A Case-Based Approach to Knowledge Management

Massoud Moussavi

The World Bank*
1818 H Street N.W.
Washington, D.C. 20433
mmoussavi@worldbank.org

Abstract

What distinguishes the World Bank from commercial banks and bilateral aid agencies is its wealth of development knowledge gained through economic and social development projects. Much of the available development information, however, continues to be scattered across a large number of documents and reports despite recent efforts of the knowledge centers to consolidate information. To retrieve a desired piece of information, the current Knowledge Management System (KMS) provides users with a powerful Web-based full-text search engine that allows them to perform a key-word search. A problem with the key-word search is that the users should be able to articulate exactly what they want. In addition, as the amount of available information increases, the searches would invariably lead to retrieval of more and more documents and one may need to perform many searches. This paper outlines a *case-based approach* to overcome some of the limitations of the current system and raise its operational impact. The proposed approach establishes a unified framework to represent the development knowledge and provides a better search and browsing mechanism to retrieve relevant and useful information.

Keywords - Knowledge Management, Case-Based Reasoning, Indexing, Information Retrieval.

1. Intelligent Assistant

> *"... The goal is not smart appliances and cars that talk to us. The primary goal is to develop systems that aren't stupid."* Christopher Riesbeck [Riesbeck, 1996].

Suppose you are planning an education project and wish to find some information about early child development in Pakistan. Using the facilities of the full-text search engine, you perform a key word search on the topic and retrieve the desired documents. For now, this may be sufficient since we have barely started encoding all the information in the knowledge base. But it will not be satisfying in the long-term. As the amount of available information increases, the searches would invariably lead to retrieval of more and more documents. Often, the search criteria have to be modified and a new search with a different combination of words submitted. Worse yet, after two or three searches, the system may retrieve a document whose content is worthless to you. At this point you might give up on the system and wish that it were not so stupid.

There are three issues involved here. *First*, we would like to have access to the entire development knowledge that exists in the collection of related experiences gained from Bank projects in a specific area. This includes staff decisions and judgments in taking certain actions as well as reasons for successes and failures of projects. *Second*, we would like to describe the problem (i.e., the search criteria) only once. *Third*, retrieve only relevant and meaningful information. The Knowledge Management System in the Bank can be improved in all three areas.

With regard to the first issue, notwithstanding the fact that a large portion of development knowledge is tacit, what has been codified is mostly scattered across documents that generally fall into two different classes:

(a) Documents that were mostly designed for management purposes rather than for sharing information.
(b) Key readings, lessons learned and best practices that have been collected and catalogued by different thematic groups. There are, however, no clear links between these documents and the project-related ones that the staff produce while working on projects.

Concerning the issues of describing a problem and searching for relevant information, the current KMS employs a Web-based full-text search engine that allows users to perform a key-word search. Hence, the problem of having to perform multiple searches that may or may not lead to finding the desired information as discussed earlier.

This paper outlines a *case-based system* for storing development knowledge, sifting through the encoded information, summarizing and correlating similar experiences gained through implementing economic and social development projects. Such a system will function as an intelligent assistant for Bank staff in planning and

* The findings, interpretations, and conclusions expressed in this paper are those of the author and should not be attributed in any manner to the World Bank, to its affiliated organizations, to the Board of Directors or the countries they represent.

implementing their projects. It would raise the operational impact of the KMS by providing:

- A unified framework to represent the knowledge that a task manager puts into the work, the lessons learned and good practices, as well as the information that a help desk staff collects in response to an inquiry.
- A search and browsing mechanism that is superior to traditional search engines in retrieving relevant and useful information.

2. Knowledge Management in the World Bank

The stated goals of the Bank's KMS are to increase staff effectiveness in doing their work and to capture and share development knowledge and experience with others. As discussed in [Hansen *et al.*, 1999] there are two very different knowledge management strategies: *codification* where knowledge is codified and stored in databases and *personalization* where knowledge is shared through direct person-to-person contacts. The knowledge management efforts in the Bank are in fact aimed at achieving the right balance between codification and personalization strategies. Recent discussions in the Bank [Denning, 1999] have centered on the role of the Bank as a storehouse as well as a broker of knowledge. The KM strategy of the World Bank is not limited to codification of knowledge or building a Web-site; knowledge sharing does indeed take place also through seminars, task team missions, and informal interactions. In this paper, we deal, of course, only with knowledge that can be automated.

In order to understand the role of the KMS, it is useful to distinguish between *development knowledge*, which has to do with the content of the project related work (one's experiences and stories) and knowledge *of processes*, which has to do with the flow of work. Staff who work on a project, say a child education project in India, need to learn about the relevant best practices or failures and successes of similar projects (i.e., *development knowledge*). They also need to know who to contact for certain questions, what the Bank's policy is concerning a particular issue, how to meet deadlines, how to write a report, how to meet the expectations of their managers, etc. (i.e., *knowledge of processes*).

Much of the knowledge management effort in the Bank has focused on providing staff with knowledge of processes. For example, staff might want to use a Terms of Reference document so that they can modify it to suit their need; or they might want to study Bank guidelines concerning loans to client countries. While the system also provides access to a large number of documents, it does not provide a mechanism to extract development knowledge from these documents. But focus on content is one of the important features of a KMS [O'Leary, 1998].

3. A Case-Based Framework to Represent Development Knowledge

Developing a World Bank project belongs to a class of "weak theory" problem solving domains such as medicine, social work, and law in which tasks are not algorithmic and one may not always be able to reason and plan based on a set of first principles. Rather, one deals with a new situation by remembering, modifying and applying prior experience. That is, one uses case-based reasoning.

The case-based nature of the knowledge about economic and social development has also been discussed in [Denning 1999; Ellerman, 1999]. According to Ellerman, a "best practice" might work well in some countries but fail miserably when recommended in other contexts.

What constitutes a case? Each development project embodies a certain number of characteristics which, taken together, represent a case. The content of a case should include:

- Lessons learned, key readings, and the information that a help-desk staff collects regarding a project. Note that in response to an inquiry, a help-desk staff gathers information for a particular situation (or case). It is important to capture and update this knowledge in a database of cases.
- Informal development knowledge shared among staff.
- Development information contained in different reports and documents.
- A course of action or solution to the problem.

A case is like a container that would hold the relevant development information from several sources. In addition, since most development knowledge is contained in the collection of related experiences, the cross-links between different experiences (cases) are highly relevant and can be established to reduce the retrieval overhead.

4. System's Architecture

The proposed system would allow the staff working on different development projects to navigate through a library of prior development projects and case studies to find an example that is similar to their situation, using a "conversational CBR" framework [Aha and Breslow 1997]. The users are prompted to answer questions as the system searches the case library for partially matching cases, and presents best cases to them. The users can then suggest changes and this cycle is repeated until a most similar case to their problem is found. As shown in Figure 1, the system would be composed of three integrated components:

- A *case library* containing concrete cases of development projects. It includes description of the approach taken to solve a problem in a country, lessons learned, pictures, sample video and audio clips, etc. Associated with this library is an *indexing vocabulary* that captures important features of development cases and thereby distinguishes one case from another.
- A *knowledge base* that contains general domain knowledge in the form of relations among features and heuristic rules. Armed with the domain knowledge, the system would be able to provide some general prescriptive advice concerning what methods to use in certain conditions (i.e., best practices).
- A *situation assessment user interface* that derives more meaningful features before attempting retrieval of a useful case. What distinguishes this component from standard information retrieval systems and Web search engines is its capability of using project-related

semantic vocabulary to help the user to formulate a query.

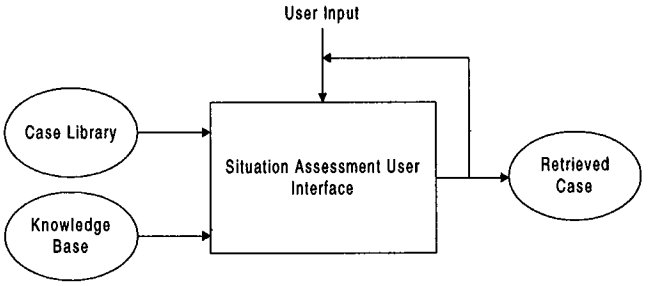

Figure 1: Situation Assessment & Retrieval Component

It is important to point out that the proposed system leaves adaptation of a case to the user. The system only helps to retrieve a case that can be adapted and when possible it suggests a "best practice". But as described in [Ellerman, 1999], it is difficult to know *a priori* just how general is a "best practice" and often the local adaptation amounts to reinventing the "best practice" in the new context.

5. Search and Retrieval: An Example

The retrieval of cases is closely related to the indexing method used. Indexes are labels that designate circumstances under which the case suggests a solution or anticipates a failure. One can distinguish between two types of case information: Indexed information that is used for retrieval, and non-indexed information. For example, in a system for retrieving information about child education we may use *type of community* and *economic condition* as two of the indexed features that can be used for retrieval and we may include graphs, photographs, and most of the text as features that are not indexed. In this example, a search can be performed based on the values of the indexed attributes (e.g., *type of community*: rural, *economic condition*: poor). Clearly, this is different from a keyword search in which one would search for words such as "rural" or "poor" and retrieve all documents (including those that are not relevant) that contain these words.

To illustrate further, consider a Bank document on education in Pakistan. In its conclusion, the paper states:

> "... It is better for the Government to give the village education committee the money directly. They will choose better...Although early evidence from Balochistan seems promising, the process may not easily be replicated in all rural communities. The absence of qualified female teachers and the community unwillingness to form a village education committee and establish a school in some communities will be the two main obstacles for further expansion of the pilot programs... There are also some orthodox Muslim families who do not send their girls to school and are of the view that education spoils the character of girls."

Assume that a user is planning an education project and is interested to find information about scholarship programs for Muslim girls in rural areas. The information retrieval mechanism employed by the existing KMS, leaves the problem of formulating the right query to the user who may search for "scholarship", "stipend" or "female education", etc. But invariably, one will find more documents than one would like; some of the retrieved documents may not be relevant at all and several searches may have to be carried out to find the relevant information.

A case-based system, on the other hand, retrieves information based on its indexes. Considering the above document (as well as other documents in the area of education), we can select the following features for indexing: *religion, cultural atmosphere, existence of segregation problem, economic condition, gender discrimination, type of stipend, type of community, female teachers, etc.* These are in fact part of the indexing vocabulary discussed earlier.

Armed with such a vocabulary, CBR assesses and interprets a new situation (situation assessment) to derive more meaningful features before attempting retrieval of a useful case [Leake 1996]. The system would lead the user through a number of questions based on the indexes described above. That is, it inquires about religion, cultural atmosphere, type of community, etc. Given that one is interested in a scholarship program for Muslim girls in rural areas, the system would retrieve the mentioned document plus some other documents like "Bangladesh: Female Secondary School Assistance Project". In addition, it may determine that the following Bank documents are also relevant: "Girls' scholarship pilot in Guatemala" and "China: Third Basic Education Project". (The latter document has a section on Muslim minority in rural areas in China.) Figure 2 shows a prototype of such a system using the Inference Corporation's CBR tool.

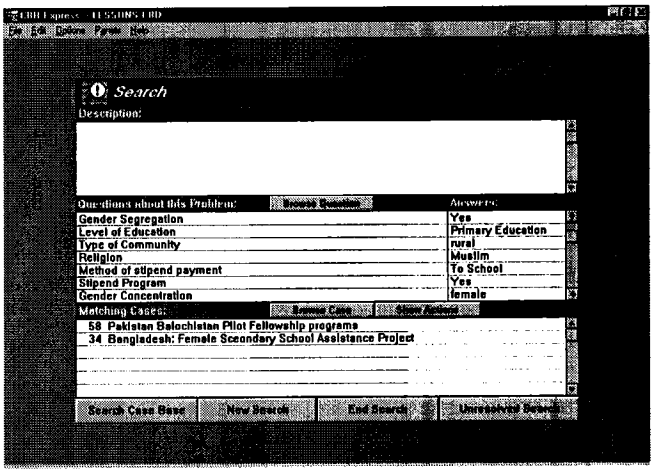

Figure 2: An example user interface

The retrieval process described above uses semantic-similarity metrics to retrieve most "similar" cases. Most CBR systems (including the Inference CBR tool) employ this technique. However, the most similar case may not necessarily be the easiest to adapt. Adaptation-guided retrieval is a novel technique that assesses the adaptation requirements of cases during retrieval [Smyth and Keane,

1995]. Adaptation of cases in the proposed system is left up to the user as adapting development knowledge to local conditions and culture requires considerable effort [Ellerman, 1999]. However, if the information concerning the cost of adaptation of a case is available, one might be able to link similarity assessment directly to adaptation knowledge [Leake *et al.*, 1996].

6. Summary

Building a case-based system to represent development knowledge is a big challenge because it implies indexing every document – a large task. But we can start with a portion of the existing documents, say the ones on most recent projects, sift through them, correlate similar experiences, and create a database of cases containing development information.

One successful implementation of a case-based system is at the NEC Corporation [Kitano *et al.*, 1992]. There is, of course, a major difference between knowledge management in the Bank and at the NEC Corporation. Quality of development projects is much harder to measure than the quality of software products and capturing development knowledge is considerably more complex than software development skills. Nevertheless, NEC's experience in developing a large-scale case-based system is a promising application of the CBR concept to creating a corporate memory.

A case-based system as an integral part of the KMS offers the following advantages:

- Summarizes all relevant project development information in one place (i.e., a case-base).
- Provides a search mechanism that is superior to the traditional search engines employed by the KMS, in retrieving relevant and useful information.
- Provides a better user interface so that the users do not need to have extensive prior knowledge about the domain in order to find what they want in the knowledge base.
- A deeper level of knowledge is obtained because of the cross-links between similar cases. A system that links related experiences would have significantly higher operational value.
- Once the cross-links are established, browsing of relevant information becomes simpler.
- It can significantly improve the quality of training and is particularly useful for distance learning.

A long-term opportunity would be to augment the retrieval component with a browsable case-base where the cases are richly indexed with inter-case links. This means that there will be no predefined retrieval involved: the user selects an initial case and then follows the links to other relevant cases based on the user's interests. In this regard, the general approach of the ASK systems [Schank *et al.*, 1991] for building a conversational advice-giving system can be employed.

Acknowledgements

I would like to thank Ulrich Hewer and Roberto Chavez for their encouragement and many valuable discussions regarding this paper. I would also like to thank Mohamed Muhsin and Steve Denning for their support and encouragement; Debra Basilis, Jahangir Boroumand, Ana Flavia Fonseca, Cheryl Kariya, Evelyn Anna Kennedy, Bill Lange, David Leake, Ali Semsarzadeh, and the AAAI reviewers for their comments and productive feedback.

References

[Aha and Breslow, 1997] Aha, D. W. and Breslow, L.A. (1997). Refining Conversational Case Libraries. In *Proceedings of the Second International Conference on Case-Based reasoning*, pages 267-278. Providance, RI: Springer-Verlag.

[Denning, 1999] Denning, S. (1999). KM Brainstorming of April 9. A World Bank Internal Document.

[Ellerman, 1999] Ellerman, D. (1999). Global Institutions: Transforming International Development Agencies into Learning Organizations. *The Academy of Management: Executive*, V. 13 (1, Feb. 99), pages 25-35.

[Hansen *et al.*, 1999] Hansen, M.T., Nohria, N., and Tireney, T. (1999). What's Your Strategy For Managing Knowledge. *Harvard Business Review*, March-April 1999, pages 106-116.

[Kitano *et al.*, 1992] Kitano, H., Shibata, A., Shimazu, H., Kajihara, J., and Sato, A. (1992). Building Large Scale and Corporate-Wide Case-Based Systems. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 843-849. Menlo Park, CA: American Association for Artificial Intelligence.

[Leake, 1996] Leake, D. (1996). CBR in Context: The Present and Future. In D. Leake, editor, *Case-Based Reasoning: Experiences, Lessons, & Future Directions*, pages 3-30. Menlo Park, CA: AAAI Press.

[Leake et al., 1996] Leake, D., Kinley A., and Wilson, D. (1996) Acquiring case adaptation knowledge: A hybrid approach. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 684-689, Menlo Park, CA: AAAI Press.

[O'Leary, 1998] O'Leary, D. E. (1998). Knowledge Management Systems: Converting and Connecting. IEEE Intelligent Systems, May/June 1998, pages. 30-33.

[Riesbeck, 1996] Riesbeck, C. (1996). What Next? The Future of Case-Based Reasoning in Post-Modern AI. In D. Leake, editor, *Case-Based Reasoning: Experiences, Lessons, & Future Directions*, pages 371-388. Menlo Park, CA: AAAI Press.

[Schank *et al.*, 1991] Schank, R., Ferguson, W., Birnbaum, L., Barger, J., and Greising, M. (1991). ASK-TOM: An experimental interface for video case libraries. In *Proceedings of the Thirteenth Conference of the Cognitive Science society*. Northvale, NJ: Erlbaum.

[Smyth and Keane, 1995] Smyth, B., and Keane, M.T. (1995). Experiments on adaptation-guided retrieval in case-based design. In *Proceedings of the First International Conference on case-Based Reasoning*, pages 313-324. Sesimbra, Portugal: Springer-Verlag.