Management of clinical XML documents
A pragmatic approach

Schweiger R, Buerkle T, Hoelzer S, Dudeck J

Department of Medical Informatics, University of Gießen, Heinrich-Buff-Ring 44, 35392 Gießen, Germany

In the healthcare domain we often find textual documents that are difficult to exploit with respect to their content. We discuss a gradual approach that uses the standard XML technology to insert more structure and more codes into textual documents.

PROBLEM

The information retrieval of medical documents is still a problem due to the fact that many data are stored as narrative text only. Experts in medical information science consequently postulate the insertion of more structure and more codes into medical documents (1) and database solutions are often not flexible enough. The reasons for this situation are manifold. Healthcare data are document oriented, i.e. they are organized into cohesive communication units that usually contain different kinds of data (e.g. administrative and clinical). Besides, healthcare data need to be very flexible with respect to their structure. It must be possible to put a comment into any context of the clinical document. We cannot reduce medical documents to a set of related data items since items often have a narrative context.

METHODS

With XML on the other hand, we may invent markup for the identification of single items in textual documents, i.e. we add structure to the data rather than adding data to a given structure. The data consequently remain in a cohesive form and the structure remains flexible. The problem of the XML approach reveals when managing a large set of related documents. We run the risk of losing control over the items contained in the various documents. What we need is a sort of data model or schema that keeps track of the items marked up in the documents.

The idea is to let the document author extend the document where and when necessary and to automatically update an XML schema definition (2) which represents the overall document model. The XML schema definition is a superset of the items contained in a set of related documents and can be used to establish search strategies upon the document set. XML consequently provides the means to preserve the flexibility of free text documents AND to schematize the data at the same time.

RESULTS

We have implemented our approach on an Internet platform and we have used it for the organization of drug guidelines and pathology reports. With this approach it is possible to search e.g. for all reports with a specific pathological finding and to insert new items and narrative context where and when necessary.

CONCLUSION

Unstructured data might be a result of inflexible data models. Our approach to this problem is to let the user requirements change the model, i.e. to adjust the structure to the data (bottom up) instead of adjusting the data to the structure (top down). XML provides the means to implement such an approach.

REFERENCES