INFORMATION RESOURCES

Charles P. Friedman, PhD, Associate Editor

Natural Language Processing and Its Future in Medicine

Carol Friedman, PhD, and George Hripcsak, MD

Abstract: If accurate clinical information were available electronically, automated applications could be developed to use this information to improve patient care and lower costs. However, to be fully retrievable, clinical information must be structured or coded. Many online patient reports are not coded, but are recorded in natural-language text that cannot be reliably accessed. Natural language processing (NLP) can solve this problem by extracting and structuring text-based clinical information, making clinical data available for use. NLP systems are quite difficult to develop, as they require substantial amounts of knowledge, but progress has definitely been made. Some NLP systems have been developed and tested and have demonstrated promising performance in practical clinical applications; some of these systems have already been deployed. The authors provide background information about NLP, briefly describe some of the systems that have been recently developed, and discuss the future of NLP in medicine.

Acad. Med. 1999; 74:890–895.

Many believe that the widespread implementation of computerized patient records will bring about a revolution in health care[1] because it will enable the development of technology that improves quality and lowers costs. The electronic medical record (EMR) does provide health care professionals with better access to patient records; however, to have a sweeping impact and to improve care, automated applications must be able to actively manage the vast clinical information resources. Early attempts to use computers to improve patient care included the computerization of applications that facilitated diagnosis and treatment[2,3] and the computerization of guidelines and alerts that facilitated patient management.[4–6] These applications were and still are substantially limited because they require reliable access to clinical data. Other important automated clinical applications (e.g., outcome studies, quality assurance, resource management, and clinical research) also have the potential to substantially affect health care, but these likewise require reliable access to data.

Such access is possible mainly with structured data because textual information is too varied to be easily retrievable. However, coded data are not widely available. Textual reports of patient encounters are a vast source of clinical information, but even when such reports are online the valuable information they contain remains locked within the text. This is because people have very diverse ways of writing notes and because the meanings of words vary depending on their context. Consider the different meanings of the following expressions that contain the word *pneumonia*: *evidence of pneumonia, pneumonia cannot be excluded, rule out pneumonia, pneumonia is not appreciated, pneumonia in 1985.* If a clinician wanted to know how many active pneumonia cases were on record and if he or she did a search using the keyword *pneumonia*, clearly too many reports would be retrieved for patients who did not have pneumonia.

Natural language processing (NLP) systems potentially offer a solution, because they not only extract individual words but also represent well-defined relations among words. If the clinician in the example above searched for *pneumonia* in an NLP system, the system would extract the main finding *pneumonia* for each expression and would also include the appropriate modifier for each record. Modifiers and their values could subsequently be accessed to determine whether or not a patient actually had pneumonia. A *certainty*-type modifier that has the value "no" means that pneumonia was ruled out, whereas a certainty modifier with the value "evidence for" would mean pneumonia was a possible diagnosis. Similarly, a date modifier with the value "1985" would mean that the episode of pneumonia occurred in that year.

Although natural language is easy for humans to understand, it is difficult for computers to comprehend. Natural language embodies an enormous amount of expressiveness, variety, ambiguity, and vagueness. For example, the same concept may be expressed in different ways (congestive heart failure, heart failure, CHF); the same word may have different meanings in different contexts (*discharge* from hospital vs. *discharge* from wound); relations among words may be ambiguous (*no acute infiltrate* may mean there is no infiltrate or that if there is an infiltrate it is not acute); the same concept may have different meanings depending on where it is found (*pneumonia* in the clinical information section of a chest radiograph may mean *rule out pneumonia*); and the meanings of certain expressions lack specificity (*may represent* pneumonia). To enable accurate access, NLP systems must encode the information using terminology from a well-defined vocabulary, represent the relations among the concepts in an unambiguous formal structure, represent contextual information, and formally represent vague concepts.

Understanding natural language requires different kinds of knowledge; humans are generally unaware of the extensive amount of knowledge they use to do so. Computer processes that extract and organize information in text do not attempt to achieve understanding, but the task is still knowledge-intensive and complex. In spite of the underlying difficulties, NLP in the medical domain has definitely begun to show promising results. For example, there are two NLP systems[7,8] that have been integrated into operational clinical information systems. Some NLP systems have been used for decision-support purposes based on radiograph reports. Evaluations showed that these systems were able to identify: (1) abnormalities in chest x-rays[7,9,10]; (2) patients suspected of having tuberculosis; and (3) findings suggesting breast cancer.[12] Most important, the evaluations demonstrated that NLP systems performed as well or almost as well as medical experts in identifying abnormal conditions. Other NLP applications have been developed that encode admission diagnoses,[13] manage patients with asthma,[14] translate findings to SNOMED codes,[15] and automate severity assessment for community-acquired pneumonia.[16]

In this paper, we present an overview of natural language extraction systems. We briefly describe their underlying goals, how such systems are deployed, the types of knowledge they require, and their present stages of development. We then discuss the future of NLP in health care.

Background

Research concerning NLP encompasses many applications other than extraction and encoding, including generation of natural language, summarization, voice recognition, question-answering systems, knowledge representation, knowledge acquisition, literature searching and indexing, clinical vocabularies, domain models, computerized translation, and grammar and spelling correction. Our focus in this article is on systems that extract clinical information from textual patient documents.

Understanding textual language involves several components.[17] Three of the most important are syntactic, semantic, and domain-knowledge components. Understanding natural language involves understanding (1) syntax, or the structure of sentences (which words are subjects, verbs, objects, etc.); (2) semantics, or the meanings of words and how they are combined to form the meaning of a sentence (e.g., *patient complained of excruciating pain in chest* means that the patient had pain, the pain was located in the chest, and the pain was severe); and (3) domain knowledge, or information about the subject matter (e.g., understanding that pneumonia is a lung disease). Not all NLP systems incorporate the same types or amounts of knowledge, and the manners in which the components are integrated into the systems can vary considerably. A few systems offer multilingual capabilities and therefore require additional knowledge components. Some are associated with a conceptual model of a particular domain or a controlled clinical terminology (in other words, narrowly focused domain knowledge).

A controlled vocabulary helps to improve access to clinical information because each vocabulary concept is associated with a unique well-defined meaning, reducing variety and ambiguity. Its presence requires the NLP system to enumerate different ways of expressing the same concept (*heart failure*, CHF, and *congestive heart failure* would be mapped to the same concept). The system must use context whenever possible to map ambiguous words to well-defined terms (*discharge* in *discharge from hospital* would be mapped to a management-type concept and *discharge* in *discharge from wound* would be mapped to a sign/symptom concept).

A domain model organizes terminology into a hierarchy, and delineates well-defined relations among the concepts. This facilitates retrieval because it allows the system to make inferences. For example, if *pneumonia* is defined as a lung disease, a system with an inferencing capability would be able to automatically infer that the patient has a lung disease when *pneumonia* is asserted. A domain model could also associate a finding with *pneumonia* (*consolidation*). NLP systems that do not include inferencing presume that the applications that use the extracted information will provide that capability. In such a case, an application looking for reports associated with lung disease would have to enumerate all the specific lung diseases as well as related findings, or use a different knowledge source that is outside the NLP system.

Although general language processing is still elusive, there has been definite advancement in NLP in medicine because its domain is restricted and constitutes a sub-language. The concept of sub-language grammar[18] was first proposed by

Harris[19] and subsequently incorporated in a text-processing system by Sager.[21,21] Sub-language domains have less variety, ambiguity, and complexity because they are more defined than the general language domain and involve only the specific information and relations relevant to the particular subject. In medicine it is possible to define suitable informational categories, to identify co-occurrence patterns among the categories, and generally to interpret the patterns unambiguously. For example, there is an informational category associated with body location (e.g., chest), with symptom (e.g., pain), and with severity (e.g., severe). The co-occurrence pattern of severity + body location + symptom (e.g., severe chest pain) can be interpreted to mean the symptom "pain" is associated with the body location "chest" and that the severity "severe" is associated with the symptom "pain."

After information is extracted from text, it must be saved in a well-defined format so that subsequent processes can use it. Some systems use a graph-like form, called a conceptual graph (CG),[22] whereas others use frames closely resembling database tables. Recently, some researchers[23–25] suggested using Extensible Markup Language (XML) to represent the processed output because XML is a standard format ideal for Web-based applications.

The State of the Art

There is a large body of literature concerning NLP extraction systems in the medical domain. Below we review 11 systems, which we selected because each has been described more than once in peer-reviewed journals within the past five years. (An overview of NLP systems in medicine published prior to 1995 has been presented by Spyns.[26])

- Sager, who heads the Linguistic String Project (LSP), is a pioneer both in language processing and in medical language processing[21,27] who has greatly influenced the field. The LSP has been involved in the development of one of the first comprehensive NLP systems for general English. Subsequent work involved adapting the system to medical text and to other languages. The system has very comprehensive syntactic and semantic components, and has been applied to numerous clinical domains, including discharge summaries, progress notes, and radiology reports.
- The SPRUS system[7] is a special-purpose radiology-text processor and was one of the first systems that functioned as a module within a working clinical information system, the HELP[28] system. A later version, SYMTEXT, incorporated a syntactic component[29] and was applied to automatically obtaining codes for admission diagnoses. This application has been independently evaluated.
- The MedLEE system[30] also operates as an independent module of a clinical information system at New York Presbyterian Hospital, and is used daily. It has been independently evaluated several times,[9,10,12,31] and studying evaluations of NLP systems is a focus of the development group. MedLEE was the first NLP system used for actual patient care that was shown to improve care.[11] It has also been integrated with two different voice-recognition systems.[32,33]
- An NLP system[34,35] developed at the Geneva Hospital was designed as a multilingual system that could process French, English, and German documents. The developers aimed to create a normalized language-independent representation of medical information. Because the domain modeling was laborious, the system was restricted to patient discharge summaries of patients admitted for gastrointestinal surgery.
- MENELAS[26,36–38] was created by a consortium that aimed to provide better access to information in patient discharge summaries. Two prototype applications were developed for the domain of coronary diseases, a document-indexing system (parts of which have been realized in French, English, and Dutch) and a consulting application that provides users with access to the information in the documents via the indexing system.
- There are several NLP systems in Germany, all in early stages of implementation. One system, MediTas,[39,40] is under development in Georg August University in Göttingen. Another system, developed at the University of Hamburg, is MeTexA,[41] and a third, called MEDSYNDIKATE,[42] is being developed at Freiburg University Hospital.
- A group associated with the Chiba University Hospital in Japan[43,44] developed a prototype system to translate findings in reports into SNOMED and ICD9 codes. It was developed based on case reports described in the *New England Journal of Medicine*.

Resources for NLP in Medicine

A significant amount of work in developing an NLP system concerns extending lexical knowledge. Since the number of words and phrases associated with clinical concepts is very large (over 100,000), the task of adding entries to the lexicon is considerable. The National Library of Medicine has undertaken a large-scale effort to facilitate access to biomedical information. The development of the Unified Medical Language System (UMLS)[45] and the release of the SPECIALIST Lexicon[46,47] will substantially benefit NLP systems. In the UMLS each concept is given a unique identifier, and all synonymous concepts have the same identifier. This feature provides a substantial body of knowledge that NLP systems need to link words in text to a controlled vocabulary (the UMLS or one of the other source vocabularies). The UMLS also has a semantic network and assigns semantic categories to all concepts. For example, *fever* is assigned the category SIGN/SYMPTOM. The categorization provides the semantic knowledge needed by NLP systems to identify relevant units of information. The SPECIALIST Lexicon, which has over 84,000 entries, assigns syntactic categories to words and phrases in biomedical text. The lexicon is useful not only for NLP extraction tasks, but also for indexing and vocabulary development.

Other nomenclatures are also important knowledge sources. Some work investigating the use of SNOMED[48] and ICD10[49] as knowledge sources for lexical work has been published. Like the UMLS, these nomenclatures are also effective for identifying relevant clinical terms and for semantic categorization. Both SNOMED and ICD10 are particularly useful to groups involved in multilingual work because they are available in other languages and because the codes provide a way to link a concept in one language to a similar concept in other languages.

Other types of knowledge sources needed by NLP systems, such as grammars and domain models, are generally developed by individual groups because they are more complex and interrelated than nomenclatures. They are also typically very difficult to adapt to other systems.

FUTURE DIRECTIONS

Natural language processing is likely to become more important shortly because of health care economics. Increasingly, clinical information will play a very important role in determining which health care organizations make higher profits. Reliable information about the process of care and patient outcomes will be critical for attracting more patients and lowering costs. It has been shown that ICD9 codes manually assigned to patients at discharge time for billing purposes are not clinically accurate, and that additional manual encoding solely for clinical purposes is not practical. Therefore, NLP will be a very attractive way to provide the necessary clinical information.

A second reason NLP is likely to become important concerns the Web. Since Web-based technology is becoming pervasive, it will be used more and more by administrators, caregivers, patients, and medical students for accessing clinical information. Search engines now retrieve information by keywords, but this method is not accurate enough. Language processing will be a reasonable way to provide information more accurately. When there is access to encoded clinical data, it will be possible for a caregiver to request a patient's profile or summary of pertinent information, for a medical student to retrieve similar cases for training, for a patient to ask a question and receive an answer tailored to his or her condition, for an administrator to investigate outcomes of a procedure or analyze resource utilization, or for a researcher to find patients for a clinical study.

Because of the explosion of Web-based applications, it is very likely that more Web-based clinical applications will soon be deployed. Up to now, Web browsers such as Netscape and Microsoft Explorer have read files that contain HTML (HyperText Markup Language) tags. These tags are added to the information in the file to tell a browser how to present the contents of the file. XML (Extensible Markup Language) is the latest general markup language developed for the Web that provides more general functionality for Web-based applications than HTML provides. NLP systems will be very desirable to use because they will be able to enrich the information in clinical reports with XML tags, thereby preparing the information in the reports for Web applications.

Another technology that will further NLP is continuous voice recognition. Continuous voice recognition systems are now becoming commercially viable for clinical applications. The availability of these systems means that physicians themselves will enter clinical data into patient records because voice recognition systems are more cost efficient and more timely than transcription services. Integrating a voice recognition system with an NLP system will substantially enhance the functionality of the voice system. It will enable physicians to dictate their reports in their usual fashion while the natural language processor translates the textual report into a structured encoded form in the background. The coded data obtained by the NLP system could then be stored in real time along with the original text in a clinical repository. This would enormously increase the functionality of the voice system.

Voice recognition systems depend on large body of text for training because they are based on statistical methods. Because collections of clinical information are becoming available electronically, we are also likely to see the development of statistically based language processors and processors that combine statistical and knowledge-based methods.

Future NLP systems will likely produce standard terminology and standard output forms. We will see the generation of standard terminology because the large-scale vocabulary integration efforts of the UMLS facilitate mapping both from text to UMLS concepts and from one terminology to another. In addition, we will see standard output forms because NLP systems will likely generate XML, an output form suitable for the Web. The use of XML will allow for processing by multiple systems because the structure of XML output is well defined when used in conjunction with a document-type definition. The use of XML tags should also cause an increase in the development of text-processing systems that are specialized or layered. For example, one specialized processor could identify and tag vital signs, whereas another specialized processor could read previously tagged text and proceed to identify and tag another type of specialized clinical information.

Although medical language processing is complex, effective systems are becoming a reality. Because of the Web, current economic factors, and the availability of commercially viable speech technology, we believe NLP will soon become a very important technology in health care.

This work was supported by grants R01-LM06274 and R29-LM05397 from the National Library of Medicine; a Center for Advanced Technology grant from the New York State Science and Technology Foundation; and a grant from Pfizer, Inc., to study community-acquired pneumonia management.

***Dr. Friedman** is professor, Department of Computer Science, Queens College, City University of New York, and associate professor, Department of Medical Informatics, Columbia University; **Dr. Hripcsak** is associate professor, Department of Medical Informatics, Columbia University; both in New York, New York.*

Address correspondence and requests for reprints to Dr. Friedman, Department of Medical Informatics, 161 Ft. Washington Ave., DAP-1310, New York, NY 10032; e-mail: <friedman.carol@columbia.edu>.

References

1. Dick R, Steen E (eds). The computer-based patient record: an essential technology for health care. Institute of Medicine. Washington, DC: National Academy of Sciences Press, 1991.
2. Shortliffe EH, Buchanan BG. Knowledge engineering for medical decision making: a review of computer-based clinical decision aids. Proc IEEE. 1979;67:1207–24.
3. Miller RA, Pople H, Myers JD. Internist-1: an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med. 1982;307:468–76.
4. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectibility of man. N Engl J Med. 1976;292:1351–5.
5. Gardner RM., Cannon GH, Morris AH, Olsen KR, Price G. Computerized blood gas interpretation and reporting system. Computer. 1975(January):39–45.
6. Barnett GO, Winickoff RN, Morgan MM, Zielstorff RD. A computer-based monitoring system for follow-up of elevated blood pressure. Med Care. 1983;21:400–9.
7. Haug PJ, Ranum DL, Frederick PR. Computerized extraction of coded findings from free-text radiologic report. Radiology. 1990;174:543–8.
8. Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. Natural Language Engineering. 1995;1:83–108.
9. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports. Ann Intern Med. 1995;122:681–8.
10. Hripcsak G, Kuperman GJ, Friedman C. Extracting findings from narrative reports: software transferability and sources of physician disagreement. Meth Inform Med. 1998;37:1–7.
11. Knirsch CA, Jain NL, Pablos-Mendez A, Friedman C, Hripcsak G. Respiratory isolation of tuberculosis patients using clinical guidelines and an automated decision support system. Infect Control Hosp Epidemiol. 1998;19:94–100.
12. Jain NL, Friedman C. Identification of findings suspicious for breast cancer based on natural language processing of mammogram reports. In: Masys DR (ed). Proceedings of the Fall 1997 American Medical Informatics Association Conference, 1997 Oct 25–29, Nashville, TN. Philadelphia, PA: Hanley & Belfus, 1997: 829–33.
13. Gundersen ML, Haug PJ, Pryor TA, et al. Development and evaluation of a computerized admission diagnoses encoding system. Comput Biomed Res. 1996;29:351–72.
14. Lyman M, Sager N, Tick LJ, Nhan NT, Borst F, Scherrer JR. The application of natural-language processing to healthcare quality assessment. Med Decis Making. 1991;11:S65–S68.
15. Sager N, Lyman M, Buchnall C, Nhan NT, Tick LJ. Natural language processing and the representation of clinical data. J Am Med Informat Assoc. 1994;1:142–60.
16. Friedman C, Knirsch CA, Shagina L, Hripcsak G. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. In Lorenzi NM (ed). Proceedings of the Fall 1999 American Medical Informatics Association Conference, 1999 Nov 6–10, Washington, DC. Philadelphia, PA: Hanley & Belfus, 1999 (in press).
17. Allen J. Natural Language Understanding. New York: Benjamin/Cummings, 1995.
18. Grishman R, Kittredge R. Analyzing language in restricted domains: sublanguage description and processing. Hillsdale, NJ: Lawrence Erlbaum, 1986.
19. Harris Z. Mathematical Structures of Language. New York: Wiley Interscience, 1968.
20. Sager N. Sublanguage: linguistic phenomenon, computational tool. In: Grishman R, Kittredge R (eds). Analyzing Language in Restricted Domains: Sublanguage Description and Processing. Hillsdale, NJ: Lawrence Erlbaum, 1986: 1–17.
21. Sager N, Friedman C, Lyman M, et al. Medical Language Processing: Computer Management of Narrative Data. Reading, MA: Addison–Wesley, 1987.
22. Sowa J. Conceptual Structures: Information Processing in Mind and Machine. Reading, MA: Addison–Wesley, 1984.
23. Sager N, Nhan NT, Lyman M, Tick LJ. Medical language processing with SGML display. In: Cimino JJ (ed). Proceedings of the 1996 Fall American Medical Informatics Association Conference, 1996 Oct 26–30, Washington, D.C. Philadelphia, PA: Hanley and Belfus, 1996:547–51.
24. Zweigenbaum P, Bouaud J, Bachimont B, Charlet J, Seroussi B, Boisvieux JF. From text to knowledge: a unifying document-oriented view of analyzed medical language. Meth Inform Med. 1998;37:384–93.
25. Friedman C, Hripcsak G, Shagina L, Liu H. Representing information in patient reports using natural language processing and the extensible markup language. J Am Med Informat Assoc. 1999;6:76–87.
26. Spyns P. Natural Language Processing in medicine: an overview. Meth Inform Med. 1996;35:285–301.
27. Sager N. Natural Language Processing: a computer grammar of English and its applications. Reading, MA: Addison–Wesley, 1981.
28. Pryor TA, Gardner R, Clayton PD, Warner HR. The HELP System. Med Syst. 1983;7:87–102.
29. Haug PJ, Koehler S, Lau ML, Wang P, Rocha R. Experience with a mixed semantic/syntactic parser. In: Gardner R (ed).Proceedings of the Symposium on Computer Applications in Medical Care, 1995 Oct 28–Nov 1, New Orleans, LA. Philadelphia, PA: Hanley & Belfus, 1995: 284–8.
30. Friedman C, Alderson PO, Austin J, Cimino JJ, Johnson SB. A general natural language text processor for clinical radiology. J Am Med Informat Assoc. 1994;1:161–74.
31. Jain NL, Knirsch CA, Friedman C, Hripcsak G. Identification of suspected tuberculosis patients based on natural language processing of chest radi-

ograph reports. In Cimino JJ (ed). Proceedings of the 1996 American Medical Informatics Association Annual Fall Symposium, 1996 Oct 26–30, Washington, DC Philadelphia, PA: Hanley & Belfus, 1996: 542–16.

32. Starren J, Friedman C, Johnson SB. The Columbia Integrated Speech Interpretation System (CISIS). In: Gardner R (ed). Proceedings of the Symposium on Computer Applications in Medical Care, 1995 Oct 28–Nov 1, New Orleans, LA. Philadelphia, PA: Hanley & Belfus, 1995:985.
33. Rosenthal D, Friedman C. Continuous-speech structured reporting. In: Chute CG (ed). Proceedings 1998 American Medical Informatics Association Annual Symposium, 1998 Nov 7–11, Orlando, FL. Philadelphia, PA: Hanley & Belfus, 1998:1116.
34. Baud RH, Rassinoux AM, Scherrer JR. Natural language processing and semantical representation of medical texts. Meth Inform Med. 1992;31:117–25.
35. Rassinoux AM, Michel PA, Juge C, Baud RH, Scherrer JR. Natural language processing of medical texts within the HELIOS environment. Comput Meth Prog Bio. 1994;45:S79–S96.
36. Zweigenbaum P. MENELAS: an access system for medical records using natural language. Comput Meth Progr Biomed. 1994;45:117–20.
37. Zweigenbaum P, Bachimont B, Bouaud J, Charlet J, Boisvieux JF. A multilingual architecture for building a normalized conceptual representation from medical language. In: Gardner R (ed). Proceedings of the Symposium on Computer Applications in Medical Care, 1995 Oct 28–Nov 1, New Orleans, LA. Philadelphia, PA: Hanley & Belfus, 1995:357–61.
38. Delamarre D, Burgun A, Seka LP, Le Beux P. Automated coding of patient discharge summaries using conceptual graphs. Meth Inform Med. 1995;34:345–51.
39. Pietrzyk PM. A medical text analysis system for German—syntax analysis. Meth Inform Med. 1991;30:275–83.
40. Pietrzyk PM. Free text analysis. Int J Biomed Comput. 1995;39:139–44.
41. Schroder M. Knowledge based processing of medical language: a language engineering approach. In: Ohlbach HJ (ed). Proceedings of the Sixteenth German Workshop on AI (GWAI). Berlin, Germany: Springer-Verlag, 1992; 221–34.
42. Hahn U, Romacker M, Schulz S. Why discourse structures in medical reports matter for the validity of automatically generated text knowledge bases In: Cesnik B, et al (eds). MEDINFO 98 Proceedings, 1998 Aug 18–22. Seoul, Korea. Amsterdam, The Netherlands: IOS Press, 1998; 633–8.
43. Do Amaral M, Satomura Y. Associating semantic grammars with the SNOMED: processing medical langauge and representing clinical facts into a language-independent frame. In: Greenes RA (ed). MEDINFO 95 Proceedings, 1995 August. Vancouver, British Columbia. Amsterdam, The Netherlands: IOS Press, 1995; 18–22.
44. Satomura Y, Do Amaral M. Automated diagnostic indexing by natural language processing. Meth Inform Med. 1992;17:149–63.
45. Lindberg DAB, Humphreys B, McCray AT. The Unified Medical Language System. Meth Inform Med. 1993;32:281–91.
46. McCray AT, Srinivasan S. Automated access to a large medical dictionary: online assistance for research and application in natural language processing. Comput Biomed Res. 1990;23:179–98.
47. McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. In: Ozbold JG (ed). Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care, 1994 Nov 5–9, Washington, D.C. Philadelphia, PA: Hanley & Belfus, 1994:235–9.
48. Zweigenbaum P, Courtois P. Acquisition of lexical resources from SNOMED for medical language processing. In: Cesnik B, et al (ed). MEDINFO 98 Proceedings, 1998 Aug 18–22, Seoul, Korea. Amsterdam, The Netherlands: IOS Press, 1998;586–90.
49. Baud RH, Lovis C, Rassinoux AM, Michel PA, Scherrer JR. Automatic extraction of linguistic knowledge from an international classification. In: Cesnik B, et al (eds). MEDINFO 98 Proceedings, 1998 Aug 18–22, Seoul, Korea. Amsterdam, The Netherlands: IOS Press, 1998: 581–5.