Collaborative Knowledge Management and Ontologies
The ONTO-LOGGING Platform

Stelios Gerogiannakis, Marios Sintichakis, and Nikos Achilleopoulos

Archetypon S.A., 236, Sygrou Av., 17672, Athens, GR
{sgerogia, mms, nax}@archetypon.gr
http://www.ontologging.com/

Abstract. Corporate memories (stored information and internal processes) in both private and public organizations grow at an exponential rate. This growth is not only quantitative but also qualitative, in the form of increasing interdependencies between processes and information bits. Although the quantitative growth is relatively easy to handle, increasing information complexity is constantly pushing existing information systems to their limits. It is slowly becoming a self-proving fact that organizations will have to transition from the traditional model of searchable/updatable repositories of “facts and figures” to self-organizing, self-adapting corporate knowledge management systems. Ontologies and Semantic Web principles are the most promising relevant technology, now entering their mature age, allowing the creation of extensible vocabularies able to describe any semantic area. Project ONTO-LOGGING is an attempt to harness the full potential of ontologies as a flexible tool of knowledge management within any knowledge-driven organization, such as corporations and public administrations.

1 Introduction

Entering the third decade of the “information age”, more and more private and public organizations are shifting their operational focus towards the service provision sector. Service provider organizations adapt their operational outcome to personalize treatment of individual customers (or citizens, in the case of the public sector). Fixed internal processes are modified in some degree to adjust to each particular case, in the form of dynamic adaptation of employee teamwork. The outcome of each process is not discarded, as it may be re-used in similar future cases. This is what is commonly referred to as working experience.

Increasing, or even maintaining at the same level, an organization’s competence is a never-ending process of combining external stimuli with internally accumulated experience. All the different combinations of data cause the exponential increase of the organization’s corporate memory, in the form of both structured (relational DB entries, spread-sheets) and unstructured (documents, memos, notes, etc.) data.

An interesting fact, about the proportion of structured and unstructured knowledge in an organization is contained in \[1\]. According to the author, almost 80% of an organization’s knowledge is presented and stored in unstructured format, usually free text. What is even more interesting, in the author’s arguments is that, on average, only
20% of organizations’ investments go in the management of that knowledge. This is especially important if we consider that, on average, EU public administrations have spent only 122 Euros per citizen on IT infrastructure [2]. Considering the volume and diversity of knowledge held in public authorities, this figure is a far cry from private IT spending, although it is expected to increase by 6.8% in 2003 [3]. These facts show a lingering threat for the continuing effective operation of organizations, especially public authorities. Only a small part of accumulated knowledge can be immediately accessed and exploited via data processing software, while the larger part of an organization’s corporate memory becomes increasingly chaotic and, hence, harder to locate and extract useful information from.

The rest of this document is organized as follows:

Section 2 provides some background on knowledge management in the organization. Section 3 gives a brief introduction to ontologies and their applications in the area of knowledge management. Section 4 provides a bird’s-eye view of the ONTO-LOGGING platform, while Section 5 delves into a more technical description of the system. Section 6 presents the findings of the project so far in relation to existing technologies and relevant efforts. Finally, Section 7 summarizes the contents of this paper.

2 Knowledge Management

According to one definition, knowledge is

*the fact or condition of being familiar with something (knowing something); this familiarity may have been gained through experience or association.*

Knowledge may be recorded and stored in a variety of ways: in a person’s brain, in corporate processes, in documents or electronic storage systems. In [4], knowledge is categorized in two general forms: *tacit* and *explicit*. The authors argue that, depending on the direction of knowledge transformation, knowledge creation can be categorized in *socialization*, *combination*, *externalization* and *internalization*.

Applying the previous definitions in the area of organizations, we may say that corporate knowledge (a.k.a. *corporate memory*) is an explicit and persistent representation of an organization’s information and knowledge. Corporate knowledge is accessed, shared, re-used and updated by the organization’s employees during the process of completing their assigned tasks [5].

Corporate knowledge requires solid foundations, as it is built upon (and around) the existing infrastructure of an organization. Figure 1 shows graphically the generic “knowledge pyramid” found in any organization, private or public [6].

Corporate knowledge is composed of hard and soft data. The different forms that knowledge comes in, result in its uneven propagation among the different units and members of the organization. Knowledge distribution is the background process of sharing existing and new pieces of information among the organic parts of an organization, with a long-term result of increased productivity. It is interesting that, from the employee point of view, tacit knowledge is equally important with explicit knowledge in the workplace. [4]
A knowledge management system (KM) should be, ideally, equally focused on both knowledge storage and knowledge distribution, preventing, thus, “corruption of knowledge”, by keeping everyone up-to-date [5]. Related studies ([7], [8], [9], [10]), have shown that at an abstract level knowledge management systems are composed of four core processes which are interdependent.

- Generation of knowledge based on new facts and previous work
- Formal organization of collected information
- Development and refinement of existing material
- Distribution of existing knowledge

The task of creating an efficient KM system is still considered an open issue. On one hand, KM systems have to effectively formalize knowledge based on changing organizational needs. On the other hand, tacit knowledge is especially hard to capture and transfer, as organizations are reluctant to shift towards more human-centered practices.

3 Ontologies

As seen in the previous section, knowledge is distributed in many forms/formats and across many locations inside an organization. Ensuring seamless knowledge distribution requires a mechanism of unified description of knowledge sources, using a common vocabulary. Semantic Web principles in general and ontologies in particular are the technology that fits this purpose best.

The most quoted definition of ontologies is that *an ontology is an explicit specification of a conceptualization* ([11], [12], [13]). Ontologies provide formal semantic representation that describes a particular, real-world knowledge domain. Ontologies are the most promising technology for KM systems as they can be used as the back-bone structure, integrating access to diverse knowledge items. The main
advantage comparing to traditional keyword based indexing/searching (relational DBMSs) is that ontologies are common-base, shared descriptions of domains. Therefore, they provide a set of a-priori assumptions about the intended meaning of used terms in communication.

Despite their advantages, ontologies are not the silver bullet for all KM problems. As real-world applications of ontologies have shown, there are two important problems that need to be overcome to harness their full potential. The first problem occurs when trying to model a semantically diverse domain, such as a public authority with different departments, using a single ontology. This approach results in large, monolithic ontologies that have proven difficult to employ for reasons of complexity and maintainability.

This problem has been addressed early on ([14], [15], [11], [12]) by adopting parallel development / operation of multiple, “compact” ontologies, each one focused on a particular knowledge domain of the organization. Mapping techniques between the terms in these ontologies have also been developed [14], providing a semantic link. This technique allows the unified access to diverse knowledge sources without the need to change the ontological commitments of the application or the individual.

This has introduced the second problem in the practical employment of ontologies. As a domain’s needs and business processes change (because of natural evolution and/or administrative decisions), the domain ontology will also have to evolve so as not to be rendered irrelevant. However, having interdependent ontologies makes ontology evolution a complex process, as a modification in one part of the ontology may generate subtle in-consistencies in other parts of the same ontology, in the ontology-based instances as well as in depending ontologies and applications [16].

Although there has been a lot of activity in the semantic web and ontology area, the question of effective evolution strategies in multiple ontology environments is essentially open.

4 The ONTO-LOGGING Platform

Giving an outline of the platform, we may say that ONTO-LOGGING is:

- Motivated by increasing demand in both private and public sector for more effective KM systems,
- Adapting of its functionality based on the importance of documents as sources of knowledge in both corporate culture as well as public sector activities,
- Targeted at existing company infrastructures and technological background,
- Focusing on human-centered practices.

These high-level aims of the project are being realized by employing and extending Semantic Web standards (RDF, OIL, etc.) and combining innovation of open-source software with stability and reliability of commercial platforms.

The ONTO-LOGGING platform is a technological and functional pilot that leads the way for the next generation of commercial KM systems. The outcome of the process will be used as feedback to relevant standardization bodies, especially in the area of ontology management and interoperability.
The main characteristics of the system are

- Scalable, decentralized infrastructure
- Ontology representation and maintenance tools, supporting evolution of multiple ontologies
- Tight integration of the KM with the human resource management system of the organization, to provide personalized functionality

The ONTO-LOGGING system is focusing on all phases and aspects of the knowledge lifecycle in the organization, providing tools for both high-level/administrative and low-level/every-day work.

Beginning from the definition and design phase of a knowledge domain’s ontologies, ONTO-LOGGING provides a number of graphical tools to assist knowledge engineers and domain experts in designing and combining together multiple ontologies.

Going to the aspect of every-day usage, a native application provides end-users with a familiar user interface to query and browse the knowledge repository and locate the desired information. A set of plug-ins for popular backend suites, like MS Office, is also being developed. These tools, being domain ontology-aware, integrate with the rest of the system, enabling one-click classification of documents during the authoring process.

Focusing on the human-aspect of such a system, a complete user-behavioral model has been developed within ONTO-LOGGING. This model allows the personalization of the system in both an explicit and an implicit way. On one hand, the user can explicitly state his/her preferences, while, on the other, usage patterns extracted by the system are used to both refine the domain ontology(ies) and to improve the way domain knowledge is presented to the user.

Finally, a network of collaborating software agents are employed, providing automated intelligent services in different system tasks, acting on behalf of system users. Knowledge propagation in the organization is handled automatically by agents, based on users’ individual profiles. Distributed agents are also used in semantically intensive tasks, as ontology mapping or ontology modification. Agents automate the process by acting collectively on behalf of their respective knowledge engineers to achieve common consensus and reach the optimal solution.

5 ONTO-LOGGING Technical Description

ONTO-LOGGING has adopted a modular, 3-layered architecture for maximum extensibility. The system is divided in the storage layer, the middleware and the presentation modules. Figure 2 gives an outline of the system layers and components.

ONTO-LOGGING is attempting to utilize the best features of both commercial (Microsoft .NET platform) and freeware (Java) platforms.
Namely, the storage layer is based on the popular MS SQL Server, while middleware components are implemented in Java and reside inside the well-proven Jboss application server. The distributed system of collaborating agents has been implemented using the open-source JADE [17] platform. Finally, recognizing the need to provide non-technical end-users with a familiar look & feel, the systems front-end is implemented in the emerging .NET platform. ONTO-LOGGING makes a distinction between expert and normal users, allowing different operations for each type of user.

As seen in Figure 2, the middleware is logically separated in the back-office modules and the integration layer. Back-office modules are:

- **Document management (DMS):** This is a commercial system, used for document storage and retrieval. The current version of ONTO-LOGGING has already been integrated with Meta4 KnowNet and future versions will provide support for other commercial systems, like Lotus Notes.

- **Ontology management (OMS):** It is based on the open-source ontology server KAON [18], providing persistent storage and management for multiple ontologies, accessed by multiple users. Stress tests of this sub-system have shown that it can easily scale up to hundreds of thousands of concepts and instances per ontology.

- **User management (UMS):** This custom module maintains individual user profiles and preferences. It is also in charge of keeping track of user actions and extracting useful information from ontology usage patterns.
The integration layer components are physically placed inside the application server and provide an abstraction of the back-office component functionality. These interfaces are exposed in the form of published web services. This approach has two advantages:

- Ontology and content repositories are represented in a platform-independent way, thus making the system completely decoupled from the presentation layer. This makes possible future integration with third-party applications at almost zero cost.
- Clients need not be aware of proprietary data transportation protocols, as they do not maintain physical data connections.

The integration layer is also host to the agent system’s central containers.

The presentation layer is a suite of different productivity tools, integrated in a single working environment with a uniform look and feel. Presentation modules allow querying of the domain knowledge structure, as well as retrieval of individual knowledge items (docs, notes, etc.). Being a native application, allows the seamless integration of the presentation layer with the user’s office applications, in the form of plug-ins. This capability allows single-click classification and storage of documents in the corporate memory, as they are created. The end-user environment is complemented by a set of dedicated semi-autonomous agents, acting on the user’s behalf for tasks such as reception and filtering of interesting knowledge updates.

Figure 3 shows different components and the processes they live in.

At the heart of the ONTO-LOGGING system, there are a number of ontologies, the structure of which dictates the functionality of the system.

The domain ontology(ies) are custom-made, created by the expert users of the system, using the provided visual modeling tools, capturing the domain knowledge structure. The user ontology models system users and their preferences, while the logging ontology captures user actions and is used by the user management system to infer useful information about system usability and implicit user preferences. The evolution and mapping ontologies allow system experts to apply evolution strategies and maintain mappings between seemingly irrelevant ontologies of different domains.
It is worth pointing out that the actual information pieces (documents) are managed separately from the corresponding ontology instances. Namely, documents are stored and managed by the DMS, while domain ontology instances by the OMS. Although this approach introduces a degree of complexity, it has two important benefits:

- Even if the domain ontology becomes corrupted or unusable for some reason, knowledge bits will still be available through conventional means, provided by the DMS.
- Integrating ONTO-LOGGING with an existing DMS already installed in the organization, or transiting to a new DMS can be done with minimal effort, as the integration layer abstracts away their functionality. In the same way, clustering and failover support can also be introduced.

6 Discussion and Related Work

The problem of effective knowledge management ontology formalization is still an open issue, despite several efforts and advances over the last years. Several research initiatives (OntoKnowledge [19], KnowNet [7] [20], RICA [6]) and standardization efforts (KQML, KIF, XML derivatives) have taken place, but most of these results have not yet been integrated with current commercial products. As a result, almost every new research KM installation seems like a one-off adventure, in which the KM system must be manually customized to the characteristics specific in each organization. This comes in contrast with market evolution dynamics, pushing towards a better integration of emerging formal ontology definition methods within commercial tools, in order to allow for the exchange of existing knowledge, between real-world applications.

ONTO-LOGGING does not re-invent the wheel but, instead, builds on top of foundations laid by various preceding efforts. The importance of ontologies in KM has already been stressed with the best way by OntoKnowledge project results. Issues such as ontology evolution based on changing organization dynamics have also been addressed. Moreover, OntoKnowledge’s findings conclude that a KM system cannot be considered complete without a detailed user-modeling scheme. However, the focus of the effort is more on solution prototyping than integration with existing infrastructures.

Project RICA stresses the need for intelligent agents in the KM process of the organization, by employing semi/fully-autonomous agents for all tasks in the KM lifecycle. Although, this approach obviously has many advantages, it has one major disadvantage: there is almost no ad-hoc integration with legacy systems, as agent technologies are considered interesting for lab-work but still immature to be used “all alone” for time-critical tasks.

KnowNet is the most interesting of existing approaches as it addresses all issues of building an effective KM system: integrating with commercial products on the back-end, detailing user roles and processes in the organization, providing intelligent assistants for end-users. We feel that ONTO-LOGGING builds on top of KnowNet’s achievements, providing better integration with commercial back-end systems, a more robust OMS and wider usage of collaborative intelligent agents, based on the widely-adopted JADE platform.
Focusing in the area of government and public administrations, so far they have been considered as non-areas for the application of KM principles, just because they are “public” [22]. However, this view is changing rapidly. It is now a widespread belief in relevant literature that public administrations are true knowledge-intensive organizations ([21], [22], [23]), with typical examples of ministerial departments, judiciary and regulatory agencies. Indeed, similar to private sector, human actors in the aforementioned authorities cooperate to process existing and produce new information, whether this is legislative actions or service provision to citizens. Some interesting efforts, like smartGOV ([24], [25]) and eGOV ([26], [27]), show the way of things to come. The next generation of public administration information systems, shall provide the domain experts with the tools necessary to create dynamic, one-stop government portals for all aspects of a citizen’s life. This will not be a trivial task as the amount of information found in public authorities is not only vast but also complex. Creating such an infrastructure cannot be a one-shot transition; instead, it will require a complete back-office infrastructure and well-organized, up-to-date domain knowledge, things that only a KM system can offer.

Of course, one cannot address issues and needs of the public sector by simply applying the same practices that have been proven successful in private organizations, as correctly argued in [28]. This is simply because not only the requirements are different but also because the mentality and culture are different. This is why usually simply porting successful systems from the private sector to public organizations usually fails or does not produce the expected results. The area of KM is no exception to this rule.

However, even if end-goals and internal processes change across domains, the technologies and the underlying infrastructures can still be re-used with minimal transition cost. We do not claim that ONTO-LOGGING will be the panacea for all of KM-related “ills” and issues of public organizations. We believe, though, that efforts like ONTO-LOGGING will smooth the way for the proliferation of e-Government services, acting in the back-office, enhancing productivity of public service employees. They will provide a solid foundation on which decision-makers and end-users (public servants) will be able to cooperate and transition more swiftly to the new technology era.

For example, Figure 4 shows how introducing KM systems, like ONTO-LOGGING, can transform processes in public organizations for the better. So far, editing a document involves creating/editing the document and either storing it in a public folder and notifying the interested parties or forwarding the document itself. This is not at all efficient because the creator of the document has to do additional tasks plus there is always the chance of notifying the wrong people (or not notifying the right people). However, in the case of ONTO-LOGGING, user preferences and semantic metadata allow the system to determine who is interested in what and act accordingly. The same Semantic Web principles also allow for more effective knowledge search and retrieval.

Of course, all this may sound greatly familiar, resembling existing document management systems. ONTO-LOGGING’s difference compared to existing systems is that it employs the full potential of Semantic Web in the area of KM. We believe this transition will eventually lead to the creation of more flexible systems capturing the true logic of the knowledge domain, able to adapt their internal logic to the changing environment instead of offering a set of fixed options.
7 Conclusions and Summary

The increasingly complex problem of organization knowledge management is far from solved, as it is not a linear problem, both in quantitative and qualitative terms. We believe that a solution for this problem will be two-fold. On one hand, make the organization accumulate new knowledge quickly and efficiently and on the other hand, aid end-users in their day-to-day tasks by preventing them from becoming knowledge-flooded or knowledge-starved.

ONTO-LOGGING is well placed in the area of KM, not only as a research effort, but also as an immediate ancestor of next-generation KM systems. These will combine the best from open-source efforts and commercial platforms, being based on open standards. ONTO-LOGGING is still an undergoing effort and is currently under evaluation deployment in test sites, under real operating conditions, at Indra Sistemas premises.

This paper provided an introduction to knowledge management in the enterprise and the most promising technology for this area, ontologies. Project ONTO-LOGGING is a re-search effort to harness the full potential of ontologies in KM systems, providing tools for both expert and non-technical users.

Acknowledgements. The ONTO-LOGGING project \[29\] has commenced on July 2001 and is co-funded by the European Commission under contract IST-2000-28293. Its duration is 2.5 years. The authors would like to acknowledge the assistance and co-operation between all partners. The partners of the ONTO-LOGGING consortium are: Indra Sistemas (administrative and financial coordinator), META4 (scientific coordinator), FZI, Deltatec, CALT-INSEAD and Archetypon S.A.

References

15. Maedche A: Measuring Similarities between Ontologies, Presentation at the 13th European Conference on Knowledge Engineering and Knowledge Management EKAW, Madrid, Spain, 2002