



Chapter VII

**Unifying Access to
Heterogeneous Document
Databases Through
Contextual Metadata**

Virpi Lyytikäinen
University of Jyväskylä, Finland

Pasi Tiitinen
University of Jyväskylä, Finland

Airi Salminen
University of Jyväskylä, Finland

ABSTRACT

Document databases available on the Internet carry massive information resources. To a person needing a piece of information on a specific domain, finding the piece, however, is often quite problematic even though there were a representative collection of databases available on the domain. The languages used in the content, the names of document types, their structures, the ways documents are organized and their retrieval techniques often vary in the databases. The databases containing legal information on the Internet offer a typical example. For finding relevant documents and for being able to interpret the content of the documents correctly, the user may need information about the context where the documents have been created. In this chapter we introduce a method for collecting contextual metadata and for representing the metadata to the users by graphical models. The solution is demonstrated by a case of retrieving information from distributed European legal databases.

INTRODUCTION

Document databases available on the Internet carry massive information resources. To a person needing a piece of information on a specific domain, finding the piece is often quite problematic even though there were a representative collection of databases available on the domain. The databases containing legal information on the Internet offer a typical example. Recent trend towards electronic government (e-government) has tremendously increased the amount of public information available to citizens and organizations on the Internet. In Europe, due to the development towards European integration, this information is increasingly needed also regarding foreign European legal systems and the European Union itself. The legal databases are, however, organized in different ways, their content is written in different languages and their retrieval techniques vary. Differences in legal systems aggravate the problems of information retrieval. Similarly, on the intranets and extranets of organizations, information resources are often scattered in heterogeneous repositories. In paper machine manufacturing, for example, the amount of cooperating organizations and documents involved is always immense, and the information related to the manufacturing process is more and more stored on documents available via intranets and extranets.

Document retrieval tools always utilize metadata in information access. The *metadata* is data about the documents in the repository. It can be, for example, data about their content, about the way they are organized or about their structure. It can also be data about the context where the documents have been created. Especially in cases where the language of the content, the ways documents are organized, as well as their structures, vary in the databases of the repository, the role of *contextual metadata* becomes important both for finding relevant documents and for being able to interpret the content of the documents correctly. In this chapter, we introduce a method for collecting contextual metadata and for representing the metadata to the users by graphical models. The solution is demonstrated by a case of retrieving information from distributed European legal databases.

The rest of the chapter is organized as follows. The next section describes legal information in Europe as a case and demonstrates the challenges related to information retrieval in the case. A method for collecting contextual metadata is then introduced, and we show a way to visualize the contextual metadata in a user interface. An interface for finding legal information from European legal databases is described as an example. Finally, we discuss future directions related to the metadata support in information access.

EUROPEAN LEGAL INFORMATION AS A CASE

The legal domain in Europe offers an interesting and important example of distributed, heterogeneous information resources. The case was studied in a project called EULEGIS (European User Views to Legislative Information in Structured Form) belonging to the Telematics Application Programme of the European Commission. The main practical intent of the project was to offer single-point Web access to European legal

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/unifying-access-heterogeneous-document-databases/9207

Related Content

The Impact of Network Layer on the Deadline Assignment Strategies in Distributed Real-Time Database Systems

Victor C.S. Lee, Kam-Yiu Lam, Kwok-Wa Lam and Joseph K.Y. Ng (1996). *Journal of Database Management* (pp. 24-33).

www.irma-international.org/article/impact-network-layer-deadline-assignment/51163

Extended Entity Relationship Modeling

Sikha Bagui (2005). *Encyclopedia of Database Technologies and Applications* (pp. 233-239).

www.irma-international.org/chapter/extended-entity-relationship-modeling/11152

Development of an E-Healthcare Information Security Risk Assessment Method

June Wei, Binshan Lin and Meiga Loho-Noya (2013). *Journal of Database Management* (pp. 36-57).

www.irma-international.org/article/development-of-an-e-healthcare-information-security-risk-assessment-method/84068

Novice Designer Performance Comparison Between the Entity Relationship Event Network and the Event-Based Logical Relational Design Techniques

Dinesh Batra and Nicole Wishart (2014). *Journal of Database Management* (pp. 1-27).

www.irma-international.org/article/novice-designer-performance-comparison-between-the-entity-relationship-event-network-and-the-event-based-logical-relational-design-techniques/118086

Natural Language Front-End for a Database

Boris Galitsky (2005). *Encyclopedia of Database Technologies and Applications* (pp. 403-407).

www.irma-international.org/chapter/natural-language-front-end-database/11180