ELSEVIER

Knowledge-Based Systems 13 (2000) 315–320

Knowledge-Based SYSTEMS

www.elsevier.com/locate/knosys

Short Communication

The role of artificial intelligence technologies in the implementation of People-Finder knowledge management systems

I. Becerra-Fernandez*

Florida International University, Decision Sciences and Information Systems, Miami, FL 33199, USA

Received 11 July 2000; accepted 31 July 2000

Abstract

The development of knowledge management systems (KMS) demands that knowledge be obtained, shared and regulated by individuals and knowledge-sharing organizational systems, such as knowledge repositories. People-Finder systems, a type of knowledge repository, attempt to manage knowledge by pointing to experts possessing specific knowledge within an organization. Details about such systems implemented at several organizations such as Hewlett–Packard, National Security Agency and Microsoft are presented. Insights, challenges and future development plans gained through the development of a People-Finder are discussed. Finally, concluding remarks about the role of artificial intelligence in the development of People-Finder KMS and automating the process of profile maintenance are discussed.

Keywords: Knowledge management; Knowledge management systems; People-Finder systems

1. Introduction to knowledge management systems

Knowledge management systems (KMS) have been defined as "an emerging line of systems (which) target professional and managerial activities by focusing on creating, gathering, organizing and disseminating an organization's 'knowledge' as opposed to 'information' or 'data'" [1]. The development of KMS demands that knowledge be obtained, produced, shared, regulated and leveraged by a steady conglomeration of individuals, processes, information technology applications and a knowledge-sharing organizational culture. It has been observed that KMS currently underway at most organizations fall into three categories [2]:

1. Educational KMS. To elicit and catalog tacit knowledge, and at the same time serve as an educational tool.
2. Problem-solving KMS. Organizations with significant intellectual capital require eliciting and capturing knowledge for reuse in solving new problems as well as recurring old problems.
3. Knowledge repositories. The majority of the KMS in place. Knowledge repositories themselves fall into three categories. The first category attempts to catalog organizational knowledge that exists in explicit form, for example a system to store marketing-oriented documents. A second category attempts to develop databases of employees' insights and observations, for example, discussion databases or lessons-learned systems. And finally, the third category is known as knowledge yellow pages or People-Finder systems, are repositories that attempt to manage knowledge by holding pointers to experts who possess specific knowledge within an organization.

This paper presents a survey of People-Finder KMS in several organizations. This paper also discusses insights and lessons learned from the development of a People-Finder KMS: the Searchable Answer Generating Environment (SAGE). Finally, it presents future plans for this research as well as a vision for the role of artificial intelligence in People-Finder KMS.

2. A survey of People-Finder systems

Several organizations in different business categories have identified the need to develop systems to help locate intellectual capital, or People-Finder KMS. The intent when developing these systems is to catalog knowledge competencies, including information not typically captured by human resources systems, in a way that could later be queried across the organization. A literary review of

* Tel.: +1-305-348-3476; fax: +1-305-348-4126.
E-mail address: becferi@fiu.edu (I. Becerra-Fernandez).

0950-7051/00/$ - see front matter
PII: S0950-7051(00)00091-5

Table 1
Summary of characteristics of People-Finder KMS in place at HP, NSA, Microsoft and BA&H

Characteristics	CONNEX (HP)	KSMS (NSA)	SpuD (Microsoft)	KOI (BA&H)
Access	Intranet	Intranet	Intranet	Intranet
Taxonomies	US Library, INSPEC own	Department of Labor (O*NET)	Own	Own
Self-assessment	Yes	Yes, supervisors also participate in data gathering	No, supervisors rate employee's performance	Yes
Participation	Voluntary, only those willing to share	Whole personnel	Whole personnel in the IT group	Only consultants
Levels of competencies	No	Yes	Yes	No
Data maintenance	User (nagging)	User and supervisor	Supervisor	User and system administrator
Company culture	Sharing, open	Technology, expertise	Technology, open	Personality driven
Purpose of the system	To share knowledge, for consulting and to search for experts	To staff projects and match positions with skills	To compile the knowledge and competency of each employee	To staff projects and to get consultants to work in a collaborative environment
Platform	HP-9000 Unix, Sybase, Verity	OS/2, VMS Bourne shell	SQL, MS Access	Oracle, DEC Alpha, Sun SPARCstation

hallmark People-Finder KMS was implemented, and follow-up personal interviews with the developers of such systems were held. Here we discuss some of the details about these People-Finder KMS, and we provide a set of references in case the reader is interested in further details.

Hewlett–Packard (HP), a company competing in the market of computers, peripheral equipment and other electronic equipment developed CONNEX (http://www.carrozza.com/connex), a People-Finder KMS (T. Carrozza, phone interview and follow-up e-mail with developer of CONNEX at HP Labs, September 16, 1999), [4]. The goal of the project was to build a network of experts, available online, to provide a guide to human knowledge within HP. CONNEX consists of a centralized database of user knowledge profiles, with a Web browser interface that allows users to find profiles in multiple ways. User's profiles contain a summary of their knowledge and skills, affiliations, education and interests, as well as contact information. CONNEX users can easily find experts within HP by searching the database by any combination of profile fields or by browsing through the different areas of knowledge, geographies and/or names. To support a large user base with high volume of transactions, CONNEX was built using Sybase database and Verity's Topic search engine, on an HP platform.

The National Security Agency (NSA) has also taken a step towards the implementation of a system to locate experts [10]. The NSA is part of the "Intelligence Community", and their two missions are Foreign Signals Intelligence and National Information System Security. The goal of the implementation of the knowledge and skills management system (KSMS), a People-Finder KMS, is to catalog the talent pool within the agency to allow the precise identification of knowledge and skills, and to take advantage of information technology. The NSA went through the development of the system by applying "database engineering" in order to solve the complexities of implementing an adequate, workable and successful KMS. They also divided the execution of this project into several "Work Tasks" and developed a knowledge taxonomy applicable to their workforce.

The goal of Microsoft's "Skills Planning und (and) Development", known as "SpuD", was to develop a database containing job profiles available online across the IT group, and to help match employee's competency with jobs and work teams [3]. The following are the five major components of the SPuD project [5]: development of a structure of competency types and levels; defining the competencies required for particular jobs; supervisors rating the employee's performance in particular jobs; implementing the knowledge competencies in an online system; and linkage of the competency models to learning offerings. The system was created based on a Structured Query Language (SQL) server as the database for MS Access, as well as a Web front end. Note that the validation of the data in this model rests with the supervisor, who essentially assigns the competency criteria to each of the employees under his/her supervision.

Bozz, Allen & Hamilton (BA&H) is a leading management and technology-consulting firm. By 1996 BA&H had already implemented one of the first global, web-based People-Finder KMS to support its knowledge management program, called Knowledge On-Line (KOL) [9], (L. Remeikis, phone interview, September 3, 1999). One of the key goals of BA&H KOL was to get consultants to work more collaboratively. Another goal for KOL's development was that a high percentage of the required architecture be supported by what the company already had in place, including the Netscape browser. In fact, KOL is based on a centralized, many-to-one client/server architecture, an Oracle database (1.5 gigabits) running on a DEC Alpha server, and a Sun SPARCstation acting as the front-end Web server.

Table 1 summarizes some of the major characteristics of the People-Finder KMS in place at HP, NSA, Microsoft and BA&H.

3. The SAGE KMS

The NASA/Florida Minority Institution Entrepreneurial Partnership (FMIEP) grant is funding the development of the SAGE, which is in the category of People-Finder KMS [3]. The purpose of this KMS is to create a repository of experts in the State of Florida (FL) State University System (SUS). Previous studies have pointed out that there is a void in the ability to identify the capabilities in the FL SUS [6]. As NASA-Kennedy Space Center (KSC) looks to develop new technologies necessary for the continuation of their space exploration missions, their need to partner with Florida SUS experts becomes evident.

Currently, each State University in Florida keeps a database of funded research, but these databases are disparate and dissimilar. The SAGE KMS creates a single repository by incorporating a distributed database scheme, which can be searched by a variety of fields, including research topic, investigator name, funding agency, or university. The SAGE system combines the unified data warehouse by masking multiple databases as if they were one. One advantage of this method is that there is no need to reconfigure the data to fit it into one template. This methodology provides flexibility to the users and the database administrator, regardless of the type of program used to collect the information at the source. Although the project SAGE is specific in nature, what was desired was to develop tools and techniques that would make managing these independent databases as seamless as possible. One of SAGE's advantages is that there is only one user point of entry at the web-enabled interface, allowing multiple occurrences of the interface and giving the end user deployment flexibility. The main interfaces developed on the query engine use text fields to search the processed data for keywords, fields of

expertise, names, or other applicable search fields. The application processes the end user's query and returns the pertinent information.

The purpose of the SAGE KMS is to unify myriad data collections into one data warehouse that can easily be mined. SAGE gives university researchers more visibility, and at the same time allows interested parties to identify available expertise within the SUS. This application helps to identify a researcher's expertise within a discipline, and to facilitate communication through a point of contact.

4. The technologies to implement SAGE

The development of SAGE was marked by two design requirements: the need to validate the data used to identify the experts, and at the same time minimize the impact of each of the universities' offices of sponsored research, who collect most of the required data. For this reason, we opted for taking the data structure in its native form and making necessary data cleansing at the SAGE server site. SAGE's strength also rests in its ability to search for experts by using a set of parameters instead of only a proper name. Yet another strength lies in the fact that it validates the data at the source and is built upon a searching criterion that is recognized as a valid indicator of expertise, that is, funded-research grants received. The reasoning behind the selection of experts is based on the actual participation of the researcher in activities related with a particular area of research.

Although a number of database systems exist on the World Wide Web which claim to help find people with a defined profile, most of these tools rely on people to self-assess their skill against a predefined set of keywords. Self-assessment is inherently unreliable, and the results could be biased and hard to normalize. On the other hand, while a number of search engines are available on the web, the entity seeking for an expert has to use a combination of different tools in order to find the appropriate information. With SAGE, all the information is easily accessible due to the versatility of its searching options, which allows one to refine the search until the degree of accuracy required is attained.

SAGE is built upon the integration of the following technologies:

1. Cold Fusion™ — An off-the-shelf rapid and integrated development environment.
2. Open Database Connectivity (ODBC) — allows middle-ware to interface with the database. This is important since it makes the code independent from the database platform used.
3. Verify's Search 97 — used to perform the keyword search within the field that contains the grant abstract. It also allows the use of logic operators, which enhances the power of the search engine.

The development of the SAGE People-Finder involved an initial design followed by incremental implementation phases. The initial design phase constituted a comprehensive survey of available tools and methodologies, followed by the selections of the most efficient approach to data storage and data retrieval. The two methodologies considered for data access were Common Gateway Interface (CGI) and middle-ware. At the time SAGE was designed, CGI presented disadvantages for most development environments except for very large systems. A user connecting to a CGI script would start another instance of the script, that is, another process for that user. This means that if 100 users connected to the CGI script, 100 processes started on the system, not including additional processes spawned by some of the built-in commands in the script. In essence, this disadvantage of using CGI in terms of memory requirements outweighed the advantage of using CGI, which allows using well-known programming languages like C, C++ or Perl. Cold Fusion was chosen as the middle-ware development environment because of its significant application strength and its demonstrated database interaction capabilities. Additionally, Cold Fusion provided the ability for secure transactions with Secure Socket Layer (SSL) technology, with the potential for strong encryption.

The implementation phase of SAGE involved the design of Cold Fusion modules, each with an assigned task. A robust server environment was set up along with a server operating system with remote user capabilities. Querying modules were used to provide search capabilities. Each query interface involved several modules that interacted with the database that connects to the Cold Fusion code through ODBC. A relational database model was used because of its efficiency and flexibility with data. The database contains entities representing each of the main attributes of the grant information, including project name, researcher, agency and university among others. In order to design the SAGE database it was required to prepare an analysis of each of the databases from the different universities, thus the final database model supports a comprehensive data-set. Following the design process, the next step consisted of populating the database. Subsequently, the next step consisted of alpha testing the interface both locally and remotely. The evaluation phase of SAGE included heavy end user testing and processing time optimization. SAGE has been online since August 16, 1999 at http://sage.fiu.edu.

SAGE receives approximately 20 hits per day, of which about four are new visitors. About 54% of the total hits come from commercial sites (.com and.net), 39% from educational institutions (.edu) and 2% from government and military organizations (.gov and.mil). Some of the commercial sites are companies developing advanced search engines or companies involved in the development of KMS. Our visitors come mostly from the US, but about 5% come from around the world, including countries such

as Japan, France, Austria, Switzerland, Bahamas, Mexico and the UK.

As mentioned before, one of the technical challenges faced during the design and implementation of this project was the fact that the source databases of funded research from the various universities were dissimilar in design and file format. The manipulation of the source data was one of the most important issues we had to deal with, since the credibility of the system would ultimately depend on the consistency and accuracy of the information. One of the most important research contributions of SAGE is the merging of inter-organizational database systems through the use of correspondence tables, which function much like array pointers, and allow compliance to differing database formats. Through the correspondence tables, we resolve finding similarities between database fields, no matter how different their original formats are. Manipulating the data included the process of cleansing the data, followed by the data transformation into the relational model, and ultimately the databases migration to a consistent format (in this case SQL Server 7.0).

Future developments for SAGE include the development of algorithms that will facilitate the maintenance of SAGE in a more automatic fashion. The main idea behind the design of the software is to make the data maintenance process as human independent as possible. We anticipate that these algorithms will include a daemon in each of the universities' severs, working according to a pre-scheduled transfer rule. This daemon will obtain the data from the universities' databases and transfer the information to the SAGE server, making the process human independent. After the information is in the SAGE server, the next steps will involve the migration of the data to the SQL Server format, followed by cleansing and transforming the data to a relational format. Currently, the software that will perform this process is being developed on basis of the experiential knowledge gained during the development of SAGE. The software will also include a data dictionary that will be augmented based on new information, and it's being designed to be interactive. This particular quality will be helpful in dealing with cases that were not considered during the software design phase.

5. Challenges in the implementation of People-Finder KMS

Previous research conducted to establish the parameters to design People-Finder KMS has demonstrated that one of the challenges in developing these systems is related to the inherent shortcoming of self-assessment. Most of the People-Finder KMS surveyed rely on each employee to complete a self-assessment of competency, which is later used when searching for specific knowledge areas. The issue of a self-assessment is one that offers both advantages and disadvantages. The advantage of self-assessment is that it allows building a repository of organization-wide competencies quickly. The disadvantage is that the results of self-assessment are subjective, based on each person's self-perception, the results could be hard to normalize, and employees' speculation about its possible use could 'skew' the results. Furthermore, one People-Finder implementation required supervisors to ratify their subordinates' self-perceptions and assign a quantifiable value to it, a requirement that many organizations would find too taxing on their supervisors.

Another challenge in developing People-Finder KMS deals with the development of knowledge taxonomies. Taxonomy is the study of the general principles of scientific classification. Knowledge taxonomies allow organizing knowledge or competency areas in the organization, and their use in People-Finder KMS is currently being researched in the context of using ontologies for information integration. In the case of People-Finder systems, the taxonomy is used to describe and catalog people's knowledge, an important design consideration. Furthermore, knowledge taxonomies could be critical in the People-Finder system's success. A deep analysis of the People-Finder KMS in place reveals that many attempts to create knowledge taxonomies are unsuccessful (Remeikis, phone interview, 1999) or sub-optimal (Carrozza, phone interview and follow-up e-mail, 1999).

6. Concluding remarks on the future role of artificial intelligence in People-Finder KMS

Future developments for People-Finder systems such as SAGE include the development and integration of artificial intelligence (AI) technologies to enhance the capabilities of these systems. For example, data mining could enhance the process of updating profiles by mining the authors of documents in an electronic repository and identifying a correspondence with the topic of the document. Authors of documents in an electronic repository are experts in those knowledge areas; therefore, the profile of the contributors to the repository could be automatically updated with keywords related to the subject matter contribution. This data mining effort would result in a diminished reliance on self-assessment.

Furthermore, a data mining effort could be instrumental in the clustering of similar data objects together. For example, the data in SAGE is organized by grant, and indexed by the principal investigator field. Through the use of a clustering tool [7], data can be grouped into clusters of expertise, to reveal expertise areas that may not be currently defined. In SAGE, keywords are not explicitly defined for each grant, and the way SAGE works is by performing an embedded search into the title and abstract of the proposal data in order to search for user-specified keywords. Our basic assumption is that a grant related to the field of 'cryogenics' is expected

to have that word in the grant title or abstract. The implementation of the clustering technology will create a domain dictionary that will serve to increase the semantic domain of the keyword. In the aforementioned example, the domain dictionary will be useful in identifying related keywords to the word "cryogenics", such that related keywords may be defined to the grants at the investigator level. In this fashion, relationships that may not be necessarily obvious may be identified also — a sort of "fuzzy matching." The resulting "pseudo-keywords" may be saved for future re-use.

Another application of this clustering notion is the development of a "super" concept, which would allow to group experts together, developing a group-level of expertise. Given the individual areas of expertise, these could be clustered together into groups of expertise or virtual "centers of excellence". In the case of SAGE, grouping of researchers with compatible areas of research from universities in the FL SUS would result in virtual "centers of excellence". For organizational People-Finder KMS, grouping of experts with complementing expertise areas would result in virtual "centers of excellence" in the organization. This effort could reveal areas of strength that may well otherwise go unnoticed in the organization. Additional developments in this area will be instrumental in the development of organizational training programs, that may be designed to address the gap between what "is known" and what "needs-to-be-known" in the organization.

In conclusion, our vision of People-Finder KMS fits well with the work to develop systems that seek to create an IT support environment for knowledge workers. This is done through the use of intelligent assistants in a business process environment; keeping in mind that "an IT tool may only act as a facilitator for sharing, creating or retrieving knowledge, but never as a key player in creating, evaluating or contributing knowledge" [8].

Acknowledgements

The author wishes to acknowledge NASA-KSC, the Technological Research and Development Authority and Bethune-Cookman College, for financial support in the development of SAGE, under the auspices of the "Florida Minority Institution Entrepreneurial Partnership Grant", grant number NAGO-0220. The author also wishes to acknowledge James Jennings, Dr Gale Allen and Gregg Buckingham, all of NASA-KSC, who championed this initiative. Finally, the author wishes to acknowledge the contributions of the students who work in the FIU Knowledge Management Lab.

References

[1] M. Alavi, D. Leidner, Knowledge management systems: issues, challenges, and benefits, Communications of the Association for Information Systems 1 (1999) 7.

[2] I. Becerra-Fernandez, This is rocket science: The Knowledge Repository of Kennedy Space Center. Presentation at the Delphi's International Knowledge Management Summit (IKMS'99), San Diego, CA, March, 1999.

[3] I. Becerra-Fernandez, Searchable Answer Generating Environment (SAGE): A Knowledge Management System for Searching for Experts in Florida. Proceedings of the 12th Annual International Florida Artificial Intelligence Research Symposium (FLAIRS), Knowledge Management Track, Orlando, Florida, May, 1999.

[4] T. Davenport, Knowledge Management at Hewlett–Packard. Available online at: http://www.bus.utexas.edu/kman/hpcase.htm, 1996.

[5] T. Davenport, Knowledge Management Case Study: Knoweldge Management at Microsoft. Available online at: http:/kman.bus.utexas.edu/kman/microsoft.htm, 1997.

[6] T. Kotnour, Presentation at the Partners in Education Conference, Cocoa Beach, FL, October, 1998.

[7] M. Mehrotra, S. Alvarado, W. Wainwright, Laying a Foundation for Software Engineering of Knowledge Bases in Spacecraft Ground Systems. Proceedings of the 12th Annual International Florida Artificial Intelligence Research Symposium (FLAIRS), Orlando, FL, May, 1999.

[8] H. Schurr, S. Sttab, R. Studer, Ontology-based Process Support. Proceedings of the AAAI Workshop on Exploring Synergies of Knowledge Management and Case-Based Reasoning, Orlando, FL, July, 1999.

[9] C. Tristram, Common Knowledge, CIO. Available online at www.cio.com/archive/webbusiness/090198_booz.html, September, 1998.

[10] A. Wright, W. Spencer, The National Security Agency (NSA) Networked Knowledge and Skills Management System. Presentation at Delphi's International Knowledge Management Summit (IKMS), San Diego, CA, March, 1999.