Acta Cybernetica **18** (4) 579–593.

Improving Content Management – A Semantic Approach

Barna Kovács*

Abstract

Public administration institutions – as well as citizens and businesses – have to meet challenges of the constantly changing business and legal environment. The complexity and quantity of information to be faced with by these actors is increasing at an alarming rate. Research and development projects must turn to the development of innovative, modern technologies which enable citizens and businesses to access, understand and apply complex information easily. Ontology-based content management systems can contribute to the improvement of quality and effectiveness of significant processes, requiring the application of complex information, within the public administration or in a corporation. Compared to traditional content management systems, these systems can support further functions, such as semantic enabled search, explication of relations between documents, drafting of new documents, and version management, as well. Ontologies, in addition to the definition of concepts, support the most detailed and complete exploration of semantic relations between the concepts of a given domain.

Keywords: Ontology-based content management, Semantic-enabled content management, Content management systems, Ontologies, Semantic technologies

1 Introduction

Citizens, businesses, and even public administration institutions have to meet challenges provided by the constantly changing business and legal environment. The complexity and quantity of laws and regulations is increasing at an alarming rate. Consequently, research and development projects must turn to the development of innovative, modern technologies enabling citizens and businesses to access, understand and apply complex legislation and regulations easily. Ontology-based content management systems can contribute to the improvement of quality and efficiency of significant processes of public administration requiring the application of complex laws and other legal sources.

*Corvinus University of Budapest, 1093 Budapest, Fővám tér 8., bkovacs@informatika.uni-corvinus.hu

Content management systems (CMSs) support the creation, management, distribution, publication, and discovery of corporate information. This definition is strongly coupled to knowledge management and has close ties with the management and presentation of information. In contrast with common opinion, content management systems are more than just web content management systems, which are designed to build web and community portals.

Ontology-based content management systems can support further functions, such as semantic enabled search, explication of relations between documents, drafting of new documents, and version management of documents, as well. Ontologies, in addition to the definition of concepts, support the most detailed and complete exploration of semantic relations between the concepts of a given domain. Legal sources, for example, have been coming into existence in many forms and formats, from different jurisdictions, in various languages, with diverse internal structures. An ontology-based approach provides support for capturing patterns in a single framework which is general enough to create a representation of requirements for content management of legal rules codified in multiple jurisdictions. This approach also enables establishing links with legal knowledge systems and legal resources and provides help in maintaining the knowledge base of the CMS as the law is being changed.

2 Content Management

2.1 What is Content?

2.1.1 Digital Content

Providing a precise definition of content is a real challenge. In the literature, there exist numerous definitions for content; definitions differ from each other depending on the authors' focus area. According to Wordnet, the meaning of the word *content* varies from one context to another. Content is 'everything included in a collection'. In the context of messages, subject matter and substance are considered content, i.e. anything related to communication [19]. A definition more adequate to organisational life can be found in [10]: 'The material, including text and images, that constitutes a publication or a document.'

When content is considered in today's computerised organisations, usually digital content is assumed. In this sense, content can be defined as

> 'a commonly used term with regard to the Internet and other electronic media [...]. In its broader sense it refers to material which is of interest to users, such as textual information, images, music and movies, and it generally excludes (1) formatting information, [...], (2) software that is used to provide and render [...] it and (3) unrelated advertising' [16]

The distinction of digital content is necessary, since there are lots of other types of content available in an organisation that are on paper or in other forms which cannot be easily handled by information systems – except when they are digitised.

Information systems deal with the production, processing and retrieving of digital content. The cost of producing digital content is very low compared to the production of any other type of content. This can lead to information overload very soon, which is a relevant problem of today's information sciences. Accordingly, decreasing costs of information creation increase the cost of information processing, requiring human work in most of the cases.

2.1.2 Textual Content

In a public administration environment, in the majority of cases, content takes the form of documents. Documents represent textual data in an unstructured manner, which makes their processing more difficult. Similarly, there are lots of textual data appearing in an organisation, which cannot be considered documents but can be equally important. Communication logs (email or discussion threads) are good examples of such textual data. Currently, other types of content (such as audio, video materials or pictures) can be considered less important in an organisation, except in cases where they are crucial resources.

Computer programs can process textual data, since this format can be indexed and searched easily. Processing of non-textual data, in contrast, is a white area on the map of information processing. Currently only humans can describe the content of certain data items, such as videos, audio materials or pictures. These descriptions can be stored as meta-data. However, producing such meta-data requires tremendous work and still does not ensure reliable results.

2.1.3 How does Content Differ from Regular Data?

As we have seen so far, content and data are related, moreover quite similar terms. Data itself does not have a certain meaning and usually is a broader term than content [15]. However, content itself can contain data, as well as information or knowledge[1], depending on the creator's intention. This way, similarly to data, meta-data can also be assigned to content [14]. Content usually has a specific context, name or title, and other kinds of meta-data associated with it, which can be specified using the Dublin Core standard [7], for example. As another example, consider the meta-data stored by a telecommunication company for every phone call.

2.2 Managing Content

2.2.1 Content Lifecycle

In the current context, content lifecycle and document lifecycle can be considered equal. Usually, four phases of content lifecycle are identified, although some au-

[1] We refer to the system theory approach to data, information and knowledge, as presented by [1]

thors use a different number of stages. Authors usually agree on the first phase (content creation) and on the publishing phase. In between these two there is an editorial process. Furthermore, a final stage can be added, when the document is retired [6]. Consider the following example: one or more authors create a document, which might go through an editorial process of subsequent updates, and then it gets published. In this published form it can also be updated several times before it is finally outdated or retired for some reasons, and becomes a subject of archiving or deletion. Content management systems should support these phases in a collaborative manner, since several people can work together on the same content.

Non-textual content can have the same lifecycle pattern. However, it might be more difficult to create or update the content due to the natural characteristics of non-textual data. Content management systems should support the whole lifecycle in this case as well.

2.2.2 What is a Content Management System?

The goal of content management is to help the users in creating, organising, or, in other words, managing data, information or knowledge represented in the form of content. This also includes the retrieval of information, which is one of the greatest challenges for information systems. Retrieving information means the ability of the system to find data which are relevant for a given user. If data is found to be relevant to the user's problem, then it can be considered as information relevant to the given user's work. By definition data can be considered information only when it is relevant in a given context. This is, however, not constrained to information. Retrieved relevant data can also function as knowledge of the individual or the organisation when it is used [1].

Content is usually handled by Content Management Systems (CMSs). Most definitions focus on the functionality of these systems, just like the following one, stating that a CMS is a:

> 'system for the creation, modification, archiving and removal of information resources from an organised repository. Includes tools for publishing, format management, revision control, indexing, search and retrieval' [3]

CMS vs. WCMS. The term Content Management System is often confused with Web Content Management System (WCMS). They share a common root, moreover both are content management systems, but WCMSs concentrate on managing the content of web portals, while CMSs are managing more general content. So Web Content Management Systems are Content Management Systems with more specific structure and content description formalisms.

CMS Architecture. CMSs usually consist of three major parts: a content creation, a content management and a presentation subsystem, as it can be observed in Figure 1. The content creation part is responsible for managing inputs, in other

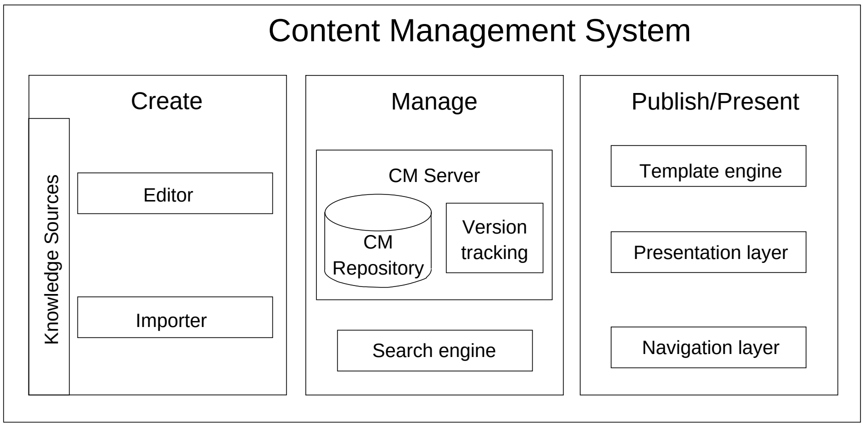

Figure 1: Content Management System Architecture

words gathering content from various knowledge sources. The task of the management part is to store content, together with appropriate versioning information; to ensure access rights management; and to offer information retrieval functions, such as full-text search. The presentation part displays or prints the content using different media types, such as computer screens, mobile devices or printed paper. This part of a CMS has a template engine offering patterns for representing content, adjusting it to the desired use. The presentation layer also handles user navigation on the pages of content – although this function is more relevant in the case of a web content management systems.

Characteristics of CMS. Ideally, the role of a content management system is to be the nerve centre of an enterprise information infrastructure [2, p.7]. It should aggregate data and information from various sources and deliver these to the appropriate recipients. In reality, there are lots of information sources in an organisation, but most of these, such as laptops of individuals or the heads of employees, are not directly connected to the content management system.

In the information system approach, content management is more than data management. For managing pure data, very sophisticated and well-developed tools are available, namely database management systems (DBMSs). However, DBMSs reach their limits quite soon, when used to process content, since DBMSs are designed to manage well-structured data. Content, in contrast, is unstructured with some more-or-less defined meta-data descriptors. This requires an approach completely different from that of data management.

Commoditisation. As a Gartner-study demonstrated in 2002, content management is not a stand-alone product any more [4]. It is integrated with other enterprise information infrastructure components into larger systems, such as Smart Enterprise Suites or corporate portals. Use of CMSs is not a competitive advantage any more, but a necessity. At the same time it is very interesting to see that, in the case of content management, the reason for commoditisation[2] was not emerging technology standards but its simplicity. Technology standardisation is still under evolution. An example is the standard 'JSR-170 Content Repository for Java Technology API' [13], which describes a standard interface for content management to be used in Java-based systems.

2.2.3 Getting the Right Information

Why is content management so relevant? What is the hidden value in content that made the industry so focused on developing content management systems and then later on developing enterprise applications using content? The answer is simple: it adds further value to corporate assets as a very potent source of knowledge.

Content and knowledge have a strong correlation; they complement and overlap each other in many ways [11]. Content is a storage of relevant data and information, which can be used and reused in a corporate problem context.

The question is how to extract knowledge from the vast amount of stored content. Information overload is a common problem in the field of content management. New content can be created in many different forms, additionally numerous distinct resources can be used for this purpose. These pieces of content can be used and reused many times, thus becoming part of the corporate knowledge body.

On the other hand, existing knowledge can be codified. In other words, knowledge can be expressed in an explicit form and put in the organisational memory as content.

Vast amounts of data, information and knowledge are stored as content in a content management system, so the real problem an organisation has to face is how to use them. As mentioned above, decreasing content creation costs cause an increase in the costs of content processing, especially if the latter requires human work. A solution can be the development of content management systems, which provide facilities for effective information retrieval. The keyword is relevance, especially relevance to that business context in which the user is working. The primary challenge in content management is to deliver appropriate content and information to users. Current content management systems provide full-text searching facilities or categorisation for enabling the delivery of relevant information. However, still too many results are provided by these solutions, meaning that the applied methods are not distinctive enough.

3 Semantic-enhanced Content Management

The main purpose of introducing semantics in content management is to retrieve relevant information. Employing a semantic-enhanced content management system (SCMS) can deliver more appropriate information than the currently used methods of classifying information stored in content. As a prerequisite for making a content

[2]Commoditisation is the process by which goods become mass products.

management system semantic-enabled one has to build a domain ontology in the background[3]. This means that the knowledge of a specific domain is processed into a structure representing both the concepts and the relationships between these concepts. Of course, the larger is the ontology, the more content can be described using it.

3.1 Content Processing in a SCMS

Having formalised the knowledge of the domain in the form of an ontology, any content that is fed into the content management system can be annotated with the concepts of the ontology. Concepts can be assigned to the content in general, or parts of the content – for example, by marking words or phrases in the text.

A problem of annotation is that it requires human intelligence in most of the cases. Currently, there are no applications or methods providing a reliable automatic way of revealing relationships between words or phrases in a text and a structured ontology. However, there are promising research activities aiming at this goal.

Annotation is a formalisation of the content, meaning a formal description of what is depicted in the content itself. As soon as the content is annotated, not only the usual descriptive meta-data (such as the meta-data specified by the Dublin Core standard [7]) are known, but, through the structure of the ontology, the meaning of the content becomes known as well. However, it must be stated, that this description relates only to one or some domains – it cannot be assumed that the whole world can be described in a single ontology in finite time.

Figure 2 summarises this process of connecting ontologies and content management systems, using an example in the legal domain. The same process can be applied to other domains as well.

In the following subsections semantic description of content is compared to classical solutions.

3.1.1 Semantic-Enabled CMSs vs. Document Management Systems

The role of Document Management Systems (DMSs) is to handle textual documents and sometimes other forms of documents. They usually treat these documents as black boxes, using only Dublin Core meta-data (such as author, title, creation date, etc.) for their description. Generally DMSs provide indexing and full-text searching facilities as well. In the semantic-enabled content management approach the domain is represented by an ontology and through this representation the content of a document can also be formalised. This means that the system has a certain knowledge of what the document is about, thanks to all those relationships which have already been established between the document and the concepts of the ontology. This information can also be used for inference or searching.

[3]The author presumes a basic knowledge of ontologies. Please refer to [8], [9], [5], [18] and [17] for more information about this topic.

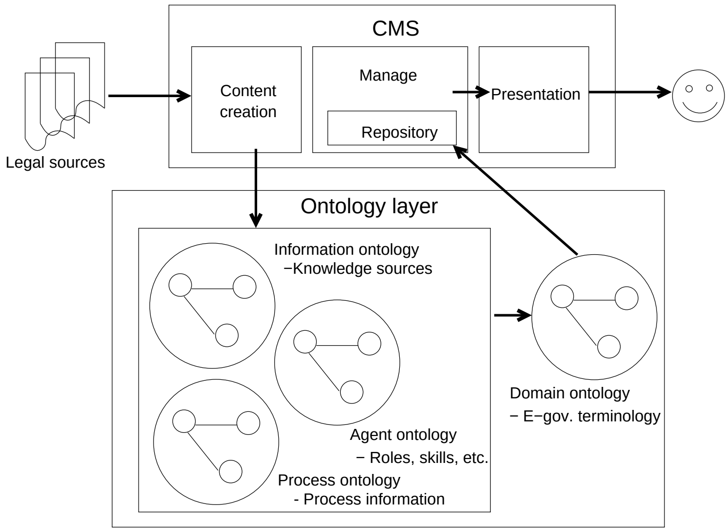

Figure 2: Operation of SCMS in the legal domain

3.1.2 Semantic-Enabled CMSs vs. Full-Text Searching

Full-text searching uses a word index of content as the source of its queries. Ranking of results is determined by different algorithms based on the frequency of words being searched. This is one of the most effective approaches of information retrieval used today. However, it has some drawbacks as well. Certain words and expressions may have one or more synonyms. When searching for a word with synonyms, those contents which only contain the synonyms may not be found. Homonyms are problematic as well, since full-text search cannot distinguish the different meanings of words, making the search results noisy.

Ontology-based content searching means that a concept can be searched independently from its linguistic representation. Actually, word forms or expressions are only instances of a certain concept in the ontology. When a concept is looked up, it will be found regardless of its actual appearance.

This feature of semantic-enabled CMSs can be easily demonstrated by a simple example from the field of Value Added Tax (VAT) regulations. In this context the concept of *natural person* is referred to by the word *customer*. Of course, this word can mean both natural and juristic persons in VAT regulations. When one is looking up content related to the phrase *natural person*, using a semantic-enabled CMS, appropriate parts of the VAT law are expected to be found. This is in spite of the fact that, in the VAT law, the phrase *natural person* is referred to by the expression *customer*.

3.1.3 Semantic-Enabled CMSs vs. Categorisation

Categorisation can be a very good approach to organise and retrieve information. It might have many forms, the simplest of which is the *single categorisation* approach. This is analogous to the folder structure of a hard disk, available on almost every operating system of modern computers, which provides a basic environment for organising files. The problem with this approach is that it uses only a single type of logic in content organisation. It is very hard to find appropriate content if the query follows a different type of logic than that used in the categorisation.

Another approach is multiple categorisation. This uses several terms, and is nothing but the very popular tagging scheme. This approach can ensure quite rich content organisation, however, it can also produce noise in the search results. A drawback is that the person designing the categorisation has to think of all possible aspects of search queries to ensure proper information retrieval. Thus, in most cases, huge amounts of search results are returned, because too many tags are assigned to the content. If sub-categories are used within multiple categorisation, the same problems will arise as in the case of single categorisation.

In the semantic approach the ontology is independent from the content. Thus content is not described by words, but, instead, domain concepts (elements of the ontology) are assigned to parts of the content. When performing information retrieval, the relationships between the concepts in the ontology can be effectively exploited. For example, in some content, such as a document requesting a passport, there is no mention of natural persons. In the corresponding ontology, however, a relationship between a natural person and a passport is defined, namely that only a natural person can request a passport, and a passport can only be assigned to a natural person (both directions can be covered). Even if this relationship is not explicit in any document, the relationship between the natural person and the passport can be discovered and exploited using the domain knowledge formalised in the ontology.

4 Realization in a Pilot Project

In this section we present an application of the principles of semantic-enabled content management systems described above, in the context of a European Union project the author is participating in.

The SAKE project[4] is an ongoing European Union project aiming at the support of knowledge-intensive and formalised processes of a public organisation in a semantic-enabled manner. This system incorporates three major, publicly available

[4]SAKE – Semantic-enabled Agile Knowledge-based eGovernment (IST 027128) is a research project pursued by an international consortium of partners, and co-financed by the 6th EU Framework Programme for Research and Technological Development. The SAKE project commenced on the 1st of March 2006 and lasts for 36 months. The author is leading the development of the Semantic-enabled content management system, one of the most crucial components of the project, which is the responsibility of Corvinus University of Budapest. For more information, please refer to http://www.sake-project.org.

components realizing content, groupware and workflow management functions via integrating existing open-source systems. Additionally, a semantic layer – including an ontology management system and its support tools – has been developed to capture all kinds of semantic information that are provided by components of the system.

4.1 Architecture

From an architectural point of view, the system is built using the classical three-tiered architecture, involving a Presentation Tier, a Business Tier and an Integration Tier, as presented in Figure 3.

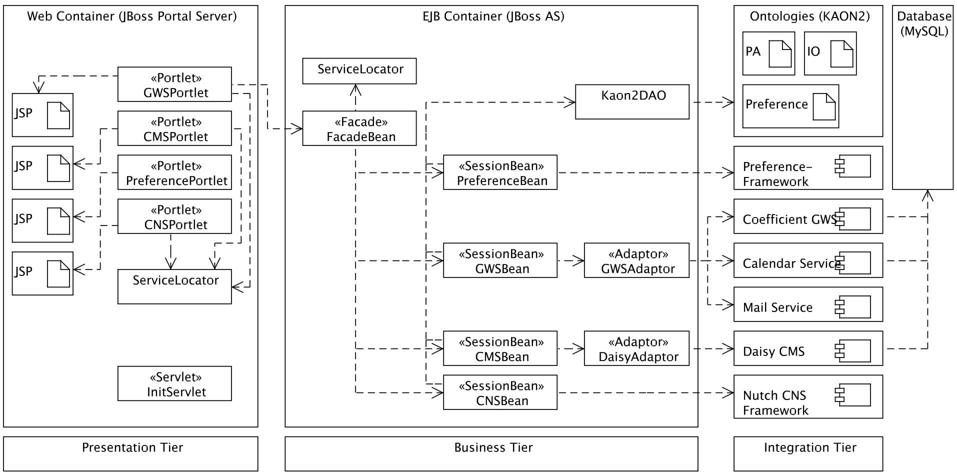

Figure 3: SAKE System Architecture

The *Presentation Tier* is responsible for displaying the content to the user, possibly annotated with additional information. Furthermore, the Presentation Tier handles user interactions using controllers, according to the Model-View-Controller (MVC) paradigm.

The *Business Tier* contains the business logic, realised by the system components mentioned earlier, such as content management or the groupware support modules. A crucial responsibility of the components in this tier is the addition of semantic functionality to the functions provided by the Integration Tier. The Java class FacadeBean provides a unified interface for components in this tier.

The *Integration Tier* consists of various open-source systems on top of which the SAKE System has been built. These include the Daisy content management system; the Coefficient groupware system and its supplementary systems (calendar and mail components); the Preference Framework for handling user-defined preference rules; and the Nutch change notification framework, which is basically a web crawler and a notification system used by the Change Notification System (CNS) component. Integration of these components is eventually realised in the business logic tier, using adaptor components. Adaptors implement the Data Access Object (DAO) design pattern, decoupling the supporting system functions in the Integration Tier from the business logic in the Business Tier. Adaptors provide therefore a homogeneous way of accessing component functionality by applying a translation between specific SAKE APIs and the APIs of components in the Integration Tier. This approach makes it possible to attach to SAKE an arbitrary software system providing the necessary functionality. Thus the SAKE System can be used in existing public administration or corporate environments, serving as a semantic integration layer built on top of legacy systems.

The *Ontology* component, within the Integration Tier, plays a special role as the storage and reasoning facility for semantic information. It contains various ontologies, such as the Public Administration (PA) ontology for capturing organisational information of Public Administration; the Information ontology (IO) representing meta-data of information sources; and the Preference ontology containing preference rules and data. The Ontology component is attached to the business logic components by a special adaptor for the KAON2 reasoner used in the project.

Business components store semantic data and information in the ontologies while storing system-specific data in the supporting systems in the Integration Tier. Business components are realised as Enterprise JavaBeans (EJBs, or beans in Figure 3), all having an appropriate back-end component in the Integration Tier.

The Presentation Tier consists of Java Server Pages (JSP) descriptor pages, constituting the user interface, and portlets, implementing control logic. Portlets govern the page flow of user interfaces and transmit user and context information to the business components. After processing the user request in the Business Tier, results are delivered to the portlets and presented by JSP pages.

4.2 System Operation

During the operation of the SAKE System, semantic aspects of user interactions are captured as presented in Figure 4.

When users work in a formalised process, they use the Workflow component of the system. The flow of activities is defined in the Process Ontology and is utilised by the workflow engine. Every activity involves some user activities in the system: entering data, creating documents, participating in discussions, etc. The first aspect of semantic information captured by the system is the activity that the user is working in. This state is called the ‘Business Context’, in the terminology of the system.

As the processes in the system are defined in an ontology, working on a specific process involves the creation of an appropriate instance of a process class in the Process Ontology. This process instance is recording all the user activities: the values she submitted, references the documents she created, and so on.

On the other hand, the user can decide to leave the formalised environment for some reasons, and use other, more knowledge-intensive functions directly in the system. Users can decide to look for documents in the content management system,

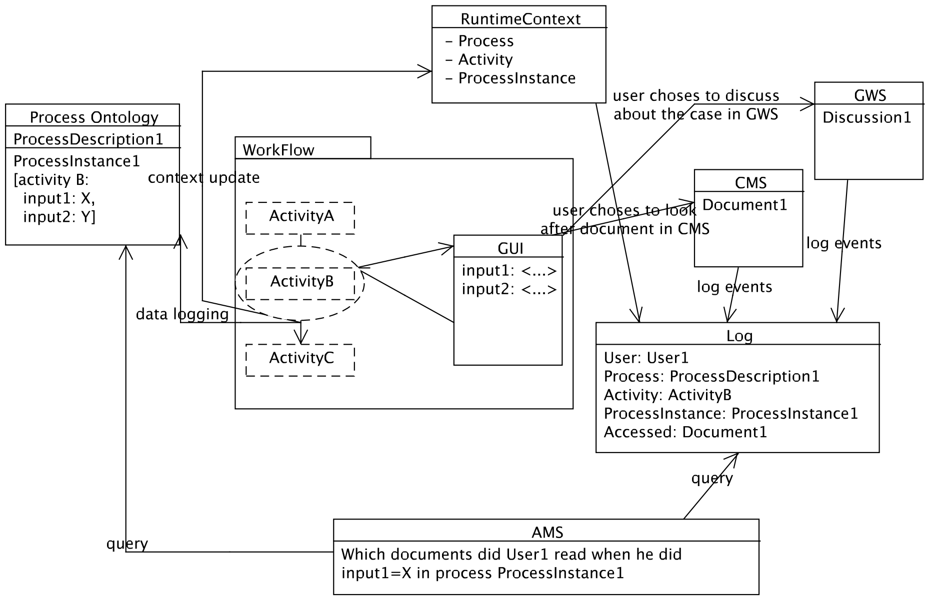

Figure 4: SAKE System Operation

or to discuss a topic in the groupware system. In such cases, the appropriate system component is responsible for recording the user activities.

4.2.1 Capturing Semantics

Various system components record activities by using appropriate ontologies. A basic principle of the system is that everything happening there should be mirrored by the ontologies as well. For example, in the case of Content Management System, when a document is created it is not only stored in the Content Management System in the Integration Tier, but, simultaneously, a new instance of the class Document is created in the Information Ontology. This process, however, provides only a static view as it does not cope with time. The time aspect of user interactions is recorded by the Logging Ontology, by generating new instances of some specific classes of this ontology, such as AccessEvent or CreateEvent, which describe the event. These instances are connected to all other relevant instances in the ontology, such as the Business Context, the User or the Document instances.

These facilities are also used when the user works in the formalised workflow environment, which operates as a kind of marshaller, using various functions of system components. This way all kinds of semantic information are gathered in specific ontologies as the system is running.

4.2.2 Using Semantics

In the previous subsection the way of capturing semantics has been demonstrated. In this process, semantic information about all user activities is stored in various ontologies. However, capturing this information is not enough: it has to be used to satisfy the goals of the system, namely to deliver more precise, more relevant information for its users.

Extracting relevant information of the stored semantic data is the task of the fourth component, the *Attention Management System* (AMS). This component employs a reasoning engine operating on the ontologies. The engine functions by applying user-defined preference rules. These can be either predefined, for example, information about changes in the local regulations, or queried in an ad-hoc way, using complex search expressions. Any data or relationship which can be found in the ontology can be defined in preference rules. This complexity enables the precise description of the user needs.

The reasoning capabilities of the Attention Management System are also used by other components, in the form of related documents or discussions, for example. This component is a crucial element of the system delivering added value to the users. The Attention Management System informs the users about the results of executing predefined preference rules. This is done at user login time, and also in the form of an RSS (Really Simple Syndication) feed, as a notification independent from the SAKE System.

5 Further Research Questions

One of the crucial points of system operation is the proper annotation of the stored documents. Currently, this task is carried out by humans. This is a huge burden on the human resources of the organisation. However, there is no fully reliable method to perform this task using automated systems. In this field, natural language processing and text mining applications show very promising approaches.

The system can be extended to deal with other application fields, involving content types gaining popularity nowadays, such as video or audio clips and pictures. The difficulties encountered in these areas are similar to those appearing in the case of textual documents.

Performance considerations are very exciting and topical problems affecting the usability of the application. Currently, ontology management and reasoning engines are quite slow compared to the well-established database management software, mentioned before. On one hand, this is acceptable since these engines perform complex operations on possibly huge amounts of data. However, the system should remain responsive and fast enough, if it is to be applied by the users in real life situations. Research on improving the speed of reasoning engines shows very promising results, building on techniques learnt from the field of database management, as it is shown, for example, in [12].

6 Conclusions

Current state-of-the-art techniques employed in CMSs are not sufficient enough to handle the the vast amounts of information created and used in an organisation in the course of everyday work. CMSs are not able to effectively manage constantly changing and expanding laws and regulations, which are crucial in public administration.

A major problem is that information in CMSs is stored in an unstructured way. Due to this problem the retrieval of information is also less effective. Moreover general information retrieval algorithms (such as full-text search or categorisation) do not provide results relevant enough.

A solution, as presented in this paper, can be the systematisation and formalisation of domain knowledge. Ontologies provide a formal representation of the given domain, which can be used for mapping content onto the conceptual structure of the domain.

At the same time, building an ontology is far from being a task easy to accomplish. Thus, as a prerequisite for the use of semantic-enabled technologies, large investment of work is needed in the formalisation of a domain. However, the development of adequate ontologies helps solving numerous further tasks. The return on investment is thus realised not only in the field of information retrieval, but also in other areas, such as system and application development, communication, and the operation of the organisation in general.

References

[1] Ackoff, R. L. From data to wisdom. *Journal of Applied Systems Analysis*, 16:3–9, 1989.

[2] APQC. Managing content and knowledge. Report on survey, APQC, 2001. http://www.apqc.org/portal/apqc/ksn?paf_gear_id=contentgearhome &paf_dm=full&pageselect=detail&docid=106067.

[3] Browne, G. and Jermey, J. *Website Indexing (2nd edition)*. Auslib Press, 2007.

[4] Caldwell, F., Gilbert, M., and Hayward, S. CM, portals and collaboration fading: Enter 'smart' suite. Letter from the editor, Gartner, 2002.

[5] Corcho, O., Fernández-López, M., and Gómez-Pérez, A. Methodologies, tools and languages for building ontologies. where is their meeting point? *Data & Knowledge Engineering*, 46:41–64, 2003.

[6] Doyle, B. Seven stages of the CM lifecycle. *EContent*, 2005. http://www.ecmag.net/Articles/ArticleReader.aspx?ArticleID=13554.

[7] Dublin Core Metadata Initiative. ANSI/NISO Z39.85, the Dublin Core metadata element set. Specification, Dublin Core Metadata Initiative, 2007. http://www.niso.org/standards/standard_detail.cfm?std_id=725.

[8] Gomez-Perez, A. Ontological engineering: A state of the art. *Expert Update*, 2(3):38–43, 1999.

[9] Gruber, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

[10] Houghton Mifflin Company. *The American Heritage Dictionary of the English Language*. Houghton Mifflin Company, fourth edition, 2004. Answers.com, http://www.answers.com/topic/content.

[11] Kő, A. and Gábor, A. Content industry and knowledge management. In Remenyi, Dan, editor, *Proceedings of the 3rd European Conference on Knowledge Management (ECKM)*, pages 255–261. Academic Conferences Limited, Reading, UK, 2002.

[12] Lukácsy, G. and Szeredi, P. Efficient description logic reasoning in Prolog, the DLog system. *Submitted to: Theory and Practice of Logic Programming*, 2008.
http://sintagma.szit.bme.hu/lukacsy/publikaciok/dlog_tplp_submission.pdf.

[13] Nuescheler, D. JSR 170: Content Repository for Java Technology API. Specification, Sun Microsystems, 2005. http://jcp.org/en/jsr/detail?id=170.

[14] Prideaux, R. What does a Content Management System do? *TechSoup*, 2005. http://www.icthubknowledgebase.org.uk/whatdoesacmsdo.

[15] Skidmore, D. What is the difference between transactional and content data in an internet packet? In Warren, Mark, editor, *Fourth Australian Institute of Computer Ethics Conference*, pages 51–59, Deakin University, Geelong Victoria Australia, 2005. Waurn Ponds Campus, Geelong, Deakin University, School of Information Systems.

[16] The Linux Information Project. Content definition, 2005.
http://www.linfo.org/content.html.

[17] Uschold, M. Building ontologies: Towards a unified methodology. In *Proceedings of Expert Systems '96*, Cambridge, December 16-18th, 1996.

[18] Uschold, M. and King, M. Towards a methodology for building ontologies. In *IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing*, Montreal, Canada, 1995.

[19] Wordnet. Content, 2007.
http://wordnet.princeton.edu/perl/webwn?s=content.