Information Sciences 179 (2009) 2643–2655

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A transformational model for Organizational Memory Systems management with privacy concerns

Sergio F. Ochoa, Valeria Herskovic, Edgard Pineda, José A. Pino *

Department of Computer Science, Universidad de Chile, Chile

ARTICLE INFO

Article history:
Received 9 August 2007
Received in revised form 12 May 2008

Keywords:
Organizational Memory Systems
Information privacy
Documents transformations
Information retrieval
Decision support systems

ABSTRACT

Collaborative activities such as coordination, decision-making and negotiation critically depend on historical information of an organization. This information is usually part of isolated legacy information systems, therefore, it can be inconsistent, redundant and difficult to retrieve and link. Previous research in CSCW has proposed the use of Organizational Memory Systems (OMS) to accumulate, organize, preserve, link and share diverse information coming from various sources, and thus support such collaborative activities. However, there is a need to provide a low-cost feeding process, to embed privacy mechanisms and to support information retrieval capabilities for all users of the OMS, in order to make these solutions useful to a broad range of organizations. As a way to deal with this need, this paper presents a transformational model able to: (a) facilitate the feeding of an OMS based on information stored in legacy information systems, (b) ease the information retrieval process, and (c) embed automatic mechanisms to evolve the information stored in the OMS, through a document privacy lifecycle. This is a low-cost solution that can be implemented using OpenSource technologies.

1. Introduction

The importance of storing and managing knowledge in an organization stems from its contribution to production and efficiency, support for decision-making, negotiation and monitoring, and as source of organizational culture for novice employees [17,30]. Typically, this knowledge is based on the information stored in legacy information systems, which usually have been developed in an isolated way. Therefore, such information can be inconsistent, redundant and difficult to retrieve and link [6].

As a way to deal with this problem, previous inquiries in the CSCW area have proposed an Organizational Memory System (OMS) as a facility for accumulating, organizing, preserving, linking and sharing knowledge [12]. An OMS also supports organizations to ensure organizational learning, increase efficiency and flexibility, overcome growth limits and integrate organizations after mergers [23]. Conklin defines two types of knowledge that can be stored in an OMS: formal and informal. The first one is the final product of knowledge work, in the form of documents, designs and procedures. Informal knowledge, on the other hand, is ephemeral and hard to capture, created in the process of producing the formal results [12].

Information stored in the OMS must be easily retrievable and readily available, so that users can access information when they need it. However, feeding the OMS from legacy information systems is not easy. Typically, this feeding process is done either by batch processes or by manual ad-hoc procedures. Neither of them is a good solution: batch processes are slow and error-prone, while filtering and structuring large quantities of information is expensive and requires

* Corresponding author.
E-mail addresses: sochoa@dcc.uchile.cl (S.F. Ochoa), vherskov@dcc.uchile.cl (V. Herskovic), epineda@dcc.uchile.cl (E. Pineda), jpino@dcc.uchile.cl (J.A. Pino).

0020-0255/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2009.01.041

much effort. The information that ends up in the OMS has a poor structure (e.g., PDF documents), which makes the OMS unmanageable and chaotic, limiting the possibility to deal with other OMS requirements, such as information privacy and fast and flexible retrieval methods. Gallupe has identified knowledge acquisition and storage as one of the open issues in this area [16].

Documents in the OMS can include sensitive information, such as ideas, discussions generated in a computer-mediated meeting and decisions that affect people. Recent works in the area of knowledge management have recognized the need to improve the access control to personal information linked to organizational information [14]. This is particularly true for network organizations [27]. If an organization does not protect the sensitive information in their OMS, then negative consequences may occur [18]. First, users might abstain from participating, either by not voicing their opinions or by not being completely honest. In turn, low quality and small quantity of generated ideas imply poor discussions and thus, decision-making becomes negatively affected. Also, users might choose not to share their information if they feel their privacy may be threatened. Finally, lack of privacy may even bring undesirable consequences to the users themselves, if what they assume to be a private system is being monitored. Therefore, it is required the OMS embeds a privacy policy to protect all users and thus stimulate their participation. The National Institute of Standards and Technology (NIST) has highlighted this issue and it has proposed several guidelines to deal with it [20].

However, the cost of including privacy mechanisms in an OMS can be too expensive for small and medium-size organizations, and the implementation process could demand an important effort in the manual processing of the information to be protected. These two issues show the need for a solution not only for large organizations but also for small and medium-size ones.

Considering the previously stated needs, this paper presents a transformational model intended to (a) ease the feeding of an OMS based on information stored in legacy information systems, so that information inside the OMS can be structured and classified correctly, (b) enhance the information retrieval process to allow for easy and fast retrieval and (c) provide an automatic privacy mechanism for information stored in the OMS. Such information will follow a privacy lifecycle based on transformations, which deals with the need to improve the access control to personal information embedded in the OMS [14]. This is a novel and low-cost approach, which has been designed to be useful for any organization size.

The rest of the paper is organized as follows. Section 2 presents the related work. Section 3 describes the work scenario and the privacy requirements involved in an OMS implementation and evolution. Section 4 presents the document lifecycle from a privacy viewpoint. Section 5 describes the proposed transformational model and its main components. Section 6 presents a discussion and the conclusions.

2. Related work

Structuring information before storing it in an OMS eases information processing afterwards, e.g., information retrieval and privacy mechanisms implementation. Unstructured information, on the other hand, is difficult to manage and support. This section presents related work with structuring (feeding process), retrieving and implementing privacy measures on information stored in an OMS.

2.1. OMS structure and information retrieval

The problem of structuring information to be stored in an OMS can be solved by automatic, semi-automatic or manual means. Manually discarding irrelevant items and adding metadata to structure the information will produce the best results. However, this process is expensive and slow.

Several authors have sought to resolve the automatic classification of information using AI [10,13,22]. For instance, Chen et al. use AI techniques such as automatic indexing, cluster analysis and Hopfield net classification, to discover the topics underlying comments made by users [10]. These topics are used for structuring and classification, concept-based management and information retrieval. Experimental results show the resulting classifications are not as good as those of a good human classifier, which suggests that AI techniques might not be ready to be the only strategy to solve this problem.

An example of semi-automatic structuring is Yaka [5], a knowledge management system based on the concept of *subjects*. A subject is a collection of views on information sources. When a user submits a document to the OMS, associating a subject to it, Yaka detects the item and automatically extracts meta-information, such as its keywords and summary. Users can subscribe to a subject or search the information repository to find relevant information.

An important OMS functionality that benefits from information being structured is information retrieval (IR). Qureshi et al. describe several knowledge management activities [25]. One of the activities, called knowledge organization, establishes relationships and context for collected knowledge. This structuring permits the next stage, which is delivery of information, to be efficient and effective. Romero et al. use a thesaurus to enhance each query and use index terms and an inverted index to locate relevant documents efficiently [26]. An important feature of their solution is that the results are presented with links to other parts of the discussion, permitting the user to navigate through a network of related items. In the case of semi-structured data stored in an OMS, a versatile and powerful query language must be used in IR to take advantage of the structure. XQuery is an example of a popular and convenient way for querying semi structured data, broadly applicable across many types of XML datasources, and allowing complex queries [7].