

F-C. Jean

AT&T Istel,
Issy-les-Moulineaux, France.

Synopsis

Information Systems

Introduction

This section of the Yearbook gathers nine articles under the heading of *Information Systems*. If one browses rapidly through their titles, one can imagine that these papers have little to do with one another, or at least that they form independent clusters dealing with different problems. However, a closer look at the addressed topics reveals that they are pivotal considerations when dealing with highly distributed information systems, (i.e., systems that grow beyond the physical limits of the hospitals to address global health networks): how to access health-care data, how to extract relevant information from those data, how the security of the data is managed, and last but not least: was it really clinically useful to access those data?

Accessing to Information

The first article in this category *Communication Services for a Distributed Hospital Information System* by Graeber [1], although focusing mainly on the hospital environment, describes a communication infrastructure able to control and perform all information exchanges between heterogeneous applications that form modern health information systems. The most original part of the paper consists of the use of the Semantic Object Model to describe the numerous business processes that interact in a complex communication schema. An

interaction model, defining the structural view of business processes, is seen as a sequence of interaction schemes involving objects interacting according to the client-server model. The global information system is thus modeled at the procedure level. The application part of the paper describes a centralized Communication Server, based on a commercially available product, that manages all the communication problems (i.e., physical and logical connections between communicants, syntactic and semantic adaptation/translation of the passed messages, security and administration). The discussion compares the Communication Server concept and implementation with the architectures used in major distributed HIS projects.

Retrieving Relevant Information

The two articles presented under this heading both deal with the problem of improving information retrieval from information systems.

The first paper *Retrieval Feedback in MEDLINE* by Srinivasan [2] describes a method that improves MEDLINE search performance without putting extra cognitive load on the user. This aspect is particularly important since end-users are generally unskilled in the use of any controlled vocabulary (in this case the MeSH) to express their queries. The author first reviews the different query-improvement strategies that have been de-

signed for various databases, focusing on query expansion ones, where a user's original query is automatically modified into one judged more effective for retrieval. Feedback approaches, where a query is modified using information in previously retrieved documents, are then discussed. The core part of the paper concerns the investigation of new strategies for query expansion within the retrieval-feedback framework, using either controlled vocabulary and/or free text, and the measurement of their effectiveness. The very important outcome of this study is that it confirms the importance of MeSH for MEDLINE searching and that it is possible to propose methods that do not require the user to be trained in the construction and selection of MeSH concepts. Unfortunately, and as for relevance ranking, commercially available retrieval systems for MEDLINE have still not included such necessary improvements.

The problem addressed of the second paper *A Concept-Based Retrieval System for Thoracic Radiology* by Taira et al. [3], although dealing with a different information medium, is very similar to the one discussed in the previous paper: how to improve the information system in order to retrieve more relevant information from it. The paper describes a prototype system that supports intelligent information retrieval, based on multimedia document content, using natural language processing agents. The described system is based on the commonly shared assumption that the natural language

processing (NLP) approach of the radiology reports is the most viable means to create large image repositories that can be used for research and teaching. The system is structured as three complementary subsystems: (1) a free-text indexing system to summarize and structure the semantic content of the reports and to create indexable abstracts; (2) an intelligent query processing engine using a lexical knowledge base; and (3) a user interface supporting the expression of high-level queries. The paper highlights the report indexing sequence: preprocessing that leads to a topicological transform of the original reports into subreports that are directed to NLP agents whose topic of expertise match the finding topic of the subreports, transformation of meaningless text strings to meaningful concepts (lexical analysis), pattern analysis to find the relations between the concepts and then construction of the frames used to create the document index. The main interest of this approach is that it does not require manual indexing. Thanks to its three-level object-oriented data model that includes a comprehensive (but narrowed to a limited part of radiology) lexicon, the system allows powerful concept-based retrieval that must be compared to the simple string-based match retrieval of many available systems.

Ensuring the Security and Confidentiality

All of the five papers of this category deal with "health data security" but are highly complementary, since each of them focuses on a different aspect of the concept. Data security, which is a major concern for any electronic medical record (EMR) designer or manager, becomes one of the hottest topics when dealing with remote access to EMR, such as in distributed information systems. But what is data

security? To what extent are the fears regarding loss of privacy justified? What are the "acceptable" trade-offs between security, accessibility, user convenience and performances? The papers in this section try to address all of these questions since both [4] and [5] propose a classification of the security problems regarding the actions to protect against the possible points of illegal access, and [6] proposes an ontology of potential intruders of information systems. Although cryptography is sometimes presented as the 'silver bullet' for protecting information, it is stated that using very strong cryptographic algorithms may be an overkill in view of some of the weakness of the information system, and that cryptography can not protect against destructive attacks.

The first article in this category *Privacy, Confidentiality, and Electronic Medical Records* is a review by Barrows and Clayton [4]. The paper clearly exposes the goals of information security and privacy in health-care, related to the different fears of loss of privacy due to the availability of health-care data in electronic medical records. Since it has been proved that the main problem with information security is not technology but the lack of a consistent security policy, the paper reviews major security holes and exposes how a strong policy could constitute a reasonable security solution.

Many papers have already discussed the means to protect health data from unauthorized users. The main interest of the paper from Miller and Cooper: *Security considerations for present and future medical databases* [5] is that it points out the necessity to also protect the information systems from authorized users. The authors position their study at the database level (and propose in their conclusion to include in the DBMS engines themselves most of the required mechanisms to enhance the security of data); however, most of the concerns could be gener-

alized to reach the level of the global information system. Besides a discussion about the granularity of the access rights linked to the different views of the database, it is stated that a less known security hole is linked to a possible unauthorized disclosure of data due to data inference resulting from the correlation of data obtained by legitimate means through statistical queries with the aid of additional, and usually external, knowledge.

Unlike in [5], where security mechanisms are proposed to be embedded in every component of the information system, the following two papers propose to centralize the mechanisms that ensure security (trust center, front-end authorization system).

In *A Cryptologic Based Trust Center for Medical Images* [6], Wong first presents a review of the different cryptographic techniques, comparing private (i.e., DES and IDEA) and public (i.e., RSA) key-based methods, both in terms of strength (how expensive will it be to crack the code through a "brute-force" attack) and of processing speed. The implementation part of the article presents an original combination of both methods. The second interest of the paper consists in the development of the trust center concept that allows to integrate those cryptographic algorithms coherently into existing multimedia information systems. The discussion measures the impact of compression and cryptography on the time-to-access information regarding both the size of data and the medium used to access it.

Vassilacopoulos et al. in *A Front End Authorization Mechanism for Hospital Information Systems* [7] describe a front end authorization mechanism developed to enhance the security features of an existing information system. A potential drawback of many HIS is that authorization and access control is done at the application level, overriding the database level authorization features (linked the con-

cerns of [5]). The paper focuses mainly on role (both permanent and temporary) based authorization mechanisms. Since 100% security is both unrealistic and usually unnecessary, the paper frames the security enhancement subject with the economic, functional and technical constraints of the real world.

In the last paper in this category, *Pseudonyms for Cancer Registries*, Pommerening et al. [8] propose the use of pseudonyms to ensure privacy. The paper poses the problem of long-term storage of person-related data and solves it by encrypting identification data that leads to quasi-anonymized records. As in other papers in this section, the implementation part describes the use of a combination of cryptographic algorithms and addresses the key management problem. The cryptographic concept of pseudonymity presented here can be adapted to all the situations where conflicts between the goals of privacy and public interest must be solved. In the core part of the paper, the authors detail (as in [6]) the role of a trusted third party in the organizational structure of the registry and describe the flows between the source of information, the trusted office and the different "information users". This third party approach, which is commonly used in most of the electronic business (considered in its broad sense) applications, will certainly be the key consideration when dealing with access to patient electronic medical records from every node of the health chain.

Clinical Value of Information Systems

The last paper in this section of the Yearbook: *The Clinical Value of Computerized Information Services*

by Balas et al. [9] is a review of 98 randomized clinical trials addressing the efficacy of clinical information systems. The debate on the value of computer systems in clinical applications is as old as medical informatics. However, world-wide health cost containment programs, with their intensive search for cost-effective methods, currently demand adequate technology assessment of clinical information systems. The objectives of the authors in this study were to review all trials that address the efficacy of such systems, using systematic and quantitative methods. The authors employ the techniques commonly used in drug trials that are relevant for testing computerized information intervention (e.g. selection of eligibility criteria). The extensive reference section (119 references) is without doubt also a major interest of the paper. The result section of the article highlights the fact that most information services were tested in outpatient care, which is surprising considering the huge investments made in hospital information systems.

Conclusion

Current medical information systems can not be limited to the hospital environment. They must provide all the health professionals involved in patient care with the right means to access the relevant information needed to improve the care given to their patients, subtended by Community Health Information Networks (CHIN). To reach this goal, both technical and non-technical issues in the fields of communication, accessibility, data organization, information retrieval and security must be solved. The papers proposed in this section, while reflect-

ing the problematic nature of modern information systems, prove that technology has reached the required state of maturity to be transferred to producers and end-users.

References

1. Graeber S. Communication services for a distributed hospital information system. *Meth Inform Med* 1996;35:230-41.
2. Srinivasan P. Retrieval feedback in MEDLINE. *J Am Med Inform Assoc* 1996;3:157-67.
3. Taira R, Johnson D, Bhushan V, Rivera M, Wong C, Huang L, Aberle D, Greaves M, Goldin J. A concept-based retrieval system for thoracic radiology. *J Digital Imaging* 1996;9:25-36.
4. Barrows RC, Clayton P. Privacy, confidentiality, and electronic medical records. *J Am Med Inform Assoc* 1996;3:139-48.
5. Miller M, Cooper J. Security considerations for present and future medical databases. *Int J Biomed Comput* 1996;41:39-46.
6. Wong S. A cryptologic based trust center for medical images. *J Am Med Inform Assoc* 1996;3:410-21.
7. Vassilacopoulos G, Peppes D. A front end authorization mechanism for hospital information systems. *Med Inform* 1996;21:93-103.
8. Pommerening K, Miller M, Schmidtman I, Michaelis J. Pseudonyms for cancer registries. *Meth Inform Med* 1996;35:112-21.
9. Balas E, Austin S, Mitchell J, Ewigman B, Bopp K, Brown G. The clinical value of computerized information services. *Arch Fam Med* 1996;5:271-8.

Address of the author:
François-Christophe Jean, MD, PhD,
AT&T Istel,
158 Avenue de Verdun,
92130 Issy-les-Moulineaux,
France.
e-mail: fcj@att.fr