Application of Information Technology ■

A Web Terminology Server Using UMLS for the Description of Medical Procedures

ANITA BURGUN, MD, PHD, PATRICK DENIER, MD, PHD,
OLIVIER BODENREIDER, MD, PHD, GENEVIÈVE BOTTI, MD,
DENIS DELAMARRE, BRUNO POULIQUEN, PHILIPPE OBERLIN, MD,
JEAN M. LÉVÉQUE, MD, BERTRAND LUKACS, MD, FRANÇOIS KOHLER, MD, PHD,
MARIUS FIESCHI, MD, PHD, PIERRE LE BEUX, MD, PHD

A b s t r a c t   The Model for Assistance in the Orientation of a User within Coding Systems (MAOUSSC) project has been designed to provide a representation for medical and surgical procedures that allows several applications to be developed from several viewpoints. It is based on a conceptual model, a controlled set of terms, and Web server development. The design includes the UMLS knowledge sources associated with additional knowledge about medico-surgical procedures. The model was implemented using a relational database. The authors developed a complete interface for the Web presentation, with the intermediary layer being written in PERL. The server has been used for the representation of medico-surgical procedures that occur in the discharge summaries of the national survey of hospital activities that is performed by the French Health Statistics Agency in order to produce inpatient profiles. The authors describe the current status of the MAOUSSC server and discuss their interest in using such a server to assist in the coordination of terminology tasks and in the sharing of controlled terminologies.

■ **J Am Med Inform Assoc.** 1997;4:356–363.

In France, a national survey of hospital activities is performed every 10 years by the governmental Health Statistics Agency (SESI) in order to produce inpatient profile statistics concerning patient demographics and clinical information. A representative sample of inpatient coded discharge summaries from public and

private hospitals is collected. Over 80,000 records were collected from 423 hospitals during the last survey (1993). The medical information was coded using the French Catalogue des Actes Médicaux (CdAM) for the medico-surgical procedures. Information about medico-surgical procedures is needed by several participants in the health care delivery system, including clinicians, hospital managers, health policy makers, and clinical researchers. Each of them has a specific viewpoint on the information. The CDAM is a catalog; that means that data analysis based, for example, on the anatomic sites on which the procedures are focused, the types of surgical approaches, or the nature of the actions, requires construction of specific computer programs due to absence of a more sophisticated structure within the nomenclature. A representation based on the semantics of the procedure labels is a fundamental requirement in this context.[1]

In order to be usable by several participants in the health care system, information dealing with medical procedures must be sharable. By sharable, we mean

two things: the first one deals with the conceptual model, the second one with the availability of the data. The model must include existing concepts and a standard representation format. Therefore, we have explored the possibility of reusing the UMLS knowledge-base,[2,3] and we have integrated the recommendations of the European prestandard defined by the CEN/TC251 workgroup.[4] The availability of controlled terminologies on wide-area networks allows a collaborative development of the terminologies, makes possible their updating by several users, and facilitates their evaluation. A terminology system that includes network services can allow multiple applications built upon the unique database and can serve a wide range of health-care needs.[5,6] Therefore, we have been developing since 1995 a terminology server for medico-surgical procedures in a Web server-accessible form, available to any workstation on the Internet.

There have been two phases in the development of the MAOUSSC (Model for Assistance in the Orientation of a User within Coding Systems, in French Modèle d'Aide à l'Orientation d'un Utilisateur au Sein de Systèmes de Codage) application. The first phase occurred before the development of the Web server. The system in use was a 4D prototype running on the Macintosh.[7] There were as many databases as medical domains. Each database was initially filled with a relevant list of procedure codes and labels and included a subset of the UMLS Metathesaurus[8] captured by an automated process. The second phase is the one described here which implements the Web server.

Methods

Our approach is based on a conceptual model, a controlled terminology, and Web server distribution.

A Conceptual Model

Medical terminology requires a deeper representation than the traditional tree-structured hierarchy, which is insufficient to precisely define what a term is and how it differs from the others. The aim of the MAOUSSC project is to provide a conceptual model for medical and surgical procedures. It is expected to have the independence and the flexibility that will accommodate several points of view.

A medical procedure may be more or less complex. There are four levels of complexity. The lower level corresponds to MAOUSSC procedures that are described by a multi-axial model. The basic units of information which describe procedures have been categorized in eight axes. The mandatory axes conform with the CEN TC251 recommendation.[4] Four of them are mandatory:

- Nature; i.e., what is done (ectomy, insertion, etc.)
- Topography; i.e., the anatomic site the action focuses on
- Instrumentation
- Approach

Four complementary dimensions must be added to the description, depending on the nature axis value:

- An additional topography (i.e., axis topography complement) is required for procedures such as anastomosis.
- The matter or device (i.e., matter/device) involved must be specified as the direct object of ectomy or insertion verbs.
- A body process (i.e., BodyProcess) is mentioned when the action focuses on a biologic function instead of an organ.
- The disease axis is instantiated only for expressions like *treatment of female genital prolapse.*

The second level consists of MAOUSSC procedures which may be qualified by modifiers; these are called elementary procedures. Modifiers can be used to indicate that a procedure is altered by some specific circumstance but is not changed in its definition. Modifiers are either general modifiers (e.g., "operation on child") or modifiers for billing, such as information about when the procedure is performed (week-end, etc.). Whereas an axis is instantiated with a single value, several modifiers may be included for one procedure.

The third level, associations, allows representation of several elementary procedures that are performed during a single operative session. The conjunction "AND" is used to connect the elementary procedures. Associated procedure terms may be combined into more complex sentences with the "OR" connective. Such sentences refer to composed procedures that correspond to the upper level in existing nomenclatures, in particular in the CDAM. It is obvious that a sentence like "A or B" is ambiguous, since we do not know what procedure was actually performed. Whereas ambiguous terms must be absent from a well-built terminology, the objective for the SESI study was to represent existing CDAM codes. Thus we had to take into account the presence of terms like "A or B" in the CDAM.

In the example given in Figure 1, we have represented the label "graph of tendon or muscle (one site) with vascular anastomosis, forearm." With such a sentence,
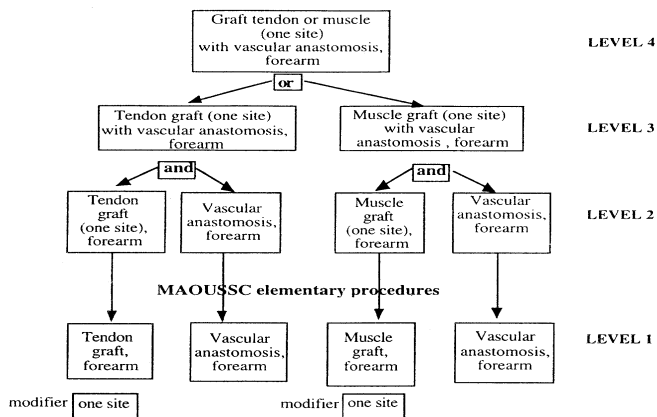
Figure 1 Example of decomposition.

it is not exactly clear what procedure has in fact been applied. The label is qualified as ambiguous. It is a level four term decomposed into two associations:

(A) Graft of tendon (one site) with vascular anastomosis, forearm
(B) Graft of muscle (one site) with vascular anastomosis, forearm

Each association is made of several procedures combined as part of one overall procedure. Association (A) is itself the association of two elementary procedures, which are as follows:

(A1) Graft of tendon (one site) forearm
(A2) Vascular anastomosis, forearm

Associated term (A1) is a MAOUSSC procedure with a multiaxial representation—the Nature axis being instantiated with graft—modified by the modifier "one site." Associated term (A2) is another MAOUSSC procedure with the Nature axis being instantiated with anastomosis (generic SHUNT) and the Topography being the vessels of the forearm.

Constraints Associated with a Controlled Set of Terms

We have explored the possibility of reusing existing terminologies (the UMLS knowledge base[8] and SNOMED[9]) as a foundation for the MAOUSSC base of concepts. The MAOUSSC version in use is associated with the fifth release of the UMLS knowledge sources which includes some of the SNOMED International terminology. Each concept in the MAOUSSC knowledge base is assigned to one or more clinical

domains (e.g., cardiology, urology, etc.). A method has been developed in order to create an automated partition of the UMLS Metathesaurus into clinical domains.[10] For each medical domain, the first step consists of selection of a few terms, specific for the medical domain in question, from one of the nomenclatures incorporated in the UMLS, like the Anatomy terms of SNOMED. The initial set of terms is then extended by a computerized process that explores the interconcept links in the UMLS Metathesaurus. The result is a set of UMLS concepts that are assigned to the medical domain in question. This partition is used to constitute the initial lists of terms of the domain that are proposed to the user as possible values for the MAOUSSC axes. Additional concepts can be created or assigned to other domains by the user. Thus, a concept in the MAOUSSC application may be an exclusive MAOUSSC concept or a Metathesaurus concept.

Each concept is assigned to one or more (up to several) semantic types. The MAOUSSC conceptual network and the semantic categorization of the concepts are based on the UMLS Metathesaurus and Semantic Network. For instance, the axis Instrumentation is associated with the Medical Device type. Topography is associated with Body Part, and/or Body Space and some other UMLS Semantic Types. We had to create a specific type for procedure description for the surgical or investigation "approach." We also needed to split up the UMLS Semantic Type Medical Device in order to distinguish devices that may be left inside the body (prosthesis for example) and belong to the Matter/Device axis from Instrumentation such as endoscopes.

The semantic categorization of the terms is coupled with syntactic patterns in order to guide and control the description of MAOUSSC procedures. Each action term belongs to a class corresponding to a generic action. The model uses a limited set of generic actions in procedures, which are:

- Amputation
- Delivery (for obstetrics)
- Destruction
- Endo-exo: taking something out of the body (e.g., ectomy)
- Graft
- Incision
- Insertion
- Liberation
- Measurement of process

- Mobilization
- Modification of process
- Pexy (fixation)
- Reconstruction
- Repair
- Shunt
- Transfer
- Treatment
- Visualization (scopy or imagery).

A pattern is designed for each action class in order to restrict the set of complementary axes for which values must be supplied. Given this set of axes and the relevant semantic types for each axis, the description of MAOUSSC basic procedures becomes constrained. For example, for the generic action "Incision" the allowed semantic axes are the mandatory axes: Topography, Instrumentation, Approach, and the non-mandatory axes: Complement of topography, and Matter/Device. Furthermore, for a given action pattern, each axis is associated with UMLS semantic network types. A second level of the semantic constraints is to allow more specific actions for a given generic action. For instance, the generic action Repair corresponds to the following action concepts: remodeling plasty, dilatation, suture, raphy. They will inherit the semantic constraints of their generic action. Figure 2 is an example of the constraints for the action Modification of Process.

Web Server Development

In 1995 we began to develop an application with the following set of specifications: it had to support the MAOUSSC model; information had to be accessed from several sites and multiple client platforms; a given medical procedure could be described by more than one user, so there would be several instantiations of descriptions; a user-friendly interface was required; and the information stored in the central system would consist of inpatient discharge summaries (at least the 80,000 records of the SESI survey) and descriptions of medical procedures from several nomenclatures and terminologies using a semantic network model.

The Web offers a solution with its multiplatform clients. It provides access to a database for a variety of users and therefore supports multiple applications and facilitates the evaluation of the terminology. Accessible on the Internet, the terminology becomes available for anyone involved in the health-care system.

We tried to divide our application into the three usual layers: data management, treatments, user interface. For the data management, we chose a Relational DataBase Management System (ORACLE) to provide a secure and maintainable multi-user data server. For the user interface, we chose the World Wide Web instead of using a fourth generation language such as SQLFORMS. Our database contains essentially text information, so we chose a language which allows easy text manipulation: PERL (Practical Extraction and Report Language). This language has an extension ORAPERL which facilitates access to an ORACLE database. The intermediary layer manages forms for the creation and updating of the database items.

We already had good experience in using the Web and for several years had relied on the performance of PERL,[11] so we knew that it was possible to build an interface that searches an existing database. The challenge for the MAOUSSC project was to build a program that allows interactive database updating. Fortunately, new versions of HyperText Markup Language (HTML), which is the Web standard format, provide a useful form interface. The aim was therefore to create these HTML forms dynamically.

Because the client and the server are not connected permanently (the main innovation of the Web system), it is difficult to preserve the context of each user. We therefore had to simulate a real session. We chose to build only one PERL script to manage all the forms presented to the users, instead of writing as many scripts as there are forms (creation/update of concepts, of MAOUSSC procedures, etc.).

Generic action : MODIFICATION OF PROCESS		
Axes	*Mandatory Yes/No*	*Allowed Semantic Types*
Topography	Yes	Body Part, Organ or Organ Component Body Location or Region Embryonic Structure Body Space or Junction
Instrumentation	Yes	Medical Device
Approach (Voie d'Abord)	Yes	Approach (Voie d'Abord)
Topography (complement)	No	Body location or Region
Biological Process	Yes	Biologic Function Physiologic Function Organism Function Organ or Tissue Function Cell Function

Terms of generic action
Modification of Process
Interruption
Augmentation
Diminution
Trigger

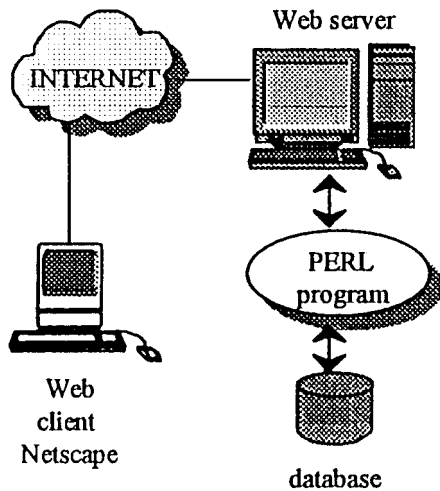**F i g u r e 2** An example of semantic constraints.

Figure 3 Accessing database via the Web.

The main advantage of our technical choices is platform independence. There are versions of ORACLE, PERL, and the Web (client and server) for various platforms. The next upgrade of PERL will include DBPERL, which allows the program to be independent of the database management system. Our PERL script can now be used to update any other database via the Web (Fig. 3).

Use of the Model on the Webserver

The system has been used to describe the existing procedure wordings in order to decompose and describe the semantics of medico-surgical procedures. We have tested our model on the French CDAM.

The following steps have to be taken for a given clinical field:

- Extract the relevant UMLS subset for a given domain using the method described in.[10]

- Identify all the wordings and terms used in a given procedure clinical section.

- Search for all the terms in the UMLS (either in French or English). If the English term exists, a French "synonym" may be created if it does not exist in the Metathesaurus.

- If the concept does not exist, then it must be created, and it must be linked to an existing semantic type of UMLS.

- Then the next step is to create MAOUSSC elementary procedures. In each axis the only terms that can be selected are those which are compatible with the UMLS semantic type which has been associated with the semantic grammar. This way we can en-

force the use of a controlled terminology set for describing a medico-surgical procedure.

- Create the whole procedure as a logical combination of MAOUSSC procedures using the connectives AND, OR.

The MAOUSSC System

Besides development of the model, we had as an objective to show that on-line cooperative work could be productive for creating and maintaining a terminology server using Web technologies on the Internet. The target users are medical informatics specialists and statisticians in cooperation with the expert clinicians of a given medical domain.

A first version of the MAOUSSC system is currently in use. The development time was mainly used for the software behind the HTML forms. As stated previously, programming for the World Wide Web requires professional computer scientists, especially for database updates, because the server loses the context between client requests. But with the powerful text, file, and database manipulation of PERL, it took 6 man-months to create the complete MAOUSSC server prototype.

We have been using the server for several months either internally or externally through a cooperative network of several Medical Informatics Laboratories in France. The prototype was validated by the users. The description of the CDAM procedures involved two users for each medical domain. The 17 subsections of the CDAM were explored. The description of

Figure 4 Looking at the concepts in UMLS browser.

the CDAM procedures for the SESI survey was made by physician experts from the medical domain, from several cities in France, assisted by Medical Informatics specialists in four university hospitals.

The terminology server has facilitated cooperative terminology development and use. There is less need for correction of mistakes or misunderstandings and consequently fewer meetings for coordinating the production work. The 80,000 discharge summaries collected by the SESI are stored in the same single database with the terminology. With the MAOUSSC terminology server accessible on the Internet, the French Health Statistics Agency can exploit the database with immediate access to information and has a user-friendly Web interface.

The application currently runs on the Web server in the Laboratory of Medical Informatics of the University of Rennes I in France. It is accessible from any microcomputer PC, Macintosh or UNIX station using popular Web clients such as MOSAIC or Netscape.

The decomposition of procedures from several source vocabularies by multiple users had to be coordinated and maintained in a unique data base. The main functions of the procedure terminology server are the creation and retrieval of elementary MAOUSSC procedures, the retrieval and decomposition of procedure nomenclatures, searching and browsing through the UMLS knowledge base, creation of new concepts that must be linked to UMLS semantic types, and administration of the model and database. These functions

Figure 6 Example of creation of a MAOUSSC procedure.

are shown in Figures 4 through 7.

The example given below is composed of four typical screens: UMLS browsing, searching for terms and concepts (Fig. 4); a screen to create new concepts, new terms, or synonyms with links to the UMLS semantic network (Fig. 5); a screen to create MAOUSSC procedure (Fig. 6); and, finally, a screen to describe nomenclature procedures (Fig. 7) by combining the existing MAOUSSC procedures.

The example is the surgical procedure concerning the wrist joint: "capsulectomy and arthrolysis of wrist joint." The first phase is to look in the UMLS browser for concepts related to Wrist Joint ("Articulation Poignet" in French). By looking at "Wrist Joint" we find that the concept is known as a French translation in the domain of musculoskeletal system, and it can be used in the semantic axes "Topography" and "Matter/Device" (in French, Matériel Mobilisé). The UMLS semantic types are also indicated. The concept is known and we can look at its semantic description. The next step is to create the MAOUSSC procedures. In this case, two MAOUSSC procedures are necessary: capsulectomy of wrist joint and arthrolysis of wrist joint. Figure 6 gives the first MAOUSSC procedure: capsulectomy of wrist joint. The different axes are completed by using concepts according to the semantic types. Then the last step is to create the complete nomenclature procedure using two MAOUSSC associated components: capsulectomy of wrist joint and arthrolysis of wrist joint. The previous decomposition gives the user the possibility of using the same elementary MAOUSSC procedure for different nomenclatures or even several languages.
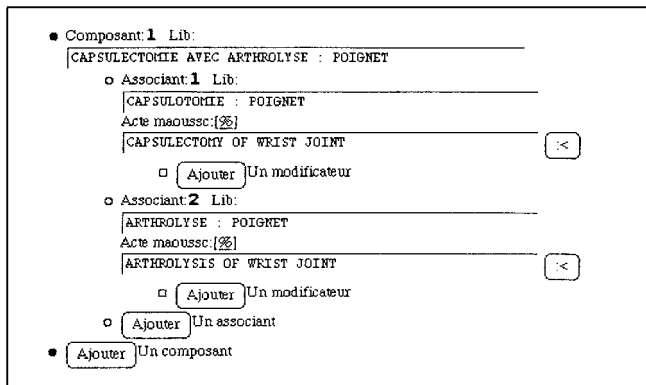

Figure 5 The concept Wrist Joint is known in English and in French.

Figure 7 Decomposition of nomenclature procedure.

Discussion

Although we share some goals with GALEN,[12] our initial approach is different. The design of the MAOUSSC server is driven by concrete application needs. The service is developed by a network of physicians, medical informatics researchers, and public health researchers. We have focused on medical procedures because there is a demand for a sharable terminology that adequately covers clinical content, serves public health needs, and is suitable for health insurance. An additional difference is our choice about the reuse of the UMLS knowledge base. When we started the MAOUSSC project in 1994, GALEN IN USE was not operational for production users, whereas the UMLS thesaurus was available for computerized research developments. The fact that the UMLS knowledge base contained much of the SNOMED International terminology was an additional argument for experimenting with it.

As mentioned in a previous paper,[13] the main limitation of the MAOUSSC application is still the time required for the description of the medical procedures. We had 2,985 distinct procedures to be described in order to cover the set of the 80,000 inpatient records of the national survey of hospital activities. The description results from joint work by both medical information specialists and physicians. We estimate that about 100 days of work were required for the whole set of terms. This time-related limitation is an incentive for us to perform a more automated description of medical procedures. We are currently working to make the UMLS interconcept links operational for a more automated description with respect to the syntactic and semantic constraints.

A major problem is the granularity of procedure descriptions (or, more generally, of any medical description). Therefore it is difficult to define the level at which the concepts had to be represented. Of course, neither our procedure terminology nor the UMLS Metathesaurus in each clinical field had the same level of granularity description, and the matching of both created some difficulties. The notion of semantic distance between terms is also an open research topic in medical linguistics and terminologies.[14]

We now have the conviction that the use of a Web server offers some clear advantages. Four points must be emphasized. The first point is related to the lack of completeness of the domain-related set of terms captured from the UMLS. To compensate for the lack of completeness, the former 4D application user created concepts without previously searching for them in the UMLS. Several users might create several terms with the same meaning. The result was redundancy. Two mechanisms in the current MAOUSSC application can help prevent creation of multiple concept records for the same concept. First, the whole terminology is online. Therefore, the concept C that exists in the Metathesaurus and that would not have been assigned to the domain D can be added to D: instead of creating a concept, the user just adds a link between two existing data. Second, sophisticated tools for term retrieval and similarity detection can be developed on such platforms. Lexical tools for looking at strings and terms have already been implemented.

The second problem deals with evaluation. In order to validate the MAOUSSC model, it is desirable for several medical experts to describe the same procedure terms and then allow the descriptions to be matched in order to measure the agreement between the experts. As there were as many 4D databases as experts, such a task was very tedious. But with the current application, we can manage several instantiations of a description based on a unique thesaurus of concepts. Furthermore, we can have automated matching processes.

The third point deals with the application design. The Web simplifies many aspects of programming. Especially, it frees the developer from worrying about the user front-end platform. Web client programs are freely available for research use. The ergonomy of Web applications using Hypertext links is quite natural for users to browse through large sets of terms and objects.

The last point involves the cooperative work. The MAOUSSC terminology has been built collaboratively with advantages for both developers and users. Some difficulties arose as expert clinicians attempted to understand the semantic model and the links with the UMLS Metathesaurus and the semantic network. For

a given medical domain, the ideal is a team of one medical informatics specialist and one expert clinician, who may become autonomous after a few days of training. Each new clinical field gave us new concepts and semantic links to create, and we feel now that for procedure terminologies our set of terms is rather comprehensive and sufficient.

Conclusion

The development of Web servers in the medical field is now quite extensive and attractive.[15] The concept of a medical terminology server is also recognized as a necessary step for building controlled vocabularies.[16,17] Although MAOUSSC has been used for decomposition of an existing catalogue of medical procedures, we feel that the system and model could also be used to build from scratch a new, coherent procedure nomenclature.

The same concept of terminology server can be similarly used either on the Internet or an intranet. Following the Health Statistics Agency, several authorized centers involved in the health system could obtain information from terminology services. We are now confident that, for other medical terminology that involves extensive cooperative work on a large geographic basis, the same approach and technologies could be applied.

References ■

1. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Inform Assoc. 1994; 1:35–50.
2. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth Inform Med. 1993;32:281–91.
3. Musen MA. Dimensions of knowledge sharing and reuse. Comput Biomed Res. 1992;25:435–67.
4. European Committee for Standardization. Terminology and coding systems for medical procedures. Part I: Surgical procedures. European Committee for Standardization, TC251, 1993.
5. McCray AT, Scherrer JR, Safran C, Chute CG. Concepts, knowledge and language in health-care information systems. Meth Inform Med. 1995;34:1–4.
6. Gennari JH, Oliver DE, Pratt W, Rice J, Musen MA. A Web-based architecture for a medical vocabulary server. Proc AMIA Annu Fall Symp. Philadelphia: Hanley & Belfus, 1995;275–9.
7. Burgun A, Botti G, Lukacs B, et al. A system that facilitates the orientation within procedure nomenclatures through a semantic approach. Medical Informatics. 1994;4:297–310.
8. National Library of Medicine. UMLS Knowledge Sources: Fifth Experimental Edition, Bethesda, MD: NLM, 1995.
9. Cote RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L. The Systematized Nomenclature of Human and Veterinary Medicine. SNOMED International. College of American Pathologists, 1993.
10. Burgun A, Delamarre D, et al. Designing a sub-set of the UMLS Knowledge Base applied to a clinical domain: methods and evaluation. Proc AMIA Annu Fall Symp. Philadelphia: Hanley & Belfus, 1994;968.
11. Pouliquen B, Riou C, Denier P, Fresnel A, Delamarre D, Le Beux P. Using World Wide Web multimedia in medicine. In: Proceedings MEDINFO'95, IMIA. Greenes, Peterson, Protti (eds.). 1995;1519–23.
12. Rector AL, Salomon WD, Nowlan WA, Rush TW, Zanstra PE, Claassen WMA. A medical terminology server for medical language and medical information systems. Meth Inform Med. 1995;34:147–57.
13. Burgun A, Botti G, Bodenreider O, et al. Methodology for using the UMLS as a background knowledge for the description of surgical procedures, Int J Biomed Comput. 1996; 43:189–202.
14. Petiot D, Burgun A, Le Beux P. Modelisation of a Criterion of Proximity: Application to Medical Thesauri. Medical Informatics Europe '96 Proceedings, 1996;1:149–52.
15. Hersh WR, Brown KE, Donohoe LC, Campbell EM, Horacek AE. CliniWeb: managing clinical information on the World Wide Web. J Am Med Inform Assoc. 1996;4:273–80.
16. Humphreys BL, Hole WT, McCray AT, Fitzmaurice JM. Planned NLM/AHCPR large-scale vocabulary test: using UMLS technology to determine the extent to which controlled vocabularies cover terminology needed for health care and public health. J Am Med Inform Assoc., 1996;4: 281–7.
17. McCray AT, Razi AM, Bangalore AK, Browne AC, Stravi PZ. The UMLS Knowledge Source Server: a versatile Internet based research tool. Proc AMIA Annu Fall Symp. Philadelphia: Hanley & Belfus, 1996;164–8.