CalSWIM: A Wiki–Based Data Sharing Platform

Yasser Ganjisaffar, Sara Javanmardi, Stanley Grant,
and Cristina Videira Lopes

University Of California, Irvine, USA
{yganjisa,sjavanma,sbgrant,lopes}@uci.edu

Abstract. Organizations increasingly create massive internal digital data repositories and are looking for technical advances in managing, exchanging and integrating explicit knowledge. While most of the enabling technologies for knowledge management have been used around for several years, the ability to cost effective data sharing, integration and analysis into a cohesive infrastructure evaded organizations until the advent of Web 2.0 applications. In this paper, we discuss our investigations into using a Wiki as a web–based interactive knowledge management system, which is integrated with some features for easy data access, data integration and analysis. Using the enhanced wiki, it possible to make organizational knowledge sustainable, expandable, outreaching and continually up–to–date. The wiki is currently under use as California Sustainable Watershed Information Manager. We evaluate our work according to the requirements of knowledge management systems. The result shows that our solution satisfies more requirements compared to other tools.

Keywords: Knowledge Management, Wikis, Data–Sharing.

1 Introduction

In today's increasingly unstable and competitive environment, knowledge is widely considered the main source of competitive advantage for organizations. As a consequence, organizations have become progressively more concerned with the concept of knowledge management, since nurturing organizational assets potentially could enable long-term competitive advantage [1,2].

Traditionally, knowledge management systems (KMSs) are controlled by a small number of individuals and any non-trivial changes to the system require intervention of software experts. Only users who feel more comfortable with web technologies are more likely to engage in electronic knowledge exchanges [3]. The problem gets worse when organizations have some publicly available data, but there is no simple interface for data sharing or any tool for data aggregation and analysis. To access the data, researchers or professionals usually need to download the data, which is not always the most updated data, and use their own analytical tools for data aggregation and analysis. In this way, it is not easy to share and discuss the results with others.

With the emergence of web 2.0 tools, such as wikis, mashups, blogs and RSS feeds, theorists have explored these social and collaborative web tools as knowledge management enablers. They believe that web 2.0 tools specifically wikis are undemanding and affordable tools for communication and knowledge management, and they have the potential to offer the features of complex and expensive IT-solutions [4]. One of the key characteristics of the wiki software is that it allows very low cost collective content creation through a regular web browser and a simple markup language. This feature makes wiki software a popular choice for content creation projects where minimizing overhead in creating new or editing and accessing already existing content is of high priority. Although Openness, participation and decentralization in wikis have made wikis suitable targets for effective knowledge management [5], not all of the requirements of KMSs can be satisfied through the current wiki engines – specifically when user needs to access organizational data for data integration and analysis purposes.

In this work, we show how a wiki can be used in organizations as a web-based interactive KMS in order to make their knowledge *sustainable*, *expandable*, *outreaching*, and *continually up-to-date*. We describe our investigations into enhancing an open source wiki called XWiki [1] with advanced knowledge management features, including databases, GIS layers, and other forms of digital data. In this way it is possible to (a) fetch and process historical and real-time data from different sources such as local or remote databases; (b) provide data analysis tools integrated with the wiki, which permit the seamless merging, analysis and presentation of the data; (c) share and discuss the reports and the results of analysis. The enhanced wiki is currently used as California Sustainable Watershed Information Manager (CalSWIM) [2]. We evaluated our solution with other solutions to knowledge management in organizations according to the KMS requirements [6]. The results show that our solution satisfies more requirements compared to other knowledge management tools.

The remainder of this paper is as follows: Section 2 summarizes some of the relevant related work. Section 3 illustrates CALSWIM wiki and the extended features. In section 4, we describe the evaluation of the work according to the knowledge management requirements. Finally, section 5 draws some conclusions and future research work.

2   Background and Related Work

With the emergence of the Internet and increase in the availability of information, web–based KMSs enabled organizations to take advantage of heterogeneous data: structured (e.g. relational), semi–structured (e.g. HTML, XML) and unstructured (e.g. plain text, audio/video) data [7]. In the web 1.0 era, using web mark–up languages such as XML reduced the obstacles to data sharing among diverse applications and databases [8]. Ashly *et al.* [6] explains some of the significant information challenges facing organizations and discuss how web–based

---

[1] http://www.xwiki.org/xwiki/bin/view/Main/WebHome
[2] http://calswim.org

technologies such as content management servers, portals, or document management systems lower information barriers. Alavi *et al.* [9] have addressed the requirements of KMSs; their survey shows that KMSs in general require a variety of technological tools in three areas: database and database management, communication and messaging, browsing and retrieval. They also highlight the need for seamless integration of the various tools in these three areas.

With the advent of web 2.0, collaborative and social technologies have made a significant impact on knowledge sharing. In contrast to the previous web 1.0 tools that passed on information to an inactive and receptive users, collaborative applications enable users to create and publish their own content [10]. Collaborative technologies can enhance collaboration between employees as well as sharing the organizational information and knowledge [2]. According to Sauer *et al.* [4], blogs and wikis are undemanding and affordable tools for communication and knowledge management. It is suggested that their rather unsophisticated structure can facilitate integration into already existing intranet and internet solutions. They go further by stating that web 2.0 tools such as blogs and wikis have the potential to offer all the features of complex and expensive IT-solutions.

Some research have studied the use of wikis as KMSs (so called corporate wikis), in corporations [11,12]. Wei *et al* [13] define corporate Wiki as an open community process that encourages multiple iterations in the creation of a knowledge repository. Pfaff *et al.* [12] explain participatory organizational processes of creation, accumulation and maintenance of knowledge, and analyze Wiki as a mediating employee-based knowledge management tool for democratizing organizational knowledge. They also have addressed quality issues in corporate wikis. They also believe that unlike radically open wikis (e.g. Wikipedia), which raises questions whether the information is authoritative and credible, corporate wikis contains more reliable information; employees who make contributions to the corporate Wiki are employed by the organization as specialists whose opinions will be highly regarded by their organizations as trusted and authoritative. Some work like [14] also shows that how in open wikis trustworthiness of users can be estimated by tracking users' actions over time.

The idea of using web 2.0 applications such as wikis, blogs and RSS feeds has been implemented in a commercial Mindroute Software [3], where organizational users use blogs for writing their comments, and wikis for sharing their knowledge (e.g. derived from meetings) in the organizations.

In this work, we discuss our investigations on enriching an open source wiki, XWiki, for knowledge management purposes. The wiki is designed for both organizational and volunteer users (e.g. researcher and professionals) in public. The new features integrated into the wiki enable users to retrieve data from local/remote relational databases or data spreadsheets. In addition, users can do data integration and analysis in a seamless way. The wiki (CalSWIM), which is currently used as California Sustainable Watershed/Wetland information Manager satisfies the requirements of an effective KMS.

---

[3] http://www.corporatewiki.us/?id=689

3   CalSWIM Wiki

CalSWIM is an ongoing project with the goal of facilitating open, coordinated, integrated, and informed decisions for those interested in water resources. With initial focus on California, CalSWIM provides a framework for engaging water management professionals, scientists, and the general public in the process of sharing information about water resources; therefore benefiting from critical open supervision of the content, and from valuable volunteer contributions that cannot occur in a closed infrastructure. The current work aims at creating a public and easily update–able information infrastructure of "all things watershed" that includes all watersheds in California.

CalSWIM is designed both as a public forum for exploring local watersheds and as a web location to help coastal managers make cost-effective and scientifically justifiable decisions regarding the monitoring, management, and alteration of coastal urban wetlands and their associated watersheds. To enable CalSWIM to scale up to its visionary goal, the first design goal has been being easily editable by environmental experts who are not software experts.

CalSWIM has three types of stakeholders:

1. *Organizations*, who have databases of historical and real–time data from their monitoring sites. CalSWIM enables organizations to link their databases to the wiki system in order to make the data accessible by public. Data is hosted on the organizational servers and only meta–data including read–only access credentials is registered in CalSWIM (Fig. 1). CalSWIM connects to these databases on demand to fetch the requested data.
2. *Environmental Experts*, who use the wiki system for retrieving data from the linked databases and processing it. They can share the results by creating new wiki pages, which can be accessible by other users.
3. *General Public*, who may use the CalSWIM wiki as a source of information or voluntarily contribute to it (for example by sampling environmental data and reporting the results in the wiki pages).

Fig. 1. Database Registration in CalSWIM

CalSWIM wiki aims at providing an infrastructure integrated with useful features in order to facilitate interactions among these parties. Currently, most of the organizations collecting environmental data either do not publish the data on the Web or have their own customized websites. Typically, users who are interested in accessing the data, need to go through several online forms in order to get the data.

The problem with this type of data sharing is that users need to download data and convert it to an understandable data model for their analytical tools. Then, they can share the result on the web. The goal of CalSWIM project is to streamline this Download–Analyze–Publish process. It enables users to insert standard queries into the wiki pages to retrieve the data from the linked databases. Users can then use the visualization and analysis tools integrated with the wiki in order to process the data and make the result accessible to the wiki users. The main advantage of this approach is that results are always up–to–date. As soon as data in linked databases is updated, content of CalSWIM pages is also updated. Fig. 2 shows the knowledge management stream in CalSWIM wiki.
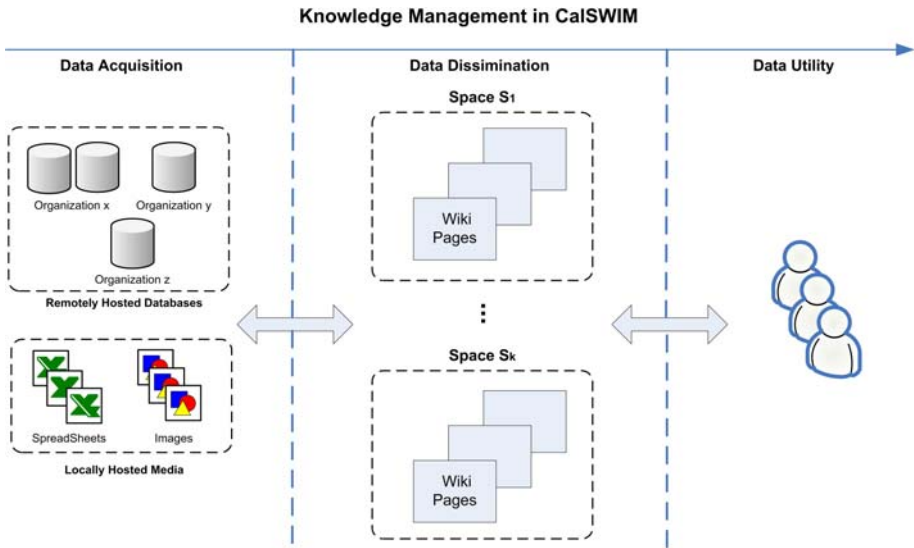

Fig. 2. Knowledge Management in CalSWIM

An example of visualization tools integrated with CalSWIM wiki is the map tool. Fig. 3 shows how a user has embedded a map into a wiki page. The map shows the location of watershed monitoring stations. The data is extracted from a linked database. The following snippet is used in a wiki page for for extracting the data from the database and generating the map:

```
#mapFromDB("StateWaterDB", "Select Name,
            Latitude, Longitude FROM Station")
```

Fig. 3. A Map Generated by a Database Query

The first parameter is the database nick name linked to the wiki and the second parameter is a standard SQL query which retrieves data from the database.

In addition to database retrieval tools, users can also extract data from attached spreadsheets and display them on wiki pages. For example, the same map as Fig. 3 can be generated if the data is in an attached Excel file using the following line of code:

```
#mapFromExcel("locations.xls", "Sheet1", "A1-C16")
```

The first parameter is the name of the attached Excel spreadsheet, the second parameter is the name of the sheet, and the third parameter is the range which contains data.

Charting tool is another visualization tool integrated with CalSWIM. Charts enable users to display real–time data extracted from online databases in the wiki pages. Inserting the following code snippet generates a chart as shown in Fig. 4.

```
{chart:
 source=type:query;
 dbnickname:StateWaterDB;
 command:SELECT Date, Result FROM sample
             Where StationCode = 'D7'|
 type=time|
 title=Sample Time Chart|
}
```

The following of this section, describes some of the other features of CalSWIM.

3.1   Content Creation

Content creation in CalSWIM is very easy. Users can use both WYSIWYG editor and wiki editor. The WYSIWYG editor is similar to popular word processors in which users see how what they type is rendered. In the wiki editor, users should use a simple wiki syntax for formatting their content. Our experiments showed
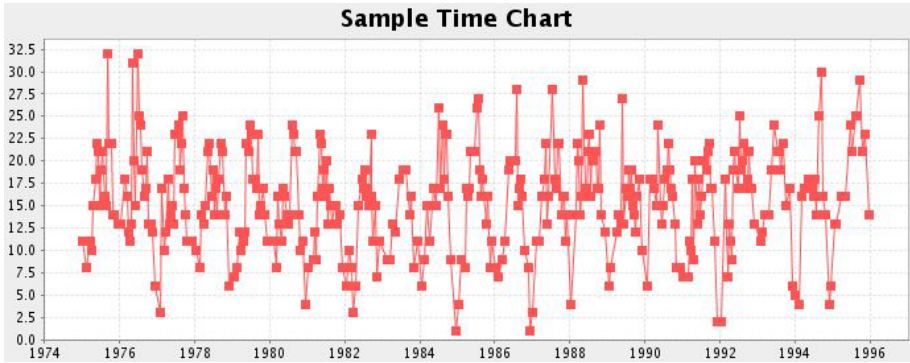
Fig. 4. A Sample Chart Generated by a Database Query

that most of the users that do not have programming skills (including most of the researchers) prefer WYSIWYG editor. These types of users are very reluctant to using codes. Even they hesitate in using very simple markups. Therefore we made the WYSIWYG editor the default editor of our wiki platform and extended it to provide simple forms for embedding maps, charts and other objects that are specific to CalSWIM.

3.2   Version Control

In CalSWIM, any change made to a document is automatically saved by Version Control component. It enables users to see how information has changed over time and gives them the ability to rollback documents to previous versions if needed.

3.3   Content Structure

CalSWIM allows users to organize pages into a set of *Spaces*. Wiki spaces are collections of related pages. Contributors can create new spaces, and add pages in them, at any time.

3.4   Access Control

CalSWIM has a three–level hierarchical access control mechanism. Access rights can be page–level right, space–level rights or global rights. Page–level rights override space–level rights and space–level rights override global rights. Rights can be assigned or denied from individuals or groups of users. Our experiments in CalSWIM showed that most of the experts only contribute in restricted spaces and pages that can only be edited by them. But are open for comments and discussion by others. Their argue is that if pages are going to be publicize their research and results of their work (with their name somewhere on them), they cannot accept them to be editable by others. In fact, while we want high school students looking at pages created by scientists and domain experts and detecting eventual mistakes, we do not want those students to change the content. Therefore CalSWIM is a mixture of Open and Closed content. But comments are always open for allowing general public to share the opinions.

4   Evaluation

To evaluate the effectiveness of our solution, in this section we compare our solution with other tools currently used in KMSs. Major requirements of KMSs have been discussed in [9,6]. Fig. 5 compares CalSWIM with other current solutions according to the requirements. The following is the summary of each requirement:

	File Server	Web Server	CMS	DMS	Wikis	Portals	Databases	CalSWIM
Easy Info. Creation	YES	NO	YES	YES	YES	PART	NO	YES
Easy Info. Access	PART	YES	YES	YES	YES	YES	NO	YES
Find-Grained Access Control	NO	NO	YES	NO	NO	PART	PART	PART
Automatic Integrated Version Control	NO	NO	YES	YES	YES	YES	NO	YES
Easy Collaboration	NO	NO	NO	NO	PART	PART	NO	PART
Fine-grained reuse	NO	NO	NO	NO	NO	NO	PART	PART
Integrated Productivity tool functionality	PART	NO	NO	NO	NO	PART	NO	PART
Full Content Search	PART	YES	YES	YES	YES	PART	PART	YES

Fig. 5. Comparison of CalSWIM with other knowledge management systems

Easy Information Creation and Access:
  Like other wikis, Calswim provides a user friendly platform to create and access information. Users can contribute with the system via almost every web browser.

Automatic Version Control:
  Like other Wikis, CalSWIM has a built–in version control component. This component enable users to view the previous versions of a document as well as the name of their editors and the timestamps. Users can rollback the current revision of a document to an older version if needed. In addition, they can use the diff functionality to view the difference between two versions of a document.

Easy Collaboration:
  CalSWIM Wiki provides some tools for data aggregation and analysis, and let users attach different types of files to a Wiki page. It make the collaboration easier compared to traditional wiki softwares; but wikis are not still perfect tools for user collaboration [6].

Full Content Search:
  CalSWIM supports two types of content (a) the content in the Wiki Pages (b) the content accessible from the linked databases. For the content in the Wiki pages, users can use full content search feature. However, the search is limited to the pages and the spaces they have access to. They can also narrow down their search to some special spaces. To search in the content available through the linked databases, users can use SQL queries, which enable them to formulate complex data retrieval queries.

Fine–grained access control:

Unlike traditional wiki software, CalSWIM provides page and space level access control. The most granular access rights can be defined on a Wiki page. Access control can be enhanced if we apply access rights on different sections of a wiki page as well as on inset and delete actions.

Integrated Productivity Tool Functionality:

CalSWIM allows users to attach any kind of files to pages and either reference to the attached file or use its content in creating part of the page content. For example, users can attach an Excel spreadsheet and use a simple interface to instruct the wiki to retrive data from this attachment and display it as a table in the page.

5   Conclusions and Future Work

In this paper we presented CalSWIM as a tool, which uses Wiki technology in knowledge acquisition, dissemination and utilization. It enables organizations to link their databases to the wiki engine, and make them accessible to public by defining suitable access rights. Using the integrated tools, users can (a) Fetch data from these databases; (b) Perform processing on the data; (c) Display or visualize the results and share them with others.

Currently, users need to use SQL queries to retrieve data from databases and represent it in the wiki pages, or to feed it to the analytical tools. According to the initial feedback of users, writing SQL queries is not easy for most of them. To overcome this problem, we are developing an easy to use query builder tool. In addition, some researchers want to import their documents in different formats such PDF, TeX and MS–Word in order to receive some comments from other users. A challenge in publicizing this type of content in CalSWIM wiki is to convert these documents to the wiki format. In addition, we aim at extending the CalSWIM to understand TeX markup, and use it as an alternative to its wiki markup.

Acknowledgement

References

1. Bolloju, N., Khalifa, M., Turban, E.: Integrating knowledge management into enterprise environments for the next generation decision support. Decis. Support Syst. 33(2), 163–176 (2002)
2. Fried, C., Edlund, A.: Web 2.0: It and knowledge management in the 21st century: A case study of mindroute software ab. Uppsala University, Tech. Rep. (2007)

3. Staples, D.S., Jarvenpaa, S.L.: Using electronic media for information sharing activities: a replication and extension. In: ICIS 2000: Proceedings of the twenty first international conference on Information systems, pp. 117–133. Association for Information Systems, Atlanta (2000)
4. Sauer, M., Bialek, D., Efimova, E., Schwartlander, R., Pless, G., Neuhaus, P.: "blogs" and "wikis" are valuable software tools for communication within research groups. Artificial Organs 29(1), 82–83 (2005)
5. Wan, L., Zhao, C.: Construction of a knowledge management framework based on web 2.0. In: WiCom 2007: International Conference on Wireless Communications, Networking and Mobile Computing, pp. 5341–5344. IEEE, Los Alamitos (2007)
6. Aitken, A.: Managing unstructured and semi-structured information in organizations. In: ICIS 2007: 6th IEEE/ACIS International Conference on Computer and Information Science, pp. 712–717. IEEE, Los Alamitos (2007)
7. Abiteboul, S., Buneman, P., Suciu, D.: Data on the web: From reltions to semistructured data and xml. Science Press (1999)
8. Seligman, L., Rosenthal, A.: Xml's impact on databases and data sharing. Computer 34(6), 59–67 (2001)
9. Alavi, M., Leidner, D.E.: Knowledge management systems: issues, challenges, and benefits. Commun. AIS, 1 (1999)
10. Abiteboul, S., Buneman, P., Suciu, D.: Coming to terms with web 2.0. Reference Reviews 21, 5–6 (2007)
11. Hasan, H., Meloche, J.A., Pfaff, C.C., Willis, D.: Beyond ubiquity: Co-creating corporate knowledge with a wiki. In: Ubicomm, pp. 35–40 (2007)
12. Pfaff, C., Hasan, H.: Can knowledge management be open source? In: OSS, pp. 59–70 (2007)
13. Wei, C., Maust, B., Barrick, J., Cuddihy, E., Spyridakis, J.H.: Wikis for supporting distributed collaborative writing. In: Proceedings at the Society for Technical Communication, pp. 204–209 (2005)
14. Javanmardi, S., Lopes, C.: Modeling trust in collaborative information systems. In: Proceedings of the 3rd International Conference on Collaborative Computing: Networking, Applications and Worksharing (November 2007)