

Classification of Enterprise Portals: A Data Mining Approach

Ahmed A. Elragal and Heba George Abouseif

German University in Cairo (GUC), New Cairo City, Egypt

Abstract

Classification of enterprise portal systems (EP) based on their features is the topic of this paper. We propose a classification based on cluster analysis which depends on features collected from the internet. Results showed that systems were found to belong to different classes based on their features.

Keywords: enterprise portals, classification, cluster analysis.

Introduction

This paper aspires to classify enterprise portals (EP). It outlines and uses Data Mining (DM) techniques to cluster EP solutions, also known as Knowledge Management (KM) tools, based on their features and capabilities, in order to help managers acquire or deploy the KM tools to choose the most appropriate tool for their organization.

From the mid-1980's onwards, many individuals and organizations began to realize, as well as, appreciate the increasingly important role of knowledge in the emerging turbulent and very competitive environments. Bechina Arntzen and Ndlela (2009) argue that no business can escape this economic turmoil untouched, and in order to cope with and overcome challenges which organizations have to face, from the rapid technological advances, to shortened product lifecycles and high market volatility, organizations need to be ready and able to manage their "highly distributed diversified knowledge".

Considered a tool for KM, EP solutions play an important role in deciding the viability and success of organizations nowadays.

In the following sections, the now burgeoning literature on EP is reviewed, where the features of EP solutions, their possible potentials as well as the advantages they bring to organizations that deploy them are discussed.

Next, the literature is reviewed in the field of DM. DM techniques are said to discover any hidden patterns, identify relationships between data variables if any, and also predict useful information from massive data sets.

This is followed by the methodology section in which a number of enterprise portals are identified and clustered according to their functions and features. The findings of this study will then be recorded and may be used by project managers to help evaluate the capabilities of portals.

Enterprise Portals (EP)

The following section discusses the concept of enterprise portals (EP), and their potential benefits, capabilities and usefulness to organizations that acquire them. Moreover, commonly known features of EP solutions are identified.

Defining EP

Enterprise Portals (EP), which are also referred to as Enterprise Information Portals (EIP) or Corporate Portals, are a form of information and communication technology (ICT) that support KM initiatives and approaches (Maier 2002). The terms Employees Portals, Customer Portals, Enterprise Intranet Portals, Business-to-Employees Portals and Business-to-Employees Systems are also commonly used to refer to the category of portals, and where each of these terms may have a different target user (Benbya, Passiante and Belbaly 2004; Hazra 2002). In addition, the terms Knowledge Portal and enterprise computerized Knowledge Management System (KMS) are sometimes used (Kotorov and Hsu 2001; White 2000). The term Enterprise Portals (EP) will be used regularly throughout this paper.

EP solutions not only support KM initiatives, but may also include business analytic applications and Business Intelligence (BI) capabilities. In effect, this forms a type of integration between KM and BI features. According to Guran (2008), EP solutions provide access to BI tools. On the hand, Hazra (2002) and Maier (2002) stated that BI tools are a form of KM. However, the authors of this paper believe that BI and KM are two different entities, although serving the same goal of decision support. In Section 5, a number of EP solutions will be examined and a cluster analysis will be performed to identify their separate BI and KM functionalities.

There has been extensive use of metaphors in the literature in an attempt to fully map out the potential and technological roots and influences of EP. On one hand, Maier (2002) believes that EP solutions play the

role of the coil that produces a magnetic field. That is to say, they act as a magnetic center and any organizational data or information that passes through them is "magnetized", integrated, reinterpreted, rearranged, recombined and managed to better serve the organization's goals and objectives. Kotorov and Hsu (2001) presented another metaphor in which EP solutions were thought to be the brain of the organization that provides employees with the necessary information for success in the ever-changing and very competitive marketplace. Another depiction is proposed by Cloete and Snyman (2003) in which EP solutions are said to act as an antidote to the problems of the information age, namely the "infoglut" and "infofamine". The first problem is that of information overload, as organizations nowadays are becoming more and more information-intensive, whereas the second problem refers to the lack of knowledge, reiterating the idea that only a small proportion of organizational knowledge is actually captured and is made available for everyone to benefit from.

Enterprise portals can be defined as single-point-of-access Web browser interfaces used primarily within organizations to support the capturing, organization, aggregation, sharing, management, and dissemination of information and knowledge assets of different structure and format that are stored in disparate sources, web file servers, and databases across the enterprise. Moreover, they provide organizations with a shared workspace that facilitates access to organizational communities and group collaborations, and promote organizational learning as well as the gradual development of organizational memory (Benbya, Passiante and Belbaly 2004; Detlor 2000; Raol, Koong, Liu and Yu 2003). Following the same view of EP technology, Detlor (2000) introduces three major components that constitute the shared information workspace namely; a content space (which provides access to a wide variety of corporate information and resources), coordination space (which handles workflows and routines to support cooperative work), and a communication

3 Communications of the IBIMA

space (which offers channels for negotiation and conversation to ensure shared interpretation of information made available. It also fosters the development and storage of new ideas for re-use in the future). As the name suggests, EP software is said to provide a "single point of entry, a single point of access, and a single point of information and knowledge interchange" (Hazra 2002).

A more technical definition of EP solutions and their approach views them as an enterprise-wide integration of business applications to the Web, in order to reap the benefits of the Internet and offer organizations the opportunity in order to form knowledge-sharing networks of employees, clients, partners and vendors. From a comprehensive viewpoint, EP present a means of joining together all the different computer technologies that cover the corporate landscape into a single system that helps its users find information unmindful of its physical location. In doing so, they provide a transparent directory of information which already exists elsewhere and do not act as a separate source of information (Detlor 2000). They more importantly promote collaboration and the notion of shared knowledge and understanding of the business across functional units.

As indicated by Remus (2007), the strength of EP solutions is that they hold the promise of providing secure, real-time, customizable, integrated and more importantly personalized and tailored access for an organization's employees as well as its customers and business partners to dynamic content from various sources, and in a range of different formats. From an operational perspective, they provide resources, applications and processes to authorized users in a neatly managed single screen or system. Also, and from a functional standpoint, they offer individual users or classes of users the opportunity to view and interact with the invaluable set of corporate digital resources anytime and catered to

their job functions, roles, or other criteria (Guran 2008).

EP Features

Guran (2008) and Bowman (2002) provided a number of common features for EP namely, content management, security management, metadata management, knowledge mapping, knowledge directories, search engine capability, text search and retrieval, standing queries, customization, personalization, simple user interface design, web-enabled, collaboration tools, affinity group filtering (to filter relevant information for specific users or user groups), performance management, tools for developing and implementing plans, and finally a gateway to enterprise applications, DW, data mining and extraction tools and other computing resources. Cloete and Snyman (2003) presented a similar analysis.

Hazra (2002) outlined a similar set of features but from a different perspective. These are: security to ensure access control and protection of organizational information, maintaining access logs to detect any violation or breach of policy, reliability to provide fail-over and crash recovery of mission-critical business processes, high availability to handle user access 24/7, scalability in order to have room for the ever increasing business functions and requirements, keyword search capability to allow navigation and retrieval of information, reporting tools, BI capabilities, friendly graphical user interface to capture the needs of users, customization, personalization to ensure that information is tailored according to preferences of the users, and finally, collaboration to support communities that allow interaction between users and the sharing of information.

Using a different perspective, Benbya et al. (2004) classified the key features of EP solutions in three categories namely; core capabilities, supportive capabilities and web services. Core capabilities include taxonomy (also called classification or categorization

schemes), publishing, search, personalization, integration and collaboration. Supportive capabilities consist of security, profiling and scalability. The most comprehensive and detailed classification of EP features is that presented by Raol et al. (2003). This classification of features and sub-features is listed in Table 1. Furthermore, it will be used in Section 5 (Methodology) to compare various EP solutions and perform a cluster analysis to group the solutions based on the availability of the mentioned sub-features. The main reason for such choice is the technical nature of the information included and its suitability for the application under question.

Data Mining (DM)

Wang and Wang (2008) refer to data mining as a powerful BI tool for knowledge discovery. Data mining is also defined as the automated extraction of hidden patterns and predictive information from massive amounts of data (2005). It involves digging deep into the data in an attempt to unravel and find any unknown relationships and associations between data variables. Nguyen, Tjoa, and Trujillo (2005) also identify DM practices as a step in the Knowledge Discovery in Databases process (KDD). KDD deals with the discovery of useful, interesting and previously unknown knowledge in databases. DM initiatives play an important role in generating an inventory of patterns and trends in the data, which are then fed into the following steps of the KDD process where they are carefully analyzed and transformed into valuable knowledge to users.

Nguyen et al. (2005) identify some commonly utilized DM tools and techniques namely; artificial neural networks (ANN), decision trees, genetic algorithms, nearest neighborhood, and rule induction. They also add that data mining initiatives deal with data exploration, pattern recognition, and time series databases.

Wang and Wang (2008) argue that some people overemphasize the power and capabilities of data mining initiatives, and may

falsely perceive that data mining tools help organizations acquire knowledge from computers and databases through the push of a button. They assert that overlooking the role of user interaction with the DM tools and technologies underlies such misconceptions. They also present a typical DM cycle which involves four phases: identify the business problem, transforming data into actionable results, acting on or applying the information, and finally measuring the results. Adding to this, the authors argue that knowledge gained from DM may not necessarily result in actions. Here, the idea of users' role in developing their own knowledge from DM outcomes and transforming this knowledge into business actions is revisited.

According to Wang and Wang (2008), and reiterating the notion of user interaction, there are two groups of knowledge workers involved in the DM process namely, data miners and business insiders. This is because no knowledge worker can do both jobs. On one hand, a data miner is usually an expert of DM and completely understands the DM techniques. A data miner must be aware of the nature of the business in order to relate the DM results to the business and its context. On the other hand, a business insider (who may be a Chief Executive Officer (CEO) or middle level manager) holds the best knowledge regarding business problem solving and decision making. Also, a business insider whose main objective is to develop the organization's business performance must have sufficient understanding of the BI, KM and DM concepts, but is not required to fully digest the DM techniques and procedures. DM results are of use and can be applied effectively and successfully when the mentioned user groups (data miners and business insiders) join forces and integrate their efforts.

The Problem

Statistics show that organizations worldwide are spending billions of dollars in acquiring the wrong systems or technologies. This paper helps in solving this

problem by classifying EP based on their features and hence, makes the technology acquisition decision a little bit easier. The problem the paper solves is “how to classify EP based on their features to facilitate technology acquisition decisions?”

Methodology

In this section, a cluster analysis technique will be adapted in order to classify EP solutions using Teradata Warehouse Miner (TWM). The Teradata product will be used because it is available and, in addition, it provides cluster analysis which is the basis of this study. Cluster analysis is based on unsupervised learning, where data has no target attribute (or where there are no predefined groups). In essence, cluster analysis encompasses algorithms for identifying homogeneous groups of data objects, where elements of the same cluster share common attributes. This is required to differentiate between the various EP solutions under question. Six EP solutions were chosen in this study for convenience, as their features were accessible.

The data gathering stage of this study (cluster analysis of EP solutions) involved two main steps. A detailed description of each of these steps is outlined in the following:

- Step 1: Identification of EP solutions that support the features mentioned in Section 2.
- Step 2: Identification and classification of the features contained in the products selected. This also includes identifying the sub-features available in each product.

The EP solutions chosen were: IBM WebSphere Portal, Microsoft SharePoint, Oracle Portal 11g, Oracle PeopleSoft Enterprise Portal, Oracle WebCenter Suite and finally, SAP NetWeaver. Information about the products was obtained from the Internet as well as through direct communication with vendors. Data on the

surveyed features of each product have been coded and stored in TWM. Table 1 has the features and sub-feature details.

Findings

Performing a K-Means clustering analysis, the six enterprise portals examined in this study were grouped in two clusters. Results are shown in Table 2, in which Cluster 1 (C1) contains four portals (SAP NetWeaver Portal, Oracle PeopleSoft Enterprise Portal, Oracle Portal 11g and Oracle WebCenter Suite) and Cluster 2 (C2) contains the other two portals (IBM WebSphere Portal and Microsoft SharePoint).

Clustering of the EP solutions was done based on the fifty sub-features mentioned earlier. Out of these fifty sub-features, forty-four were the same for both clusters. That is to say, either the six products offered this feature or they all did not supply it. The six differences between the two clusters are shown in Table 3. The differences were in the following sub-features and features: “sort order” (customization and personalization), “secure search” results (proactive/search), “security mirroring” (secure/security), “develop and execute plans” (dynamic feature), “KML rendering” (extensibility/embedded applications), and finally “open gadget standards” (extensibility/embedded applications).

By examining Figure 1, the following can be noted: 'Develop and execute plans' is offered by all C1 members and only one out of the two C2 members. 'KML rendering' is only offered by one C2 member. 'Open gadget standards' is provided by two C1 members and both C2 members. Similarly, two C1 members provide 'secure search' and both C2 members offer it. For 'security mirroring', the only portal that provides it is a C1 member. One out of four C1 members provides 'sort order' and only one member of C2 provides it.

Table 1: Features and Sub-Features of EP Solutions

Features	Sub-features								
Customization and Personalization	Personalization	Selecting Resources	Sort Order	Change in colors and layouts	Customization	e-Room Digital Workplace			
Proactive / Search	Integrated Search Capability	Embedded Search Engine	Secure Search Results	Search Collection Replication					
Collaboration and Community	Community	Task Management	Document Collaboration	Threaded Discussion					
Secure / Security	Authentication	Open Security Architecture	LDAP Integration	Access Control	Security Mirroring	Invitations	Auditing	Object Level Access Control	Full SSL Support
Dynamic	Search by Category	Publish Information	Subscribe to new content	Query and analyze Information	Develop and Execute Plans				
Extensibility / Embedded Applications	Open Gadget Standards	KML Rendering	Caching System	Gadget Factory	Repository Search Gadgets	Internet Access			
Content Management	Document Directory	Publications	Content Filtering and Routing	Metadata Management System	Content Approval	Crawler			
Scalability / Network	Intelligent Caching	Load balancing across multiple servers	Pooled Connections	Performance-enhancing techniques					
Administrative Tools	Web-based Administration	Job Scheduling	Window Administrative Client	Image Server					
Easy to Use	Portal is Web-enabled	Uses familiar graphical user interface							

Table 2 shows the details of the products and the six main differences in their features. The reason behind the outcome of the clustering is the similarities between the members of each cluster. In C1, all members are identical in two features namely; develop and execute plans and KML rendering. C2, the members are identical in:

secure search results, security mirroring, and open gadget standards. It is important to note that clustering does not indicate ranking or preference; it just groups similar items together. One cluster can be superior in a certain aspect, but the other can be superior in a different aspect. The whole process facilitates the exercise of comparison between different product.

Table 2: Clusters of Portals

	Cluster 1	Cluster 2
SAP Net Weaver	Y	
Oracle PeopleSoft EP	Y	
Oracle Portal 11g	Y	
Oracle WebCenter Suite	Y	
IBM WebSphere		Y
MS SharePoint		Y

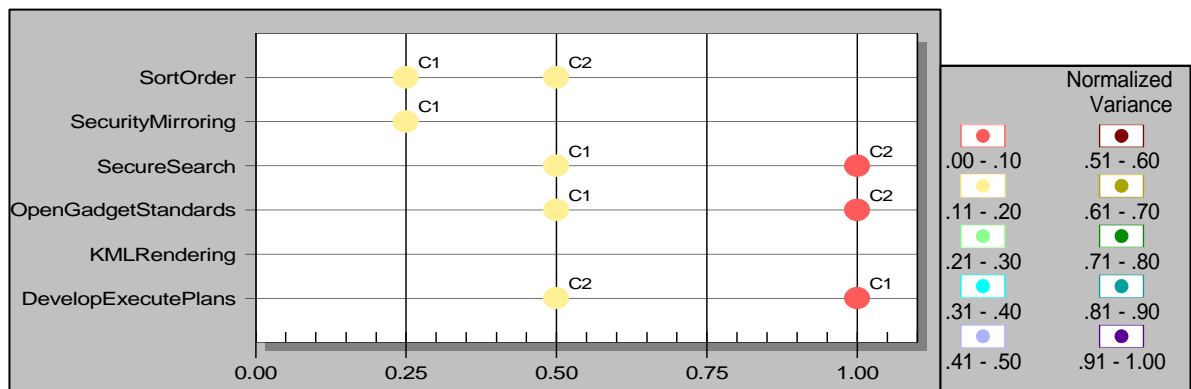


Fig 1. The Six Main Differences between the Two Clusters

Table 3: Differences among Various portals

	SAP NetWeaver	Oracle PeopleSoft EP	Oracle Portal 11g	Oracle WebCenter Suite	IBM WebSphere	MS SharePoint
Sort Order	Y	N	N	N	N	Y
Secure Search Results	N	N	Y	Y	Y	Y
Security Mirroring	N	N	N	Y	N	N
Develop / Execute Plans	Y	Y	Y	Y	Y	N
Open Gadget Standards	Y	N	N	Y	Y	Y
KML Rendering	N	N	N	N	N	Y

Limitations

A limitation in this study is the credibility of the data relevant to the EP solutions. This study is based on the reported features of the products. However, a lab test for such features may have obtained different results. A more accurate cluster analysis should rely on lab-tested products rather than reported functionalities. This involves implementing the different EP solutions and testing, comparing and evaluating their different features.

Moreover, this study was conducted for six portals only, which were available to the authors, but a larger number of portals could have been clustered in much more sets (more than two) and would have resulted in more detailed information and classification.

Finally, all the features were weighted 0 or 1, but a relative weight from 0 to 1 could have given more useful information. In other words, the features and sub-features were considered as existent or non-existent, but their real value or strength on a scale from 0 to 1 was not supplied. This, of course, limits the clustering results.

Conclusion

A clustering analysis was performed on six enterprise portals using Teradata Warehouse Miner. Similar portals were identified based on their common features. The outcome of the clustering may be useful to managers in considering different products. Limitations of this study were also discussed, and should be used in future work and EP analysis.

References

- Bechina, A. A. A. & Ndlela, M. N. (2009). "Success Factors in Implementing Knowledge Based Systems," *Electronic Journal of Knowledge Management* 7, (2). 211-218. [Online], [Retrieved November 30, 2009], http://www.ejkm.com/volume-7/v7-2/Bechina_Arntzen_and_Ndlela.pdf
- Benbya, H., Passiante, G. & Belbaly, N. A. (2004). "Corporate Portal: A Tool for Knowledge Management Synchronization," *International Journal of Information Management* 24, 201-220.
- Bowman, B. J. (2002). "Building Knowledge Management Systems," *Information Systems Management* 19, (3). 32-40.
- Cloete, M. & Snyman, R. (2003). "The Enterprise Portal-Is It Knowledge Management?," *Aslib Proceedings* 55, (4). 234-242.
- Detlor, B., (2000). "The Corporate Portal as Information Infrastructure: Towards a Framework for Portal Design," *International Journal of Information Management* 20, 91-101.
- Guran, M., (2008). "Knowledge Management Using Intranets and Enterprise Portals," *International Journal of Computers, Communications and Control* 3, 75-81.
- Hazra, T. K. (2002). "Building Enterprise Portals: Principles to Practice" *Proceedings of the 24th International Conference on Software Engineering (ICSE)*. 19-25 May 2002, Orlando, Florida, 623-633.
- Kotorov, R. & Hsu, E. (2001). "A Model for Enterprise Portal Management," *Journal of Knowledge Management* 5, (1). 86-93.
- Maier, R. (2002). "State-of-Practice of Knowledge Management Systems: Results of an Empirical Study," *Journal of the Swiss Computer Society, (Informatik / Informatique - Knowledge Management)* 3, (1). 14-22.
- Nguyen, T. M., Tjoa, A. M. & Trujillo, J. (2005). Data Warehousing and Knowledge Discovery: A Chronological View of Research Challenges, *Proceedings of the 7th*

International Conference on *Data Warehousing and Knowledge Discovery* (DaWaK). 22-26 August 2005, Denmark, 530-535.

Raol, J. M., Koong, K. S., Liu, L. C. & Yu, C. S. (2003). "An Identification and Classification of Enterprise Portal Functions and Features," *Industrial Management and Data Systems* 103, (9). 693-702.

Remus, U. (2007). "Critical Success Factors for Implementing Enterprise Portals – A Comparison with ERP Implementations," *Business Process Management Journal* 13, (4). 538-552.

Wang, H. & Wang, S. (2008). "A Knowledge Management Approach to Data Mining Process for Business Intelligence," *Industrial Management and Data Systems* 108, (5). 622-634.

White, M. (2000). "Corporate Portals: Realizing their Promise, Avoiding Costly Failure," *Business Information Review* 17, (4). 177-184.