 4 details the query workflow. The user accesses the web client by clicking on a button („full-text search") available on the computerized medical record screen.

The user makes a request in the web form and the queries entered into the web interface are submitted via HTTP to Solr.

The search engine returns XML documents containing the results corresponding to the query criteria. These XML documents are used to display results to users within the usual medical record interface.

Data security and confidentiality

A procedure was implemented to assure data security and confidentiality. This procedure defined the rights of authorized users regarding use of the search engine in a single medical record or within different medical records of the facility. Any query to the search engine requires the user to enter a password. All queries are traced in a database (including the name of the user, the search criteria, and the date and time of the search). A reminder screen regularly displays the confidentiality rules. A search within a patient record is open to all users authorized to view that record, whereas a search on the entire database of patient records (multi-record search) is only permitted to members of the medical board of the hospital, or to those granted special dispensation by the hospital management in exceptional situations. To prevent data leaks, exporting of query results (e.g., into Excel format) is only available to physicians of the Department of Medical Information who have been individually authorized.

2.2.3 Functional description

To query a single patient record, the solution returns a list of documents from the computerized patient record in which the selection criteria were found. Some descriptive data are displayed along with the query response: number of selected text documents, document type (consultation letter, pathology report, etc.), and dates of the procedures described in the report (e.g., surgery, consultation, etc.).

To query the database of patient records, the solution returns a list of patients with the documents that match the criteria.

Some descriptive elements regarding the group of patients are displayed as facets: age, living/dead status, and housing department. The user can then navigate through each document in the record of each patient selected. Different operators, filters, and facets are available, and these functionalities are detailed in ►Table 1.

All queries are stored by users, enabling them to easily repeat a previous query. This also ensures the traceability of queries issued.

2.2.4 Logistical aspects of the implementation

The ability of the IT department to act with full autonomy, along with the partnership with SWORD Group, allowed the rapid implementation of this project, which was completed in six months at a low cost (€15,000). The software development process consisted of three steps: gathering requirements and designing the architecture (2 months); development of the tool and writing the documentation (2 months); and deployment, bug corrections, and elaboration of the maintenance process (2 months). Development costs were broken down as follows: €5,000 for compensation of two trainees over 5 months (one for CLB and one for SWORD); €7,000 for a CLB developer to finish the integration project; and €3,000 for the purchase of a new server. The Solr licenses were free.

At the time of launch, demonstration sessions were offered, on a voluntary basis, to the future users of the full-text search tool on single medical records. In parallel, the user manual was distributed by e-mail to all authorized users of electronic medical records (physicians, nurses, and other paramedical professions). For future users of the full-text search of the database of patient records, a demonstration was given at a meeting of the CLB's medical board. The user manual for multi-record search was distributed by e-mail to all members of the CLB's medical board.

When new staff members arrive at the hospital (especially physicians and medical students), training sessions are held regarding the use of the electronic medical records, including the use of the full-text search tool.

2.3 Evaluation of the quality of search results

The quality of query results was assessed by evaluating the sensitivity and positive predictive value of the full-text search tool for the detection of patients eligible for an epidemiological study, ISICA [8]. This study required the selection of patients suffering from breast cancer and diabetes. The gold standard for assessing the sensitivity was the identification of eligible cases by manual search of the computerized records by a clinical research assistant. Working time required for the conventional selection of eligible cases was compared to the time required for selection using the full-text search tool.

3. Results

3.1 Description of types of use

From January to May 2013, 500 queries concerning individual patient records were launched per month by an average of 140 different users, and 210 different expressions were searched per month. Among all healthcare professionals authorized to use the computerized medical records for medical care, 50% of search engine users who queried individual patient records were healthcare professionals other than medical practitioners (e.g., nurses and other paramedical professionals).

In the same period (January to May 2013), 25% of users of medical records who were entitled to make requests on the entire database launched at least one request. On average, 114 multi-record requests are made per month.

3.1.1 Tool for the medical care of the patient

The most common use of the full-text search tool is in consultation or at the bedside. In such cases, a medical practitioner needs additional information regarding the medical care of the patient, and launches an appropriate query of the computerized patient record. The most frequently searched expressions are drug names, different steps involved in medical care (radiotherapy, chemotherapy, metastasis, multidisciplinary consultation meetings, etc.), comorbidities (allergy, diabetes, etc.), and test results (myelogram, KRAS gene, etc.).

For example, a patient was monitored for chronic lymphocytic leukemia for several years at the center. The medical practitioner could not remember whether blood cytogenetics have been performed so he made a request in the computerized patient record, entering the term „karyotype." Instantly, the date and the test results were found in the computerized patient record, avoiding unnecessary new prescriptions and procedures.

Under the Multidisciplinary Consultation Meetings, this tool is particularly useful when discussing a case of orphan disease, a difficult case, or new therapies still rarely used in the situation under consideration. Furthermore, the tool allows quick retrieval of records of patients whose medical care has already been discussed. Earlier decisions traced in the database aid with discussions of the most suitable medical management for that patient.
The tool is also used by the Pharmacy for the validation of drug prescriptions.

3.1.2 Tool for clinical research

This tool also aids with identifying patients eligible for clinical trials, a process far quicker than manual search. For example, suppose that a patient is being considered for pre-selection for a clinical trial for chronic lymphocytic leukemia, in which an exclusion condition is the absence of prior treatment. To verify this exclusion condition, the doctor queries the patient's record by entering the term „chlorambucil". The search results include a document indicating that the patient received this treatment several years ago for 2 months. Therefore, the patient must be excluded from the clinical trial.

Full-text search also facilitates the rapid generation of a database in response to a study protocol (e.g., in a thesis, or in preparation of a scientific publication). For instance, full-text search allowed exhaustive retrieval of patients treated for a very rare clinical entity (NK/T-cell lymphoma) and easy collection of data for a retrospective multicenter study of French institutions [9].

3.1.3 Improving the traceability of medical care in medical records

The Department of Medical Information and medical assistants use the tool to ensure good traceability of medical care in patients' medical records and the pricing of medical procedures.

3.2 Qualitative evaluation

The ISICA study [8] was a case-control study based on data collected in 92 centers that manage breast cancer in France, the United Kingdom and Canada. In the context of this study, the Medical Information Department selected patients diagnosed with breast cancer at the CLB between 1 January 2008 and 30 June 2009, using ICD-10 coding for diagnosis-related groups (DRGs).

A clinical research assistant analyzed each patient's medical records to assess their diabetic status. Of the 2,129 patients diagnosed with breast cancer over this period, 88 patients had diabetes and were selected for inclusion in the study. The screening of the 2,129 files required 304 hours of work (on average, seven cases handled per hour).

In parallel, the screening of 2,129 records was performed using the full-text search tool. Using the combinations „DIABETES“ OR „INSULIN“ or „trade name insulin treatment“ or „trade name treatment with oral antidiabetic agents,“ the full-text search identified 150 patients.

Screening by the clinical research assistant of the 150 preselected patients identified 73 of the 88 patients selected by the traditional method, corresponding to a sensitivity of 83.0% and a positive predictive value (PPV) of 48.7%.

The work time required was 2 hours for the Medical Information Department (conception of the query, programming the query on the server, formatting results) vs. 22 hours for screening by the clinical research assistant. If this method had been used for the study, the gain in work time of the clinical research assistant would have been 92.8%. However, in order to follow the protocol of this multicenter study, this tool has not been chosen for the selection of eligible study subjects.

4. Discussion

Thirty years ago, the CLB Cancer Center made the strategic choice to develop its own computerized patient record system and gradually develop additional tools for its use. This decision allowed the deployment of a full-text search tool to be used directly by medical practitioners in the care of their patients. When patient records were stored on paper, practitioners seeking very specific information about a patient needed to review each sheet, one after another. One of the first advances related to the computerization of medical records was the use of metadata, including the characteristics of each document (surgical report, pathology report, etc.), which limited searches to subsets of medical records. The introduction of a tool for full-text search for direct use within computerized patient records provides an additional method for searching unstructured data within medical documents. The functionalities offered by the full-text search tool have significantly changed the use of computerized patient records by practitioners. Indeed, these functionalities allow health professionals to obtain query results in a few seconds, significantly improving their daily practice as demonstrated in this article.

There are four prerequisites for implementing full-text search of computerized patient records.

- Health information consists of confidential data to which access is regulated, taking into account the rights of the patient. A procedure that describes various aspects of data security and confidentiality (see "Data Security", above) must be implemented in order to enforce these rights.
- It is necessary for patient records to be fully computerized and for health professionals to support this by not using a paper record.
- It is necessary to work with software engineers who have technical mastery of the full-text search tool and search-engine indexing.
- It is necessary for the developer of the computerized patient record to be willing to invest in this area.

Full-text search is particularly efficient at retrieving poorly structured data from medical records, due to its ability to search for words in documents without any constraint on the data structure. Uncommon medical data such as karyotypes, HLA typing, particular histories, and former treatments are examples of data for which the full-text search tool can help to retrieve information very rapidly. The searches are indeed extremely fast (a few seconds at most), saving time and providing medical information that reduces the number of redundant prescriptions and procedures (e.g., biological or radiological examinations that were already prescribed or performed recently), both of which are important parameters for improving the quality of medical care. This empirical observation has been demonstrated in the health information technology literature, which has shown that this health information technology enhances monitoring and surveillance activities, reduces medication errors, and reduces rates of utilization of potentially redundant or inappropriate prescriptions [10].

This project was met with strong approval by the CLB medical community, possibly attributed to the involvement of physicians during the initial "needs gathering" phase at the time of project start-up. Because this type of tool is very new to physicians, they asked that the first implementation phase should focus on simple functionalities such as full-text keyword-based search of medical records. More complex functionalities, such as synonym searches for medical concepts, giving similar concepts (e.g., generic names of drugs), were judged as "nice to have" additional features for the next step, but the initial features were judged to be good enough for the first step. Now that the implementation phase has been completed successfully, users have seen the limits of the basic features initially introduced and now want to refine the search capabilities of the tool. Indeed, the limitations of the current version of the tool represent areas for improvement in the coming years. For example, because the tool is not based on a semantic analysis technique, it is not possible to identify negations [11]. This leads to a significant number of false positives, as shown by the relatively low PPV (48.7%) (See "Results", above). A considerable proportion of the false positives are due to differential diagnoses listed and discussed in medical records. Moreover, in the absence of semantic analysis, no time labeling of medical entities takes place. For example, if the user searches for patients with non-Hodgkin's lymphomas located on the ovary, a large number of false positives (66 of 72 records in the current database) will be selected by the tool, because it cannot discriminate between information about the ovaries relating to the patient's history (including family) and information related to the current management of lymphoma.

Time labeling of extracted medical information is, therefore, an important feature that should be introduced to clarify the relevance of detected entities [12]. Depending on the type of query, the degree of precision of time labeling will be more or less important. The minimum precision for time labeling would permit discrimination between medical information extracted from the „history", „reason for seeking care", and „evolution" sections.

In the short term, to overcome this lack of precision, the developers decided to combine full-text search with the use of the metadata structuring computerized patient records, allowing users to exclude entities identified in the „history" section. Moreover, the information processing required for medical practice seems more sophisticated than that offered by full-text search tools found on the Internet [13]. More complex algorithms remain to be performed [14]. In addition, because of the number and diversity of potential users of health data, implementation of search modules should facilitate development of algorithms tailored to these different types of queries [15].

Development prospects in the coming months are:

- Increasing the number of types of documents on which a search is possible (medical notes during hospitalization stays; end-of-life wishes; resuscitation information; information given to the patient; information provided to third parties; reports on multidisciplinary consultation meetings; orders issued from computerized patient records; certificates issued from computerized patient records).
- Implementing advanced functionality, including NLP (for detection of negations and of documents headers), concept extraction (used to identify medical terms such as names of drugs, diseases, and multidrug-resistant bacteria), use of thesauri to extend searches to synonyms and drug equivalents, and implementation of reasoning rules to categorize the patient profile. Along with other French cancer centers, the CLB is a member of the Consore project, which will provide these functionalities. The text-analysis service developed for Consore will be integrated into the CLB search engine during the first semester of 2014. This service uses an ontology that integrates several international terminologies (ICD-10; the 'morphology' and 'disease' branches of SNOMED; ATC, a list of commercial drugs names; the Cosmic gene list; and CCAM, which contains French terminology for therapeutic procedures) that are used for concept annotation.
- Connecting this advanced functionality to aid in DRG coding, in order to optimize the cost of hospitalizations, including taking better account of comorbidities.
- Narrowing down the rules of access and use of sensitive data to facilitate their use, while maintaining strict confidentiality of patient data.
- Generating automatic alerts from pre-built queries (e.g., sending an alert when a new document mentions a patient with multidrug-resistant bacteria).

5. Conclusion

In France, hospital IT tools that can influence medical practice are uncommon. Indeed, the most frequently deployed tools are based on an old conception (from the 1990s) and have focused more on the needs of hospital institutions than on the requirements of medical activities.

The project described here, consisted in integrating a full-text search engine in the hospital information system. Using an open source search engine Solr which indexes content sources, processes query requests and returns search results in electronic health records, this project demonstrates that the evolution of computer technology in the health sector has the potential to significantly change daily medical practice in patient care.

Clinical relevance statement

The primary purposes of integrating a search engine into electronic health records are to save medical time and facilitate the use of unstructured medical information for various endeavors. This has a direct impact on the medical care, changes the practice of patient management, and reduces the time required to select patients and collect data for research activities.

Conflicts of interest

The search engine integration and configuration was performed by SWORD, under the direction of E. Barthuet. The other authors declare that they have no conflicts of interest in this research.

Protection of human and animal subjects

Human and/or animal subjects were not included in the project.

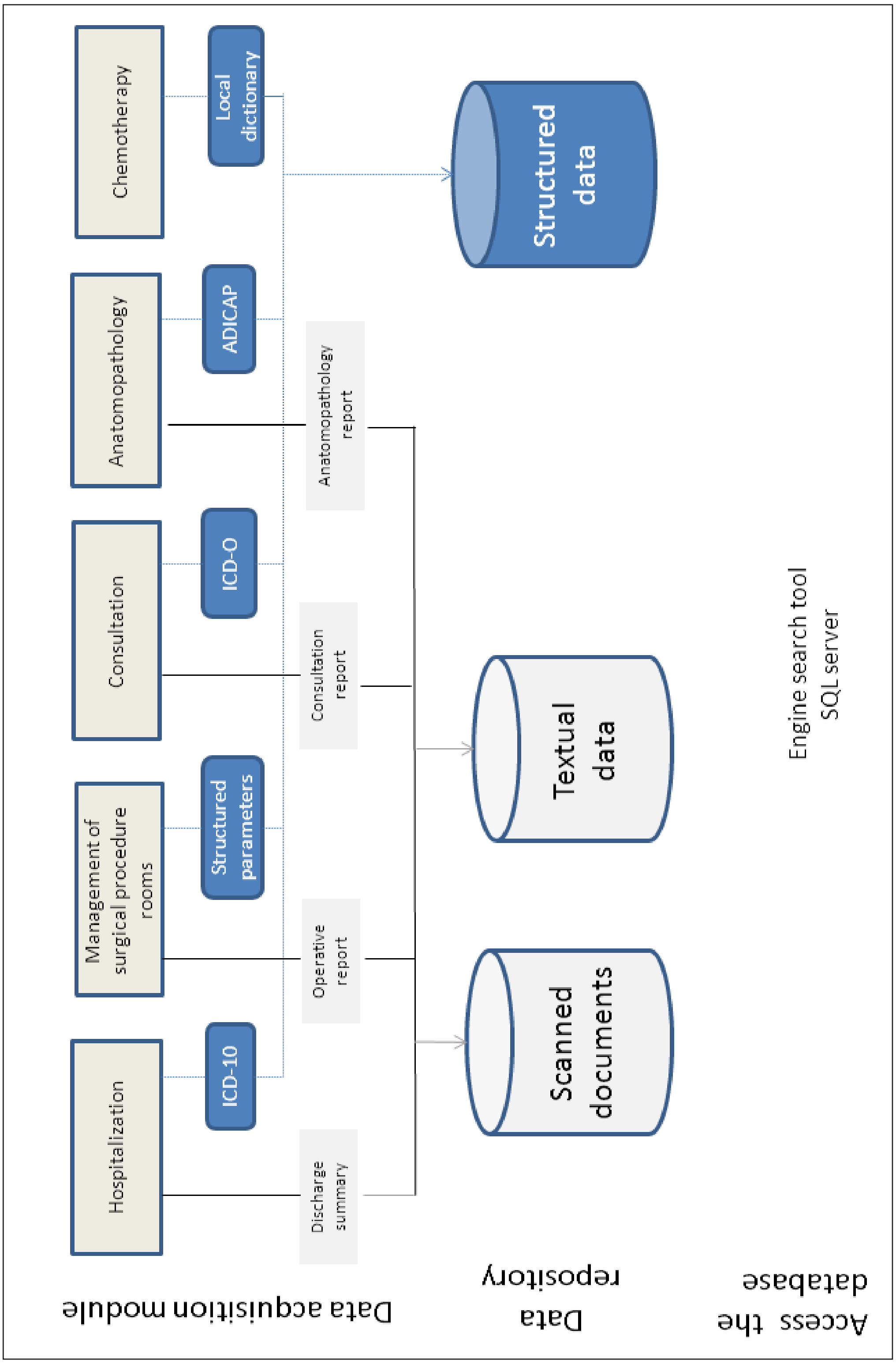

Fig. 1 Hospital Information System (Léon Bérard Cancer Center, Lyon, France)

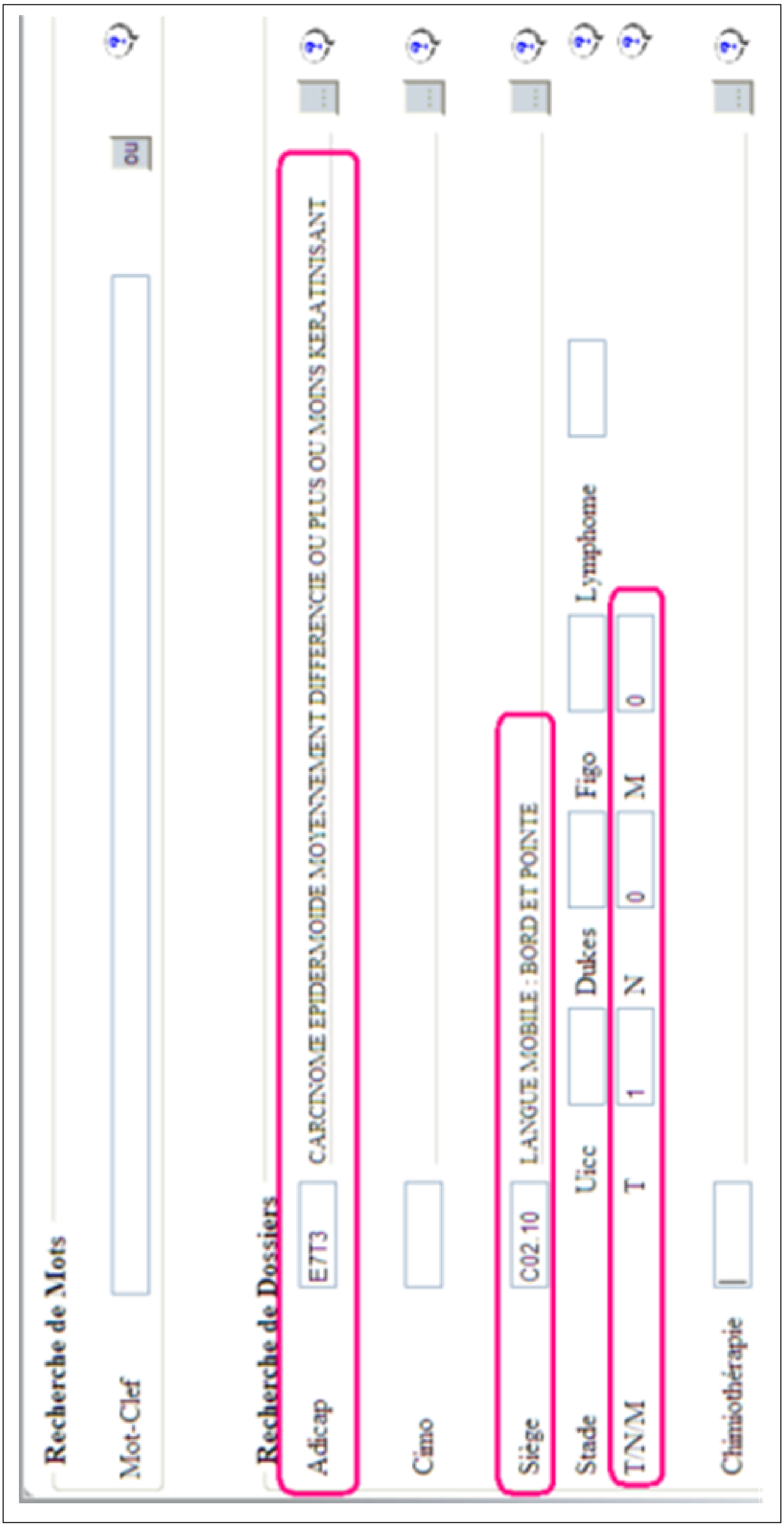

Fig. 2 Example of a query on the web form (multi-record search)

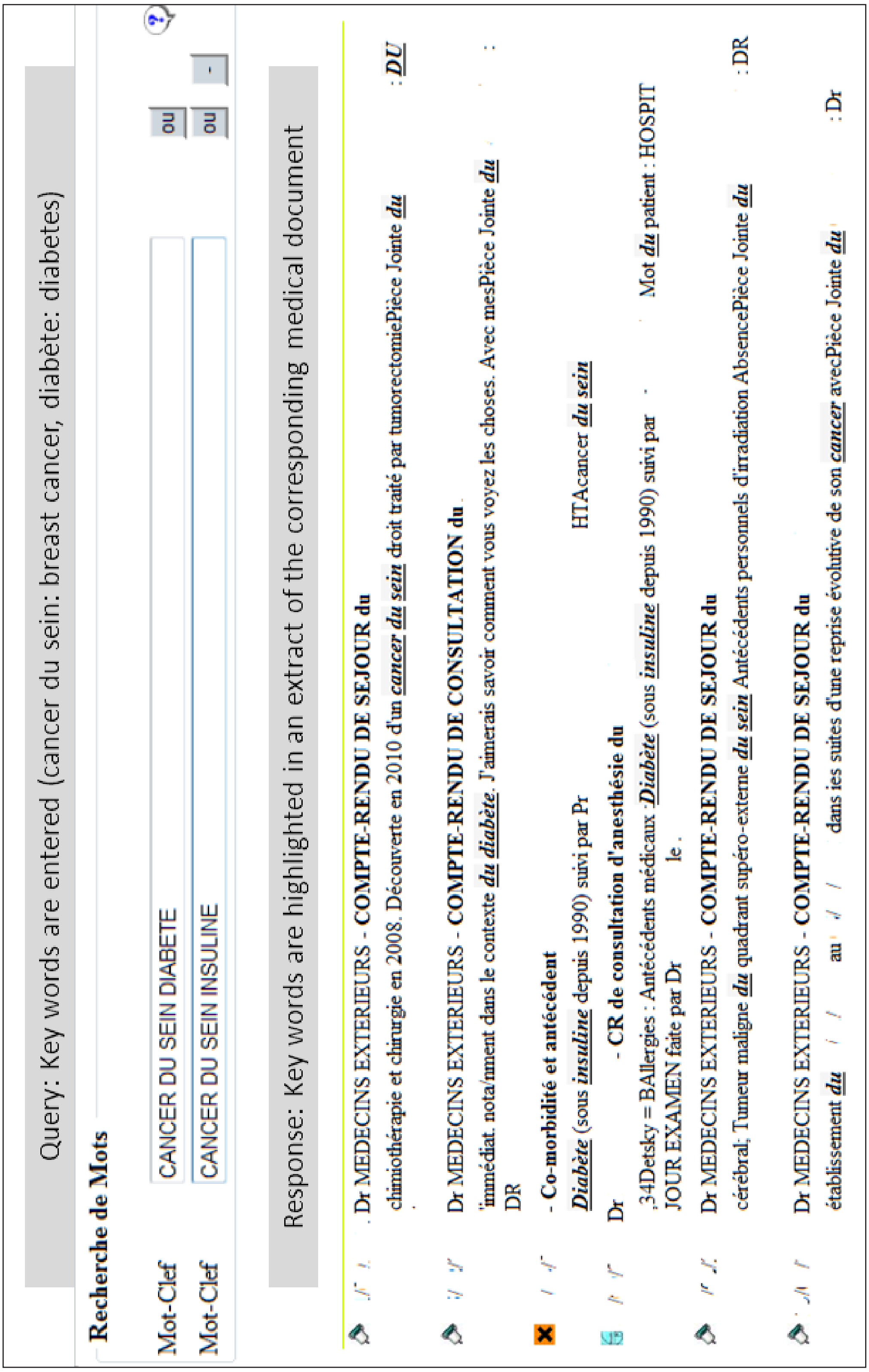

Fig. 3 Example of a query (keywords) on the web form (multi-record search)

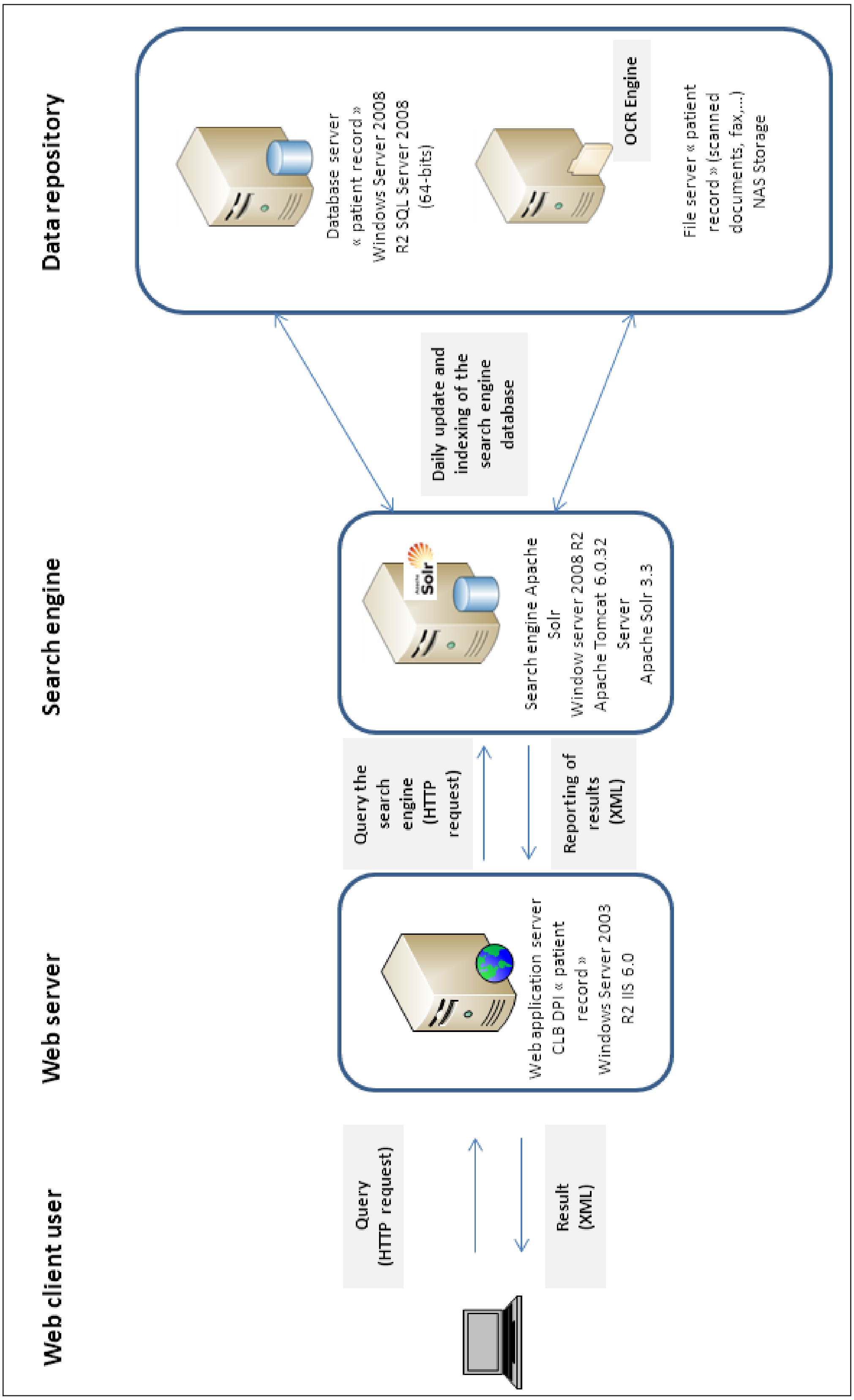

Fig. 4 Data repository search engine and user interface (Léon Bérard Cancer Center, Lyon, France)

Table 1 Available operators, filters, and facets for queries of the medical record database

Type of functionality	Description
Operators	Search records containing the expression “Word 1”
	Search records containing the expression “Word 1” and the expression“Word 2”
	Search records containing the expression „Word 1“ or the expression “Word 2”
	Search records containing an exact expression, “Word 1 Word 2”
Filters	Date of completion of the event
	Pathological code
	ICD-O
	Stage T N M
	Type of chemotherapy
Facets	Time between the effective date of the event and the date of the report
	Age of the patient at the time of report
	Age of patient (when alive)
	Living status: Alive / Dead
	Sex
	Department: Department of Housing patient
	Type of text document (consultation report, pathology report, etc.)
	Doctor who signed the report

References

1. Gregg W, Jirjis J, Lorenz iN, Giuse D. StarTracker: an integrated, web-based clinical search engine. AMIA Annu Symp Proc 2003: 855.
2. Hanauer DA. EMERSE: the Electronic Medical Record Search Engine. AMIA Annu Symp Proc 2006:941.
3. Erinjeri J, Picus D, Prior F, Rubin D, Koppel P. Development of a Google-Based Search Engine for Data Mining Radiology Reports. Journal of Digital Imaging 2009; 22(4): 348–356.
4. Thiessard F, Mougin F, Diallo G, Jouhet V, Cossin S, Garcelon N, Campillo B, Jouini W, Grosjean J, Massari P, Griffon N, Dupuch M, Tayalati F, Dugas E, Balvet A, Grabar N, Pereira S, Frandji B, Darmoni S, Cuggia M. RAVEL: retrieval and visualization in electronic health records. Stud Health Technol Inform 2012; 180: 194–198.
5. Seyfried L, Hanauer DA, Nease D, Albeiruti R, Kavanagh J, Kales HC. Enhanced identification of eligibility for depression research using an electronic medical record search engine. Int J Med Inform 2009; 78(12): e13–e18.
6. Zheng K, Mei Q, Hanauer DA. Collaborative search in electronic health records. J Am Med Inform Assoc 2011; 18(3): 282–291.
7. Association pour le Développement de l'Informatique en Anatomie et Cytologie Pathologiques. Thésaurus de la codification ADICAP. 2009.
8. Grimaldi-Bensouda L, Marty M, Pollack M, Cameron D, Riddle M, Et-al. The International Study of Insulin and Cancer. Lancet 2010; 376: 769–770.
9. Chauchet A, Michallet A, Berger F, Bedgedjian I, Deconinck E, Sebban C, Antal D, Orfeuvre H, Corront B, Petrella T, Hacini M, Bouteloup M, Salles G, Coiffier B. Complete remission after first-line radio-chemotherapy as predictor of survival in extranodal NK/T cell lymphoma. J Hematol Oncol 2012; 5: 27.
10. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, Morton SC, Shekelle PG. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006; 144(10): 742–752.
11. Hagège C, Marchal P, Gicquel Q, Darmoni S, Pereira S, Metzger M. Linguistic and Temporal Processing for Discovering Hospital Acquired Infection from Patient Records. In: David Riaño AtT, Silvia Miksch, Mor Peleg, ed. Knowledge Representation for Health-Care – ECAI 2010 Workshop KR4HC 2010, Lisbon, Portugal, August 17, 2010, Revised Selected Papers Lectures Notes in Artificial Intelligence. Vol. 6512 Springer, 2010; 70–84.
12. Proux D, Hagège C, Gicquel Q, Kergourlay I, Pereira S, Rondeau G, Darmoni S, Segond F, Metzger M. ALADIN : Développement d'un outil sémantique d'analyse des documents textuels médicaux pour la détection d'infections associées aux soins IRBM, Ingenierie et Recherche BioMedicale 2012;33(2):137–142.
13. Giustini D. How Google is changing medicine. BMJ 2005; 331: 1487.
14. Yang L, Mei Q, Zheng K, Hanauer DA. Query log analysis for an electronic health record search engine AMIA Annu Symp Proc 2011: 915–924.
15. Natarajan K, Stein D, Jain S, Elhadad N. An analysis of clinical queries in an electronic health record search utility. Int J Med Inform 2010; 79(7): 515–522.