Viewing Knowledge Management as a Case-Based Reasoning Application

W. Dubitzky[§], A.G. Büchner[§], F.J. Azuaje[‡]

[§] Faculty of Informatics, University of Ulster, at Jordanstown, Northern Ireland, UK
[‡] Northern Ireland Bio-engineering Centre, University of Ulster, at Jordanstown, Northern Ireland, UK
{w.dubitzky, ag.buchner, fj.azuaje}@ulst.ac.uk

Abstract

Knowledge management refers to the process of creating, sharing, and reusing of knowledge to improve and support the overall business strategy of an organization. *Case-based reasoning* is a general problem-solving or decision-making framework, which revolves around the processes of case retrieval, reuse, retention, and maintenance. Based on the obvious affinity of the two approaches, we are proposing a case-based model to knowledge management. The key assumption of this model is that knowledge management can be viewed as a *decision support* task. Both the notion of case-level information synthesis (knowledge creation) and the integration of multiple case bases (knowledge sharing) constitute crucial concepts in the proposed model.

Introduction

Knowledge management (KM) can be thought of as a business discipline involving the identification, capturing, organization, analysis, and processing of data and information to create new knowledge, which is made available to others in order to use and create more knowledge (Radding et al., 1998). KM solutions constitute an attempt to address the business problem that large amounts of an organization's knowledge are not reused. Benefits of successful KM include improved decision-making (faster, more accurate decisions), higher adaptability and flexibility, and prevention of knowledge loss. The key KM processes are knowledge creation, sharing, and reuse. Traditionally, KM solutions involve technologies such as computer networks (e.g., intranets, extranets), storage facilities (e.g., relational databases), capture and collection systems (e.g., document management systems, groupware), dissemination technologies (e.g., data warehouses), and knowledge processing and analysis technologies (e.g., data mining, statistical tools, on-line analytical processing, data visualisation).

Case-based reasoning (CBR) is a problem-solving method or reasoning model whose core processes revolve around the retrieval, reuse, and retention of previously encountered problem-solving episodes or cases (Lenz et al. 1998). CBR technology has been applied to a variety of analytic (e.g. classification, prognosis, decision support) and synthetic (e.g., design, configuration, planning) problem-solving or decision-making tasks.

The principal goals of CBR and KM are the same: to *capture* and *reuse* experience or knowledge. In contrast to KM, CBR work has put less emphasis on knowledge creation and sharing aspects; cases are assumed to exist and normally there is only a single case base. This paper puts forth a case-based KM framework by extending the current CBR process and explicitly putting knowledge creation and sharing on the same conceptual footing as, for example, retrieval. Key aspects of the proposed case-based KM process include the integration or *synthesis* of heterogeneous information at the level of individual cases, and the integration or *sharing* of multiple, autonomous case bases.

Although the paper has a strong positional character, some key aspects of the proposed CBR/KM architecture have been modelled and tested by the authors in various studies and experiments. Given the nature and available space, this paper is deliberately very focused. Therefore, a number of relevant aspects (e.g., knowledge as human resource, properties of knowledge and data, etc.) are ignored.

The remainder of the paper is organised as follows: To achieve some degree of self-containment, Section 2 recapitulates KM and CBR concepts important for the discussion. Section 3 is concerned with presenting the various aspects and the rationale for the proposed CBR/KM framework. Finally, Section 4 concludes the paper with a brief summary and some remarks.

KM and CBR: Setting the Scene

This section is intended to give a summary of some key concepts in KM and CBR, and to provide focus on the concepts involved in the proposed CBR/KM model.

Knowledge Management

The business problem KM is designed to solve is that the knowledge in an organization, which has been gained through experience, is not reused because it is not shared and made available to others in the organization in a formal way (Radding 1998). In order to address this problem, KM researchers and practitioners have devised various approaches that revolve around the basic knowledge processes of *creating*, *sharing*, and *reusing* knowledge.

In the basic KM process, we distinguish between knowledge creation, and knowledge dissemination and reuse. The basis for the KM process is manifest in the organization's disparate data, information, and knowledge sources, information systems, and information-oriented business processes. Up to 90% of the data held in these sources is *unstructured* information (e.g., text documents, graphics, images), and the rest is in structured format (e.g, databases).

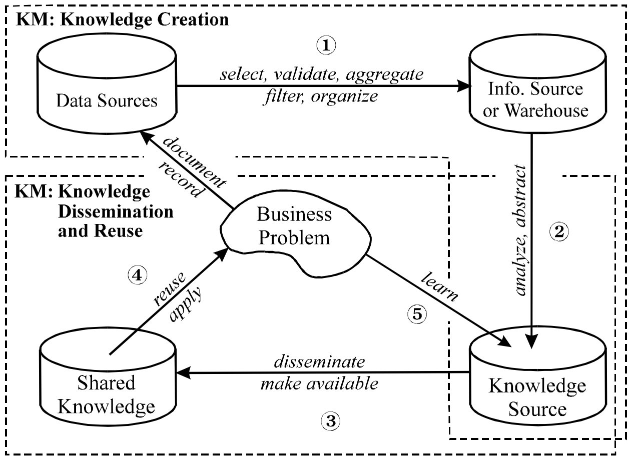

Figure 1: The KM process.

Consider Figure 1. The first step of the KM process is to filter, validate, select, and aggregate relevant data from disparate sources and organize and store it in some kind of *information source*, *data warehouse* or *staging area*. In step 2, the cleansed and pre-processed information is further analyzed, summarized, abstracted, and synthesized using powerful statistical and data mining techniques. The output of this step is considered knowledge, it is stored in some knowledge source (e.g., database, data warehouse). An example for a piece of knowledge created in step 2 is the business rule "Ignore all insurance claims of less than $750". Step 3 makes use of technologies such as groupware, computer networks, databases, and others, to share the newly created knowledge among various knowledge seekers within the organization. The shared knowledge can than be reused and applied to business problems, and the lessons learned from using that knowledge can be recorded, either as new knowledge or as raw data.

Some Potential KM Issues

The lack of systematically *integrating* the knowledge created in the KM process is arguably one of the more severe shortcomings in current KM and data mining approaches. At least for rule-formatted knowledge, this problem is partly due to the maintenance problem known from classical rule-based or expert systems. A crucial element for meaningful and efficient reuse of any knowledge is that the knowledge can be *accessed* based on the understanding of what the knowledge seeker *needs* or *wants* to know. It seems that existing KM models are more concerned with facilitating physical access, as opposed to *intelligent* access that is driven by the needs of the knowledge seeker. Consistent knowledge integration and relevance-driven knowledge access has been at the centre of CBR research. Moreover, there seems to be scope for improving the KM process with regard to consistently integrating related heterogeneous data (e.g., graphics, text) into reusable knowledge units. Recent work on second-generation CBR has been addressing issues arising from loosely structured and heterogeneous cases such as text documents (Brüninghaus and Ashley 1999), signals (Azuaje et al. 1999), and others (Gebhardt 1997).

Case-Based Reasoning

CBR refers to a reasoning model or problem-solving process that revolves around accessing and reusing previously encountered problem-solving episodes (cases) (Lenz et al. 1998). The notion of a *problem* in this context describes a very general concept ranging from simple to complex analytic (e.g., diagnosis) as well as synthetic (e.g., design) tasks. Including the design-time case base *creation* phase, the CBR process could be described by the sub-processes case *retrieval*, case *reuse*, and case *retention*.

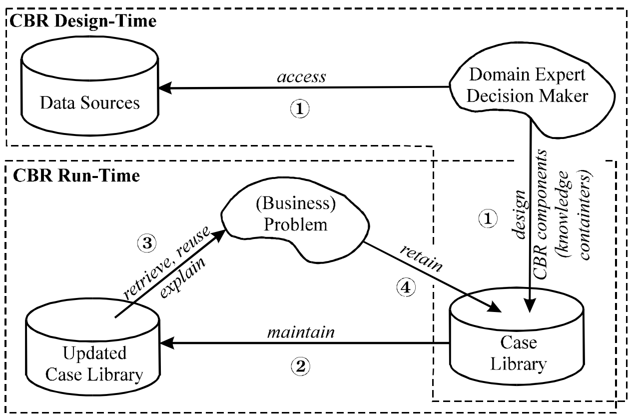

Figure 2: The CBR process.

Consider Figure 2. The first step in the CBR process is similar to the knowledge creation phase in the KM process, it revolves around constructing the actual case base. Typically, this part of the process is concerned with identifying and defining the content structure of the actual cases, a retrieval model (similarity measure, indexing structure), rules for solution transformation and case retention. Furthermore, a set of suitable seed cases must be selected in this phase. Conventional case base design is characterized by a considerable knowledge engineering effort involving subject matter specialists and knowledge engineers. Once the design is completed, the CBR process is largely governed by the processes case retrieval, reuse, and retention, and case base maintenance. Given the description of a new problem-solving situation, the case retrieval process locates those cases in the case base that are most ***relevant*** to the problem at hand. Relevance is normally approximated using a similarity measure (computational approach) or an indexing model (representational approach) (Liao et al. 1998). Based on the retrieved cases, the actual reuse or reasoning process takes place, it is concerned with adapting (the solution of) the past case to the problem at hand. This process may involve simple modification rules or more complex inferences. Case retention refers to integrating the new problem-solving episode or case into the case library, and, possibly, updating of other knowledge containers (e.g., indexing structures, adaptation rules). The maintenance process encompasses consolidation of the case base's knowledge containers (Leake and Wilson 1998). Two key issues faced by this process are the ***utility problem*** — which has to do with the trade-off of the system's performance (time complexity) and reasoning competence (Smyth and Keane 1995) — and concerns about the ***consistency*** and accuracy of the case knowledge (e.g., obsolete cases need to be updated or removed).

The KM/CBR Synergy Framework

Among other aspects, such as the prevention of knowledge loss, KM solutions could essentially be considered as ***decision support*** systems whose knowledge content is composed from various distributed and heterogeneous data sources, and made available to decision-makers according to the KM process depicted in Figure 1. Although the notion defies precise definition, decision support generally refers to ***data and information access facilities targeted to provide the information needed by (business) decision-makers***. Decision support can be as simple as providing front-line managers periodically with reports on production, or it can be as complex as extensive modelling of prospective customers using sophisticated techniques like Bayesian belief networks. Decision support systems are often used in domains where terminology is highly ambiguous, context plays a strong role, or existing domain knowledge is too incomplete or weak for stronger problem-solving methods to be applied. Decision support belongs to the broad category of ***analytic*** tasks (Lenz et al. 1998). Analytic tasks involve the analysis and interpretation of a given situation, and the drawing of inferences based on the results of the analyses and interpretations. However, in contrast to more specific analytic tasks, such as classification and diagnosis, decision support is of a more general nature. It is often not obvious, which parts of the situation describe the problem and which the solution, or even to decide if a solution proposed by the system is correct or not, or more useful than another solution or not. CBR systems have been successfully applied to both analytic as well as synthetic problem tasks. Thus, CBR can fulfil the decision-support role of a KM system.

Viewing the KM run-time as a decision-support CBR system (see Figure 3) has the advantage that powerful CBR mechanisms, including relevance-based retrieval, case reuse, and learning, could be used for building KM systems. However, the question of how the KM sub-processes of (initial) knowledge creation and sharing are handled within case-based KM framework needs to be discussed. (The items labelled in bold in the diagram indicate KM elements introduced from the CBR framework).

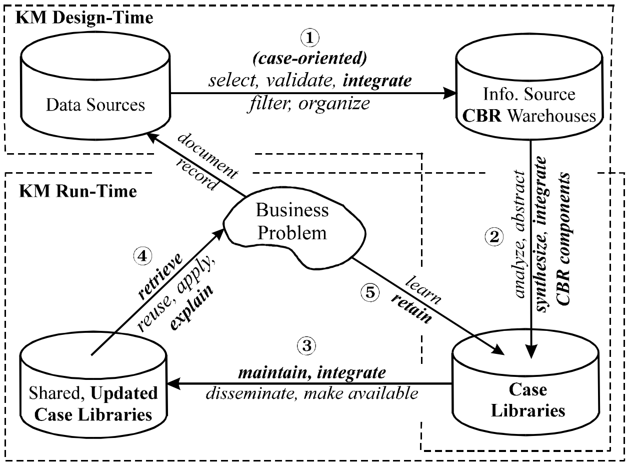

Figure 3: Case-based KM process.

The standard KM knowledge creation process is characterized by analysis and abstraction. In contrast, knowledge creation in the proposed case-based KM will focus on *synthesis* and *integration*. Given the diverse types of information sources in the organization, integration at case level will involve representing case content that is of heterogeneous nature, and effective handling (retrieval, reuse) of such cases. Sharing in the standard KM process is often achieved by simply providing "physical" access to various data warehouses and data marts. In the proposed case-based KM model, sharing will require relevance-driven access to a set of heterogeneous case bases based on the needs of the knowledge seeker.

Integration: Heterogeneous Cases

At the case level, the case-based KM approach will require the selection, validation, filtering, and organization of information that will be used to describe case content information. This information will be held in one or more *CBR warehouses* (see Figure 3). A CBR warehouse is similar to its conventional counterpart except that its content is geared towards construction of case bases. Each CBR warehouse will eventually be associated with a more or less autonomous case base or case library — called departmental case base. This sub-process will require analytic as well as synthetic tasks. It will involve decisions on what constitutes a case, what vocabulary should be used to describe the case content, what model will be suitable to describe relevance, and so on. Because graphics, text, and other types of information will be crucial in describing the content of individual cases, this process will draw upon the lessons learned with more recent second-generation CBR (see Section 2.1.1) and *data fusion* approaches. Data fusion refers to the acquisition, processing, and merging of information originating from multiple sources to provide a better insight and understanding of the phenomena under consideration (Arabnia and Zhu 1998). In (Azuaje et al. 1999) we have proposed a case-level integration model, that combines signal (ECG signals) case data with structured patient records (Figure 4b). An index knowledge discovery model, based on *growing cell structures* neural networks (Fritzke 1996), was used in these experiments to automatically generate the indexing-based retrieval model.

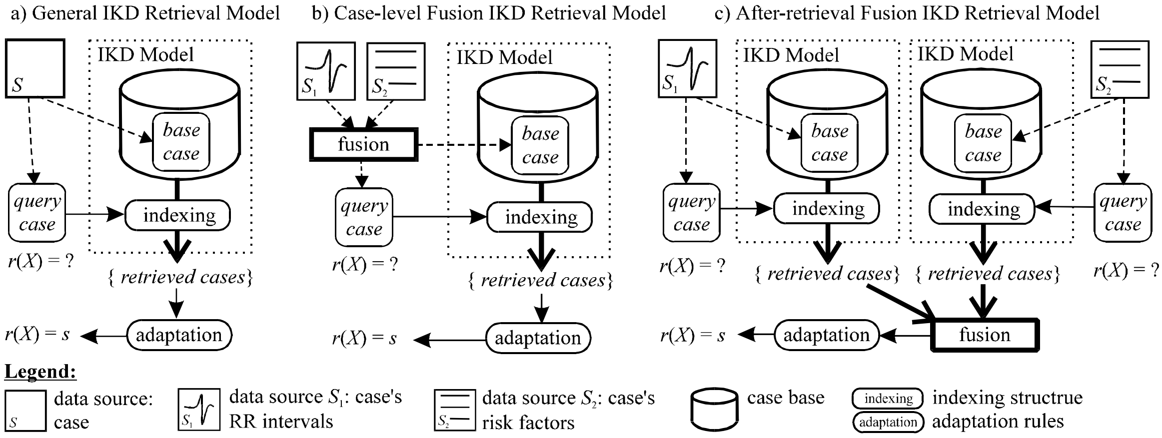

Figure 4: Case-level and inter-case-base integration (IKD = index knowledge discovery)

Integration: Heterogeneous Case Bases

Within the proposed case-based KM model, knowledge sharing will be viewed as the provision and integration of relevance-driven access to a set of autonomous, heterogeneous case bases (departmental case bases). Simply speaking, the integration at this level will facilitate the retrieval of relevant cases from the set of underlying individual case bases without requiring the user of this knowledge pool to possess detailed information on the various cases bases (meta data, case base location, etc.). Instead, the user should only be required to have some knowledge of the vocabulary relevant to his or her area of interest or expertise. To enable a coherent retrieval mechanism across a set of loosely coupled or independent case bases, a mechanism must be provided that is able to handle interoperability conflicts arising from *schematic* (syntactical description of case content) and *semantic* (knowledge content of case) *heterogeneity* across different case bases. Focusing on semantic heterogeneity (the more complex of the two types), we propose a model based on (Kashyap and Sheth 1998). This model achieves semantic interoperability using terminological relationships between terms across ontologies. This approach is based on the assumption that *terminological ontologies* are used to represent the knowledge content of cases across the entire set of case bases. The basic architecture for integrating multiple departmental case bases is depicted in Figure 5.

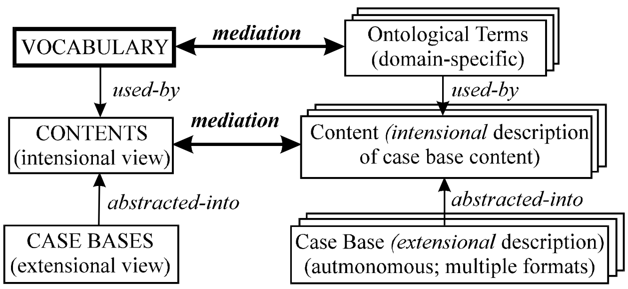

Figure 5: Shared access to multiple case bases (adapted from (Kashyap and A. Sheth 1998)).

The problem of semantic heterogeneity is the identification of semantically related objects (cases) in different case bases and the resolution of discrepancies among them. The key to interoperability is *vocabulary sharing* among the ontological terms and case content descriptions associated with the various case bases (highlighted elements in Figure 5). The *ontology level* represents domain and application-specific terminologies. To integrate the terms into a single *vocabulary*, mapping-based mediation is necessary. Appropriate relationships that allow inter-ontology interoperation are *synonyms*, *hyponyms*, and *hypernyms* (Kashyap and Sheth 1998). We have extended their approach with an *alternative* relationship and applied the concepts in the context of multiple manufacturing ontologies (Büchner et al. 1999). The *content level* represents intensional descriptions of the meaning of multiple case bases, each of which uses its own application-specific ontological terminology. In order to create an intensional view that contains the contents of all case bases participating in the case base federation, mediation is required, which is usually relying on Boolean combinations of content descriptions. The *case base level* holds extensional descriptions of each participating case base instance. The virtual synergy of all loosely joined case bases results in the extensional view. Explicit mediation is not necessary, since it is implicitly achieved through the negotiation mechanism at the content and ontology levels. Within a software design reuse scenario, we have demonstrated that the terminological ontology approach is useful to model cross-domain remindings (Lester et al. 1999). The basic concept used to bridge domain boundaries is the *synonym* relation. In (Azuaje et al. 1999) we have combined the knowledge contained within two differently-formatted case bases at the retrieval-result level (Figure 4c).

The Need for Automation and Tools

Clearly, there is a need for appropriate tools and automation techniques that facilitate the construction and integration of cases with heterogeneous content; the design and management of suitable retrieval and maintenance models; the management and maintenance of multiple case bases. Representing multi-format case content will certainly require tools that support the labelling and annotation of the knowledge contained in cases. Machine learning techniques and tools such as genetic algorithms and neural networks could turn out to be crucial for automatic or semi-automatic construction of suitable vocabularies, indexing (Azuaje et al. 1999), similarity (Dubitzky et al. 1999), adaptation (Hanney and Keane 1997), and other knowledge containers.

Brief Summary and Some Remarks

Viewing KM as a decision support exercise, the paper outlines some aspects of a possible approach to KM through CBR. It proposes the synthesis of disparate and heterogeneous information sources into cases as a key element of the knowledge creation process. Further, it suggests that effective knowledge sharing could be achieved via relevance-driven access to multiple case bases on the basis of ontology-based mediation models. The underlying character of the paper is deliberately positional, its main intention is to stimulate discussion. Having adopted a CBR-biased approach, we are particularly interested in comments, criticism, and feedback from the KM community.

References

Arabnia, H., and Zhu D. (eds.), 1998. In *Proc. of the Int'l. Conf. on Multisource-Multisensor Information Fusion, (FUSION-98)*, vol. I and II. CSREA PRESS.

Azuaje, F., Dubitzky, W., Black, N., Adamson, K. 1999. Improving Clinical Decision Support Through Case-Based Fusion. To appear in *IEEE Transactions on Biomedical Engineering, Special Issue on Biomedical Data Fusion*.

Brüninghaus, S., and Ashley, K.D. 1999. Bootstrapping case base development with annotated case summaries. To appear in *ICCBR-99*.

Büchner, A.G., Ranta, R., Hughes, J.G, Mäntylä, M. 1999. Semantic Information Mediation among Multiple Product Ontologies. To appear in *Proc. World Conf. on Integrated Design & Process Technology*.

Dubitzky, W., Azuaje, F., Lopes, P., Black, N., McCullagh, P., and Song, Y. 1999. On Local and Global Feature Weight Discovery for Case-Based Reasoning. In *Proc. ISCA 14th Int'l. Conf. on Computers and their Applications*, 107-110.

Fritzke, B. Growing Self-organising Networks — Why? 1996. In *European Symposium on Artificial Neural Networks*, 61-72.

Gebhardt, F., 1997. Survey on structure-based case retrieval. In *The Knowledge Engineering Review*, 12:1, 41-58.

Hanney, K. and Keane M.T. 1997. The Adaptation Knowledge Bottleneck: How to Ease it by Learning from Cases. In *Proc. 2nd ICCBR-97*, 359-370.

Kashyap, V. and Sheth, A. 1998. Semantic Heterogeneity in Global Information Systems: The Role of Metadata, Context, and Ontologies. In *M.P. Papazoglou and G. Schlageter (eds.), Cooperative Information Systems: Trends and Directions*, Academic Press, 139-178.

Leake D.B. and Wilson D.C., 1998. Categorizing Case Base Maintenance: Dimensions and Directions. In *Proc. 4th EWCBR-98*, 196-207.

Lenz, M., Bartsch-Spörl, B., Burkhard, H-D., and Wess S. (eds) 1998. *CBR Technology: From Foundations to Applications*, Springer-Verlag.

Lester, N.G., Wilkie, F.G., Dubitzky, W., and Bustard, D.W. 1999. A Knowledge-Guided Retrieval Framework for Reusable Software Artefacts. 1999. Submitted to *Int'l. Information Reuse and Integration Conf*.

Liao, T.W., Zhang, Z., and Mount, C.R. 1998. Similarity Measures for Retrieval in Case-Based Reasoning Systems. In *Applied Artificial Intelligence*, 12: 267-288, 1998.

Radding A. (1st ed.). 1998. Knowledge Management: Succeeding in the Information-based Global Economy. In *Report of Computer Technology Research Corp*.

Smyth, B., Keane, M.T. 1995. Remembering to Forget: A Competence-Preserving Case Deletion Policy for Case-Based Reasoning Systems. In *Proc. Int'l. Joint Conf. on Artificial Intelligence*, 337-382.