AUTOMATIC ANONYMIZATION OF JUDGMENTS IN ELECTRONIC FORMAT

Goran Sladić, Stevan Gostojić, Branko Milosavljević, Zora Konjović
{sladicg, gostojic, mbranko, ftn_zora}@uns.ac.rs
University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia

Abstract – *This paper proposes a novel method for anonymization of judgments, represented in XML format, according to the predefined anonymization rules. The method is based on the access control for XML documents enforced using the XXACF framework. AKOMA NTOSO was used as an electronic format for representation of judgments in order to enable efficient integration of the proposed solution with existing legal information systems.*

Keywords – *anonymization, judgments, RBAC, XXACF, XML, AKOMA NTOSO*

1. INTRODUCTION

Until recently, the limitations of the judicial system's ability to implement the goal of proper information management were largely concealed by the "practical obscurity" of a paper-based system of judicial records. While the records in a paper-based system of court records are technically public, all information in the court file receives a considerable amount of protection virtually by the sheer difficulty of accessing it. The introduction of electronic judicial information systems has caused both positive and negative consequences associated with the electronic medium. One of the positive consequences is the ability to make core judicial records available to a far larger audience than ever before at a far lower cost, while the main negative consequence is an issue of the privacy protection [1].

Most commonly used electronic formats for the representation of judgments are Open Office Writer, Microsoft Word and PDF documents. However, unstructured document formats have weaknesses compared to structured document formats since they are not machine readable. There are not many structured standards for the representation of judicial documents. The JuriX language [2] can be used to describe the content of judicial decisions as an XML document written according to the particular syntax. XML Schema Definition of Supreme Court Judgements [3] is an attempt to standardise the content of judgments across courts throughout Australia. Guidotti and Serrotti [4] propose a DTD compatible with the structure proposed in the feasibility plan for the Norma In Rete project to represent the Italian administrative high court decisions. Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontologies (AKOMA NTOSO) [5] recommends technical guidelines for developing and integrating parliamentary information systems throughout Africa. The OASIS Legal XML LegalDocML technical committee [6] works towards advancement of structured standards in the legal field starting from the results on the AKOMA NTOSO project.

Court of Appeals in Novi Sad in [7] prescribes the rules that must be followed when anonymizing judgments in our judicial system. Broadly speaking, those rules can be divided into two categories. Judgements can be anonymized by replacing text or by redacting it. In this paper we propose a method for anonymization of AKOMA NTOSO compliant judgments. The method is based on the access control enforcement for XML documents. The proposed solution relies on the eXtensible XML Role-Based Access Control Framework (XXACF) [8, 9, 10] which is suitable for definition of access control policies and access control enforcement.

The rest of this paper is structured as follows. The next section describes the AKOMA NTOSO document format. The third section gives a brief overview of the XXACF. Application of XXACF for AKOMA NTOSO is given in the forth section. In the conclusion, strengths and weaknesses of this approach are elaborated on and directions of further research are given.

2. AKOMA NTOSO

In order to design a solution that may be used in different judicial information systems we have chosen to use the standard AKOMA NTOSO XML-based format to represent structured textual documents. AKOMA NTOSO was chosen because it separates content, presentation and metadata, it uses XML design patterns, it has found practical applications in several jurisdictions such as EU Parliament, EU Commission, Uruguay, California, US House of Representatives, Switzerland and it has entered OASIS standardization process.

AKOMA NTOSO is the set of principles for electronic parliamentary services in a pan-African context [5]. It has several goals: to define common data exchange standard between parliaments, to specify a basic document model that can be used to build information system and to define simple citation mechanism.

The standard identifies legal documents at FRBR (Functional Requirements for Bibliographic Records) [11] work, expression, manifestation and item levels. It can be used to identify any legal document (legislative, executive or judicial) created by any organization (local, regional,

national or international) at any time (past, present or future).

AKOMA NTOSO defines two types of schemata. The general schema declares the set of elements and constrains that must be satisfied by all AKOMA NTOSO documents. Currently, AKOMA NTOSO supports six different document types: acts, bills, minutes, debate records, judgments and documents (a generic document type). The detailed schema is a schema that declares stricter constrains on the same set of elements. It is used to define concrete document types for a concrete organization.

We used one of the landmark criminal cases in our country's legal history [12] as a case study of applying anonymization rules to judgments. That judgment was made by the Supreme Court, although same method can be applied to other types of judgments.

Since existing format was used, the representation of a judgment was straightforward. Firstly, document was identified at different FRBR levels as an URL according to AKOMA NTOSO guidelines. Then, the important metadata were identified and serialized into the metadata section of the document. At the end, the document structure was marked up using standardized set of elements.

Metadata are organized into several groups. The metadata belonging to the references group (Listing 1) contains the list of locations (*TLCLocation*), persons (*TLCPerson*) and roles (*TLCRole*) referenced from the document and relevant to understanding its content.

```
<references source="#bungeni">
  <TLCLocation id="Ivanjica" href="/ontology/location/wikipedia.ivanjica"
    showAs="Ивањица" />
  <TLCPerson id="DragoljubMihailovic"
    href="/ontology/persons/rs.DragoljubMihailovic.1987-05-06"
    showAs="Драгољуб - Дражa Михаиловић" />
  <TLCRole id="Defendant"
    href="/ontology/role/defendant"
    showAs="Окривљени" />
</references>
```

Listing 1. References group metadata

Each judgment consists of the header section, the body section and the conclusion section. The body section contains an introduction (the summary of the case), background information (the description of the facts), a decision (the decision of the judge) and motivation (the argumentation of the judges). Fragment of the decision section, which is interesting from the anonymization point of view, is described in Listing 2.

The element *decision* is a structural container for the section of a judgment containing the decision. The element *party* is of particular importance since it contains the name of the party that took part in the proceedings (the defendant in this particular case) and a link to the concept in the ontology that specifies it. The content of this element as well as the content of its *refersTo* attribute needs to be anonymized.

Apart from the element *party*, other elements such as *lawyer*, *person*, *organization*, *location* and *date* contain personally identifiable information and therefore need to be anonymized.

```
<decision>
  <p>Са напред изложеног, а на основу цитираних законских прописа и
  <ref id="ref06" href="/rs/act/1945-09-01/66"> Закона о кривичним делима
  против народа и државе</ref> суд их осуђује:</p>
  <ol>
    <li>1) <party id=”par1” refersTo=”#DragoljubMihailovic” as="#Defendant">
    Михаиловића Драгољуба – Дражу</party>, рођеног <date date="1987-05-06">
    27.04.1893. године</date> у <location id="loc1" refersTo="#Ivanjica">
    Ивањици</location>, на казну смрти стрељањем, трајан губитак политичких и
    појединих грађанских права и конфискацију целокупне имовине;</li>
  </ol>
  ...
</decision>
```

Listing 2. Decision section

3. XXACF

In the Role-based Access Control (RBAC) model, access to resources of a system is based on a role of a user in the system [13]. The basic RBAC model comprises the following entities: *users*, *roles* and *permissions*, where permissions are composed of *operations* applied to *objects*. In RBAC, permissions are associated with roles, and users are made members of roles [13]. This greatly simplifies management of access rights, so the RBAC model has generated great interest in the security community. It is customary to use the *role hierarchy* [13] to aggregate permissions, i.e. a role is assumed to inherit the permissions assigned to its parent roles in the hierarchy. In addition, the role hierarchy also determines the roles that are available to a user, i.e. a user assigned to a particular role can also activate any subordinate roles in the hierarchy.

The growth of use of XML as a format for data modelling and interchange accentuates the issue of access control to XML documents. An XML document may contain data with different levels of accessibility.

eXtensible XML Role-Based Access Control Framework (XXACF) [8, 9, 10] provides the means of defining access control policies and access control enforcement based on the RBAC model. Access control policies in XXACF may be defined on different priority and granularity levels and they may be content dependent, thus facilitating efficient management of access control. The access control policy in XXACF may be separately defined for each operation on an XML document. Although the standard RBAC model supports only *granting* policies, research on the XML access control identifies a need for *denying* policies to achieve more efficient security administration. Therefore, XXACF supports both granting and denying access control policies. The concept of context-sensitive access control enables customization of access control policies depending on the environment where XXACF is being used. Therefore, XXACF can be deployed in various environments.

When the user requests to execute a certain operation on the document, the access control enforcement is performed. The process of access control enforcement is performed within five steps:

1. Loading of user properties and user roles (only when the user accesses the system for the first time) - upon accepting the request, user's properties and assigned roles are loaded from the database.
2. Selection of the applicable permissions - the goal of this task is to find access control policies that will be used for access control enforcement on the document being accessed.
3. Marking document nodes – this is a process of determining and applying each permission from the applicable permission set to the nodes of DOM representation of XML document selected by that permission.
4. Conflict resolution - it is possible that on some nodes both policy types are applied; in such cases this conflict needs to be resolved, i.e. to determine whether granting or denying policies will be applied.
5. Execution of the requested operation depends on type of operation:
   - The *read operation* retrieves only parts of the XML document (nodes) that are allowed to be read by the user. If for some segment of the document an anonymization rule is defined it will be replaced by an anonymized data.
   - The *search operation* is performed in the same way as the read operation but, it also removes unauthorised data when counting and displaying the search hits.
   - The *create*, *update* and *delete* operations are allowed if they can be applied on each node being created, updated or deleted.

4. XXACF for AKOMA NTOSO

The resources for which access control is enforced are whole judicial documents or their fragments. All permissions will apply to all documents or their fragments. The proposal has no security requirement for defining access control rules for the particular document, all permissions are defined at document definition level [9].

Resources are identified by XPath expressions which select certain elements in the AKOMA NTOSO documents. For example: XPath expression */akomaNtoso/ judgmentBody/decision* selects the *decision* element (part) of the *judgment*.

Since XXACF is designed to work only with authenticated users, i.e. it must know who is a user that is accessing the documents. Therefore, we have introduced a specific virtual user named *anonymous* that will represent all users who do not need to authenticate. All unauthenticated users will have permissions of the *anonymous* user. The role *Anonymous* is only used to define permissions for the anonymous users that are not required to authenticate.

The anonymization rules in Table 1 anonymize data by calling specified *anonymizer* functions on the fragments of the document selected by the corresponding *pattern.* Anonimization rules are implemented as defined in [7].

The function *to_initials()* convert person's/location's name to initials. The *replace()* function replaces selected fragment with the given text (in this particular case the selected fragment is replaced with the ellipsis). The *party/location* element references the proper *TLCPerson/TLCLocation* element (the attribute *refersTo* of *party/location* has same value as the attribute *id* of *TLCPerson/TLCLocation*). Since the values of those attributes usually correspond to personal names, it is

necessary to anonymize them. It is also necessary to preserve the referential integrity after anonymization. Therefore, the values of those attributes are replaced with their HMAC (Hash-based Message Authentication Code) value [14]. The content of the *date* element has to be replaced with the ellipsis.

Role	*Anonymous*	
Operations	read, search	
Propagation	*direction*: down, *level*: unlimited	
Type	grant	
Resources	/akomaNtoso	
Anonymization rules	*pattern*	*anonymizer*

	//party/text()
	//TLCPerson/@showAs
	//location/text()
	//TLCLocation/@showAs
	//date
	//party/@refersTo
	//TLCPerson/@id
	//location/@refersTo
	//TLCLocation/@id
		to_initials()
		to_initials()
		to_initials()
		to_initials()
		replace('...')
		hmac()
		hmac()
		hmac()
		hmac()

Table 1. Anonymized search and read permissions for *Anonymous*

The *href* attribute of the *TLCPerson* element refers to the proper information about person in the metadata database. Therefore, it is necessary to remove that reference in order to fully anonymize document. The permissions in Table 2 deny access to those attributes for the *Anonymous* role.

Since the permissions in Table 1 grant access and the permissions in Table 2 deny it, the attribute *href* of *TLCPerson* will be assigned both granting and denying permissions. According to the conflict resolution principle "*more specific object takes precedence*" [9] the final permission for this attribute will deny access. The denying permissions in Table 2 are more specific than the granting permissions in Table 1, because they are defined for more specific entity (the permissions in Table 1 are defined for the whole document, while the permissions in Table 2 are defined for the specific attribute), and therefore they are selected as the final permissions.

Listing 3 shows the results of anonymization of the references group metadata shown in Listing 1, while Listing 4 shows the results of anonymization of the decision section shown in Listing 2. The anonymization is performed by applying the proposed method.

Role	*Anonymous*
Operations	read, search
Propagation	*direction*: down, *level*: unlimited
Type	Deny
Resources	//TLCPerson/@href, //*TLCLocation*/@href

Table 2. Deny search and read permissions for *Anonymous*

```
<references source="#bungeni">
  <TLCLocation id="qoRh+MXPEiazsZIWW02iOrLXQIc=" showAs="И." />
  <TLCPerson id="xpkrRA/w3Cts7FT4E1yCBNEI2fA=" showAs="Д.М." />
  <TLCRole id="Defendant" href="/ontology/role/defendant"
    showAs="Окривљени" />
</references>
```

Listing 3. Anonymized references group metadata

```
<decision>
  <p>Са напред изложеног, а на основу цитираних законских прописа и
  <ref id="ref06" href="="/rs/act/1945-09-01/66"> Закона о кривичним делима
  против народа и државе</ref> суд их осуђује:</p>
  <ol>
    <li>1) <party id=”par1” refersTo=”#xpkrRA/w3Cts7FT4E1yCBNEI2fA=”
    as="#Defendant">М.Д.</party>, рођеног … у <location id="loc1"
    refersTo="#qoRh+MXPEiazsZIWW02iOrLXQIc=">И.</location>, на казну смрти
    стрељањем, трајан губитак политичких и појединих грађанских права и
    конфискацију целокупне имовине;</li>
  </ol>
  ...
</decision>
```

Listing 4. Anonymized decision section

6. CONCLUSION

This paper proposes a method for anonymization of judgments represented in the AKOMA NTOSO format using the XXACF framework. The presented solution is not dependent on any particular XML schema for judicial documents and may be used in different judicial environments. The access control solution presented in the paper represents a successful application of concepts of the role-based access control and the XML access control for anonymization of judgments.

The proposed access control solution is designed to be fine-grained in order to provide access control enforcement at different levels of granularity and content and/or context dependent access control. It can be integrated with different legal information systems that use the standard AKOMA NTOSO document format.

Similar methods to the methods used to anonymize judgments can be used to anonymize any other document type that requires anonymization (e.g. archival records). In [15] we described a system for the semantic browsing of legislation. Our intention is to integrate proposed solution into this system. Also, we plan to investigate how anonymization policies can be represented using XACML (eXtensible Access Control Markup Language) [12] and executed using XACML implementations.

REFERENCES

[1] Winn, P. (2009), “Judicial information management in an electronic age: old standards, new challenges”, *Federal Courts Law Review, Forthcoming*, Vol.3 No. 2, pp. 135-176.

[2] Belhissi, R., Moudam, Z., and Chenfour, N. (2011), “JuriX framework for XML Modeling of judicial documents: a support system for checking the regularity of judgments”, *IRACST – Engineering Science and Technology: An International Journal (ESTIJ)*, Vol. 1 No. 1, pp. 51-57.

[3] Kirk, G. and Lazberger, J. (2006) “Proposed XML schema definition of supreme court judgements”, working paper 2.0, Supreme Court of Western Australia, Australia, 2 August.

[4] Guidotti, P and Serrotti, L. (2002), “Legal drafting systems for judges”, available at: http://espejos.unesco.org.uy/simplac2002/Ponencias/Derecho/DER36.rtf (accessed 23 January 2013).

[5] United Nations Department of Economic and Social Affairs - UNDESA (2000), “AKOMA NTOSO”, available at: http://www.akomantoso.org (accessed 23 January 2013).

[6] Organization for the Advancement of Structured Information Standards - OASIS (2012), “OASIS LegalDocumentML (LegalDocML) TC”, available at: http://www.oasis-open.org/committees/legaldocml (accessed 23 January 2013).

[7] Court of Appeals in Novi Sad (2011), “Rules for data anonymizations in judicial decisions”, rule book, Court of Appeals in Novi Sad, January 05.

[8] Sladić, G., Milosavljević, B., and Konjović, Z. (2007), “Extensible access control model for XML document collections”. In *ICETE SECRYPT: Proceedings of the 2nd International Conference on Security and Cryptography, Barcelona, Spain, 2007*, INSTICC, pp. 373–380

[9] Sladić, G., Milosavljević, B., Konjović, Z. and Vidaković, M. (2011), “Access control framework for XML document collections", *Computer Science and Information Systems (ComSIS)*, Vol. 8 No.3, pp. 591-609.

[10] Sladić, G., Milosavljević, B., Surla, D., and Konjović, Z. (2012), “Flexible access control framework for MARC records", *The Electronic Library*, Vol. 30 No. 5, pp. 623-652.

[11] International Federation of Library Associations and Institutions - IFLA (2007), “Functional Requirements for Bibliographic Records, International Federation of Library Associations and Institutions”, available at: http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records (accessed 23 January 2013).

[12] Zečević, M. (2001) “Dokumenta sa suđenja Draži Mihailoviću – Izricanje presude“, http://www.znaci.net/00001/60_3_34.pdf (accessed 23 January 2013).

[13] Ferraiolo, F. D., Ravi Sandhu, R., Gavrila S., Richard, D. K. and Chandramouli, R. (2001), “Proposed NIST standard for role-based access

control", ACM Transactions on Information and System Security, Vol. 4 No. 3, pp. 224–274.
[14] Krawczyk, H., Bellare, M., and Canetti, R. (1997), "HMAC: Keyed-Hashing for Message Authentication", request for comments: 2104, IETF, Network Working Group, February 1997.
[15] Gostojić, S., Milosavljević, B. and Konjović, Z. (2013), "Ontological Model of Legal Norms for Creating and Using Legislation", *Computer Science and Information Systems*, Vol. 10 No. 1 (in print)

ACKNOWLEDGMENTS

Results presented in this paper are partialy funded as the research conducted within the Grant No. III-44010, Ministry of Science and Technological Development of the Republic of Serbia