

User Roles in Document Analysis

Pasi Tiitinen

P.O. Box 35 (Agora), University of Jyväskylä,
Department of Computer Science and Information Systems,
FI-40014, Jyväskylä, Finland
pti@cc.jyu.fi

Abstract. In a modern business process, documents are important carriers of information between organisations. Document management solutions are increasingly based on structured document formats, such as XML (Extensible Markup Language). However, deployment of structured documents is a complicated task, which requires systematic analysis of requirements of all groups either producing or using documents. In this paper, we describe a method for modelling user roles in document analysis. The method is used to support analysis of requirements of individual and organisational users of documents as well as analysis of organisational needs related to security and access control.

1 Introduction and background

The concept of *role* is used in organizational modelling in order to specify organizational structures and activities which individuals perform in organization. Role-based access control (RBAC) (e.g. [5]) is used to simplify complex security administration of large systems. Besides the security, role-based access can also improve the usability: support for roles serves as support for varying needs of users. Roles as abstractions of various types of users having different needs and tasks serve the requirements analysis, and design of systems (e.g., [4]).

In modern organisations, document management solutions are increasingly based on structured document formats, most commonly XML (Extensible Markup Language) or SGML (Standard Generalized Markup Language). The deployment of structured documents in organisations requires agreement upon the schemas that define how information is represented in documents. This process is called *document standardisation* [8]. The deployment of structured documents is a complicated task, which requires usually development of new systems as well as work practices. For this purpose, systematic analysis of requirements of all groups either producing or using documents is required. Current document management practices must be studied, after which new document structures, practices and management systems are developed. This process is called *document analysis* [8]. Especially when the document production is part of inter-organizational business processes, the analysts should be able to utilize efficient methods for elicitation of the user requirements.

Deployment of structured documents raises new possibilities and challenges also for access control. Often there is a need to control reading and modification of document portions, i.e., at the granularity of XML or SGML document elements (see, e.g., [1]). For example, eXtensible Access Control Markup Language (XACML) [6] supports defining role-based security policies for XML documents and their parts. It uses XPath [3] expressions for identifying parts of XML documents on which the access control policies are targeted. There are also other proposals for access control mechanisms for XML documents (for example [1]). The focus of our work lies in requirements elicitation, and it is thus complementary to these approaches.

A *role*, in our definition, is an abstraction referring to a group of people with similar document management needs, tasks and responsibilities. *Role modelling* is performed for identifying the key users of documents in the domain and for gathering information about their activities. This information is used for defining access rights as well as for developing new systems for document management. Methods for analysis of access rights presented here extend the methodology for document analysis originally developed in a project called RASKE, which took place during the years 1994-1998. Existing RASKE methodology has been utilised in the Finnish Parliament and ministries to elicit information about requirements of individual and organisational users of documents for developing new document management solutions [8].

There are several other approaches for role-based or user group-based application development (see, e.g. [2], and [4]). None of these, however, pays special attention to requirements that use of structured document formats necessitates. Analysis of organisational structures and work tasks for finding suitable roles, role hierarchies and constraints for RBAC has been called role engineering (see, e.g., [7]). However, there is also scarcely literature describing how the role engineering takes place [7].

2 Modelling of the domain

The document analysis starts by specifying the domain to be analyzed (for more in-depth discussion see [8]). At this point, a project council is created, which has representatives from organisations involved in the domain. Analysis of literature sources, such as project reports, written rules and system manuals, meetings of the project council and informal interviews of the domain experts gives information about the creation and use of documents in the domain. At the same time, current document management systems are examined. The results are used in *process modelling*, *information modelling* and *role modelling*. In process modelling, the processes in which documents are created are divided into smaller activities. In addition, the actors responsible for those activities and information created and used in them are identified. The process and the most important organisational actors are described graphically (see [8]).

In information modelling, documents of the domain are identified. Moreover, their content and structure, lifecycles and relationships to each other are described. The lifecycles are described graphically by state transition diagrams. The diagrams give detailed information about the phases in which people in different job positions and functions modify or use the documents and thus information which is needed for de-

fining roles and permissions. In role modelling, the most important producers and users of information in the domain are identified. People are grouped to roles based on similar needs, tasks, and responsibilities. Roles may often be derived from organisational structure. On the other hand, they may reflect the organisation's essential functions and the connection to organisational structure may not be apparent.

In the following user needs analysis, information is gathered by semi-structured interviews. The idea is to gather information about users' different work tasks, after which the needs for information management are discussed in relation to these. The interviewees are asked what documents or other sources of information they need in a specific task and what outcomes there are. More people are interviewed from roles where the differences in work tasks are large within the role. After user needs analysis, initial document structures in the form of DTDs or XML Schemas are created.

3 Analysis of the roles

After analysis of the information gathered with RASKE methodology there will be several results reported according to different roles; use of different applications, modification and use of most important documents and document parts, and descriptions of opinions and needs concerning different documents. In this phase, requirements for access rights concerning different document types are analysed. Information about requirements is gathered to a table showing what kind of requirements (e.g., creation, reading and modifying) concerning certain document type or parts of it people in a certain role have. In addition to other user needs based requirements, access rights requirements strongly affect the design of document structure, since the schema must support access control for document parts containing sensitive information. Moreover, the schema should be so designed that the document views, which consist of those document parts that the user is allowed to see or modify, always form an integrated whole when requested from the document management system. The analysts must also consider in which situations the access policies defined in some level of the document hierarchy should propagate to lower levels of the hierarchy. Propagation options may differ, for example, between document types.

Analysis of gathered information and discussions with domain experts help to define, whether access right authorizations may be separately specified for individual documents besides of document types. This is important, since often some documents may contain confidential information while other documents of the same type do not. It must also be decided, whether authorizations defined for individual documents may take precedence over document type level definitions. Different conflict resolution policies may be used to solve possible conflicts, see, for example [1].

At this point, the roles which have same kind of requirements and same needs for permissions can be merged to help the administration. However, if the roles have, for example, very different work tasks, it is better to keep them separated, since their requirements may change later. Finalising of the role model is done together with domain and security experts. Possible constraints concerning, for example, separation of duty principles are identified and carefully scrutinised with domain experts. A role hierarchy may be created after roles and related permissions are defined. Since inheri-

tance of permissions is feasible only in few situations, it is usually not practical to place all roles into same hierarchy. Therefore, the roles are usually split into separate groups.

Finally, the results of the document analysis are published in a report that is broadly disseminated. This is helpful for sharing knowledge about the current situation and ideas about possible changes especially between people working in different roles, since their requirements may be very different.

4 Conclusion

Structured document formats are increasingly used for sharing information between organisations, e.g., in electronic commerce or government. Role modelling can be a powerful tool in the development of document management. Analysis of roles assists the developers to understand the domain. It enhances usability by facilitating the gathering and analysis of requirements of different kinds of users. It also enhances security of structured documents by simplifying the design and management of access policies. The possibility of using standardised techniques, such as XACML, should ease the deployment of RBAC for structured documents in the future.

References

1. Bertino, E., Castano, S., Ferrari, E., & Mesiti, M.: Protection and administration of XML data sources. *Data & Knowledge Engineering*, 43: 237-260, 2002.
2. Casteleyn, S., & De Troyer, O.: Structuring Web Sites Using Audience Class Hierarchies, *Proceedings of DASWIS 2001 workshop*, Yokohama, Japan.
3. Clark, J., & DeRose, S.: XML Path Language (XPath) Version 1.0. W3C Recommendation 16 November 1999.
4. Constantine, L., & Lockwood, L.: Usage-centered engineering for Web applications. *IEEE Software*, 19 (2), 2002.
5. Ferraiolo, D., Sandhu, R., Gavrila, S., Kuhn D. R. & Chandramouli, R.: Proposed NIST Standard for Role-Based Access Control. *ACM Transactions on Information and System Security*, 4, (3): 224–274, 2001.
6. Godik, S., & Moses, T.: OASIS eXtensible Access Control Markup Language (XACML). Committee Specification 1.0, 8 October 2002.
7. Neumann, G., & Strembeck, M.: Design and Implementation of a Flexible RBAC-Service in an Object-Oriented Scripting Language. *Eighth ACM Conference on Computer and Communication Security (CCS)*, Philadelphia, USA, November 2001.
8. Salminen, A.: Methodology for document analysis. In A. Kent (Ed.), *Encyclopedia of Library and Information Science*, Vol. 67 (Supplement 30) (pp. 299-320). New York: Marcel Dekker, 2000.