Knowledge Management revisited: needn’t you represent what you want to manage?

Luca Gilardoni, Clara Bagnasco, Christian Biasuzzi, Massimo Ferraro, Silvia Mazza

1 Quinary S.p.A., Via Pietrasanta 14, 20141 Milan, Italy
L.Gilardoni@quinary.com

Abstract. Knowledge Management is a discipline that investigates how to consolidate and augment individuals’ expertise and know-how and how to make them part of a corporate knowledge asset. We believe that this goal can not be achieved without defining a structured representation of the enterprise knowledge asset, that should act as the explicit context inside which each individual matches and augments his own individual knowledge. Such representation is what, in the Artificial Intelligence community, has been called Knowledge Base. In this paper we compare the definitions of knowledge given in the Knowledge Management and Artificial Intelligence community and present one of the main agreed principles beyond Knowledge Management: how corporate knowledge flows and is augmented in a cyclic process, called “knowledge cycle”. We then describe our view and our experiences about how the different sub-processes of the knowledge cycle can be made concrete by means of Information Technology tools acting on an explicit representation of the corporate knowledge. We give also some hints about how such representation can be used as a gluing factor to make the whole cycle concrete.

1 Introduction

Knowledge Representation has long been a main topic in the Artificial Intelligence (AI in the following) field, and Knowledge Based Systems a major field of application of AI techniques. Recently, at least the term “knowledge” has become part of the new buzzword: “Knowledge Management”, a discipline gluing together themes from apparently disparate fields such as organization and social sciences, business management, Artificial Intelligence. Despite the large number of scientific and technical publications dealing with Knowledge Management and Knowledge Management tools, it is difficult to find in them a short statement defining the term Knowledge Management (KM in the following). Common points are anyway that knowledge is information combined with experience, context and interpretation, and that individuals are the source of two types of knowledge, tacit and explicit. Is thus agreed that a successful KM strategy should facilitate the transfer of knowledge assets from an individual to another, enabling bi-directional transformations between tacit and explicit knowledge. Information Technology can play an important role in implementing these processes across a distributed enterprise, as it can provide tools that facilitate the communication, cooperation and information sharing among the members of the organization. However, given that the matching between information and context is crucial for knowledge creation, an effective KM tool should enable to represent context (the enterprise consolidated and explicit knowledge) in a rich and structured way, so that to support the detection of new relevant information and the correct integration of it with pre-existing knowledge. We therefore believe that Artificial Intelligence techniques for Knowledge Representation could be the basis for encoding and maintaining domain and organizational knowledge, to effectively support all the activities aiming at defining, exploring and augmenting the “context”, to share information and knowledge. Moreover, as most information is available nowadays in textual form, and as natural language is the most natural way to express “anomalous states of knowledge” (i.e. vague perceptions of knowledge, that results in under-defined statements) we also believe that Natural Language Processing techniques, when properly integrated with explicit domain knowledge, can provide tools to support advanced text classification and sharing of near-explicit knowledge.
2 Perspectives on knowledge

The nature of knowledge and of intelligence has been investigated by philosophers for hundreds of years. AI researchers have traditionally taken a pragmatic view of knowledge, considering knowledge what enables an artificial system to behave intelligently. “In AI, a representation of knowledge is a combination of data structures and interpretative procedures, that, if used in the right way in a program, will lead to knowledgeable behavior” [Barr 81]. According to this, data structures in AI Knowledge Bases are intended to be “knowledge” as they represent, in a semantically sound formalism, facts or rules that can be used by a program to behave in a knowledgeable way.

Knowledge Representation (KR in the following) studies have brought to the development of several theories and techniques that focus on formalization of knowledge within a formal framework enabling manipulation and usage; long debates took place between the proponents of “declarative” and “procedural” knowledge1, that was an AI transposition of the philosophical distinction between “knowing what” and “knowing how” [Winograd 75]. However, apart from few exceptions, most notably being the work of Brooks [Brooks 91], there has been general consensus on the need of explicit representation systems as the foundation of AI research.

On the other side, during recent years there has been a rising interest within the business world for the role of knowledge, intended here as what enables people and organizations to behave intelligently. In this perspective, “Knowledge Management projects are attempts to do something useful with knowledge, to accomplish organizational objectives through the structuring of people, technology and knowledge content” [Davenport 98]. This is not surprising as the worldwide economy is currently transforming from an economy based on goods production to an economy of services, and knowledge is no more just a resource needed, as well, and with equal status, as materials, labor, capital, but is the only significant resource of today economy [Drucker 1993]. The effective managing of intangible assets has indeed a crucial role in driving present economic processes. In several sectors (such as banking and insurance) acquisitions and merging originate the need to transfer and homogenize different kind of experiences and work practices. In the manufacturing industry, market globalization and process outsourcing make it necessary to readily transfer knowledge assets to a new generation of technicians, suppliers and subsidiaries. Everywhere the increasing importance of customer care services (such as call centers) implies the need of making knowledge readily available to front-line employees. In summary, the exploitation of the enterprise knowledge asset is today considered as the key to innovate, to reduce production time while increasing product quality, to support decision making and to maintain market shares and customer liability.

The focus of Knowledge Management studies, being their main subject human organizations and not artificial systems, has however not been put on Knowledge Representation formalisms, but on knowledge creation and sharing mechanisms. Moreover, a common agreement seems to exist in establishing the need to take into account tacit, or implicit, knowledge, that is knowledge that cannot be represented.

Davenport [Davenport, 98] gives the following definition of Knowledge: “Knowledge is information combined with experience, context, interpretation, and reflection. It is a high-value form of information that is ready to apply to decisions and actions…Knowledge is fuzzy and closely linked to the people that hold it; its categories and meanings change frequently…..”. When we speak about knowledge we thus should have in mind different types of things, as information, logical models, cultural backgrounds, know-how and intuition. While explicit knowledge can be somehow formalized into a logical model or in a set of statements, tacit knowledge is personal, context specific and difficult to transmit. Types of tacit knowledge include hands-on skills, special know-how, intuitions, and the like. As Michael Polanyi [Polanyi 66], the first to distinguish tacit from explicit knowledge stated “We can know more than we can tell”. Knowing how to effectively perform a job means understanding both types of knowledge.

The goal of a knowledge management program should thus be to encourage the circulation of knowledge, both implicit and explicit, inside the organization and to stimulate the creation of new knowledge. Accordingly, the goal of any information technology infrastructure deemed to support knowledge management processes should be to make this easier, by providing suitable knowledge aware tools. We believe that the only way to make this tools really knowledge aware is to found them on suitable representation mechanisms.

1 “Declarative representations stress the static aspects of knowledge – facts about objects, events, and their relations and about states of the world. The proponents of procedural representations pointed out that AI systems had to know how to use their knowledge – how to find relevant facts, make inferences, and so on – and that this aspect of knowledgeable behavior was best captured in procedures” [Barr 81]
3 The knowledge creation cycle

On of the main agreed principle inside the Knowledge Management field is that corporate knowledge flows and is augmented in a cyclic process, that has been described by Nonaka and Takeuchi in the book “The knowledge creating company” [Nonaka 95]. In this book, the authors examine the different epistemological attitude of the western school of thought (that is dominated by the Cartesian dualism between the thinking subject, that knows, and the object, which is known) and of the Japanese one (that does not postulate a separation between subject and object). According to the authors, this is the reason why western managers are more concerned with “managing” explicit knowledge, given in forms of process models, logical forms and documents stored in the company repository, while Japanese managers tend to concentrate on how to stimulate the creation of new tacit knowledge, i.e. of know-how and shared mental models.

As stressed by Nonaka and Takeuchi, these views are complementary and must be integrated to define a successful management strategy: organizational knowledge is created and shared through a cyclic process, which involves bi-directional conversions of explicit and tacit knowledge. The overall knowledge cycle that has been adopted by all the KM analysts involves four different processes: socialization, externalization, combination and internalization (see Figure 1).

![Figure 1: the knowledge creation cycle](image)

**Socialization** is a process based on sharing of experiences that results in the creation of tacit knowledge. Tacit knowledge is acquired by direct interactions with other people, even without the use of language. An example of socialization is training on-the-job.

**Externalization** is a knowledge creation process, in which tacit knowledge becomes explicit by assuming the form of concepts or models. Writing is the most common form of externalization, but mental pictures are sometimes difficult to translate by means of linguistic expressions. The gap between mental pictures and linguistic expressions stimulate the interaction among individuals, which is often driven by non-analytical methods, as analogy and metaphor.

**Combination** is the organization of explicit knowledge in a knowledge system, in which existing explicit knowledge is reconfigured by means of clustering, categorization and combination with new explicit knowledge.

**Internalization** is the process in which explicit knowledge becomes part of the individual tacit knowledge asset. In this way, the experiences gained and formalized in the previous phases of the knowledge cycle become really usable inside the organization.

Given the agreed goal of making the knowledge flow and grow through the knowledge cycle, ideas about what kind of actions should be the backbone of a KM program differ. Part of the KM literature focus on managing people and investigates issues like how to encourage creativity and how to motivate people to share their knowledge assets [Garvin 93, Leonard 97]. In this perspective, Knowledge Management projects are qualified as efforts to re-engineer organization behavior and management strategy that are independent of technology. On the other hand, many publications to date [van der Spek 97, De Michelis 98] assert that successful KM programs require a change both in organizational behavior and in technology infrastructure. Even if technology is not the only solution to an organization’s knowledge management needs, it is however required to enable the organization’s knowledge management processes.

Support offered by current IT is however mainly focused on **Combination** activities, covered by text and data mining tools, document management systems, information retrieval engines and data warehouses tools, with some
limited extension towards **Socialization** and **Externalization** by means of collaborative environments based on workgroup packages and/or Intranets. 

Because of underlying technology, relationships among data are however generally loose: tools integrating structured data require manual modeling efforts and are generally unaware of the semantics of data, while textual material is supported by word-level indexing and retrieval mechanism, which focus on the storage and basic access mechanisms, with minimal support for content handling. These tools tend to provide a reasonable infrastructure in terms of connectivity, sharing and document management, but are generally weak in supporting content based (not to say knowledge based) integration, as they lack adequate mechanisms for structuring context and domain knowledge, for finding the relevant information for a given context and for properly correlating data. As a consequence, support for the **Internalization** process is also poor. Overall, it is evident to users, and remarked by analysts, that products currently on the market cope only minimally with the issues involved in making the cycle concrete. Gartner Group [Gartner 99] reports that IT tools should acquire knowledge representation and language technology capabilities integrated before 2001 or be faced with vast but unusable knowledge stores.

To properly manage information it is needed to work at the level of knowledge (what information is used for) [Trickey 98]; the knowledge structures behind information have to be taken into account, made explicit and exploited in access mechanisms. Only this can start a virtuous cycle in which new information, classified into a known schema, makes knowledge grow and eventually enables building a knowledge warehouse supporting effective sharing.

### 4 Encoding domain and organization knowledge

While KR studies were born within the AI community to support intelligent agents reasoning, rich knowledge representation schemas are also needed to support effectively knowledge gathering and sharing where actors are humans and organizations. The knowledge creation cycle as describe in figure 1 has in fact as a prerequisite the establishment of a common ground favoring on one side communication between actors and on the other side supporting dynamic integration of new knowledge within the common schema.

Common ‘ground’ knowledge is used at all level within the cycle. Even in socialization activities, the most ‘fuzzy’ by definition and perfectly exemplified by the ‘coffee machine chat’, the place (virtual of physical) where people meet to communicate, a common background between individuals is mandatory to make communication possible. An oil field worker and a farmer would have to synchronize and recognize each own background while talking about ‘oil’, and while the specific example could seem trivial, we have encountered the same problem in ‘conceptual alignment’ almost everyday (and often nearby a coffee machine) on much less trivial examples. Shared background knowledge about domain specific relations and entities (drilling vs. oil production, or geophysical models vs. oil prospection) is also fundamental in enabling the creation of connections between information and between individuals, whether in human to human interactions or in human interactions with information repositories.

When searching over textual sources (e.g. on the web, or in large unstructured repositories), it is this background knowledge that enable building sound queries able to sort out material. 

A search for “blowout” (a very well defined concept in the oil domain, referring to wells or pipelines accidents) with Altavista returns over 40,000 hits, no suggestion (‘related search’) referring to oil domain and no top hit in this area. When doing a more precise search adding domain specific terms (and also exploiting some meta knowledge on specific search engines), hits reduce to about 4000, and adding terms conveying the meaning in a given context (e.g. apparels to manage oil blowouts) reduces to a much more manageable and sound set, albeit still containing quotes to environmentalist newsletters (which may or may not be relevant, depending on the ultimate purpose of our search) and surely missing relevant information not properly covered in the search query.

What’s relevant here is not to highlight well known limitations of currently available search engines, but to remark that the only factor, together with a bit more sophisticated linguistic handling, enabling to overcome limitations is proper embedding of domain and context knowledge.

We adopt here a pyramid metaphor (figure 2), where the pyramid shape indicates at the same time focusing capabilities and (inversely) the amount of knowledge needed to support this focus.
The narrower and more focused the area, the more specific knowledge we have to employ, and the more focused results are. At the same time, this focused knowledge has to provide cross-domain and cross-text clues and connections, to enable establishing relations between more loosely related information and to enable sharing of knowledge between individuals.

Similarly, the same kind of knowledge about semantics of entities is what enables comparison and integration of structured or semi-structured data, to perform schema integration as in [Bergamaschi 99], where a rich representation is used to support reasoning within and over existing schemas, and/or access to semi-structured sources, as in [Singh 98]. Here too, a rich conceptual representation enables to express the semantics of entities and relationships and their links to natural language terminology. Other than enabling comparison of existing schemas based on schema descriptions and entity and attributes lexical proximity, as in [Bergamaschi 99], or providing services that integrate heterogeneous web data as in [Singh 98], such a representation could also be used to perform context aware navigation and access through structured databases, as exemplified by Quercus products (www.quercusresearch.com).

Accordingly, the pyramid metaphor used before can be made more complete putting the two worlds of textual and structured information aside (see figure 3).

In figure 3, the lower part mirrors the upper, with more advanced methods for structuring data, based on knowledge about data itself, supporting data integration and browsing. The glue factor, the knowledge that can be encoded in the system, is represented in a Knowledge Map that can be understood in terms of domain entities and relations.

Technology described here is currently behind commercial web services from Cadabra.com, previously Tesserae Inc.

---

2 Technology described here is currently behind commercial web services from Cadabra.com, previously Tesserae Inc.
The Knowledge Map should enable users to sort out data and to reason over them, to relate structured and unstructured data, and, at the same time, to share individual knowledge by integrating it with an understandable corporate knowledge representation. To reach these goals, it must exhibit three fundamental characteristics:

1. it must be well founded, and here AI techniques and research results play a key role, enabling manipulation and sound inferencing;
2. it must enable connections between conceptual models, that describe entities in terms of conceptual relations (e.g. a share ISA financial_instrument, a market_index ISA financial_variable LINKED_TO a stock_market) and linguistic models and terminology (e.g. a share is denoted by the English word “share” – which has different meanings, and specific linguistic features in terms of morphology, PoS roles, etc.);
3. it must be easily usable by human actors: this means that it must provide effective navigation tools and interfaces but also intuitive views over the overall ontology (a distinction between sortal and non sortal concepts, albeit probably relevant for reasoning purpose, could have no meaning for a ‘normal’ end user).

The latter two aspects enable building semantic networks able to effectively support browsing, integration of data sources and integration within communication packages. Semantic networks have been at the foundation of AI research, [Minsky 75] but here our interest is not in defining sound inference mechanism but in providing the glue to support human interactions. When attacking the knowledge creation process within organizations, we have to put human interaction and context into play, and to make the representation directly accessible and understandable to end users to make the system really usable.

5 Knowledge based information organization: an example.

An example of how an explicit representation of domain and organization knowledge can be used to support the integration of information contained in textual sources with structured repositories of business data comes from the result we obtained from research carried on within the framework of EC sponsored projects Facile (http://quinary.com/innovation/) and Concerto (http://concerto.pira.co.uk), and from exploitation of the developed techniques.

The FACILE project [Gilardoni 99] developed a concept-based classification and information extraction tool, whose main characteristic is that of integrating linguistic resources and knowledge representation tools. The system, described in detail in [Ciravegna 99], is able to analyze natural language texts and it is based on the integration of pattern matching and information extraction techniques, both supported by a knowledge representation system that enables to model domain concepts and their relations, and to represent linguistic entities linked to conceptual entities.

Text analysis has been implemented as a two-stage process: first domain relevant concepts mentioned in texts are recognised, and then categories (topics) are assigned to the text as a whole. Concepts in this context represent just bits of information contained in texts, that can be conveyed by single sentences (e.g. ‘Company XYZ shares bounced’) while categories classify the text as a whole (e.g. ‘Company XYZ results’ vs ’Stock market behaviour’, depending on whether the overall text in which the sentence above appears deals with the company results or with the overall market).

The first stage is a shallow analysis phase based on pattern matching techniques integrated with a knowledge representation system, modelling both domain objects (e.g., shares, market sectors, balance sheet items) and domain events (e.g., joint ventures, public offers, financial transactions). Concepts are associated with a set of weighted linguistic patterns. The pattern language is similar in spirit to those implemented by search engines (e.g., use of Boolean operators, proximity, optionality), but it is augmented to enable testing of lexical variants and it is also tightly coupled with the knowledge representation system, allowing references to concepts in patterns and exploiting inheritance and relations to reduce complexity. The result of the first phase of analysis is a set of matches of text portions with concepts, associated to scores expressing confidence factors.

The second stage is performed by a rule-based categoriser. The text’s main topic is determined using application-specific heuristics (based on both concept matches and linguistic context), as well as domain knowledge. The shallow analyser and the categoriser use similar but distinct concept lattices: the categoriser lattice directly reflects the end-user’s perception of the application-specific conceptual space.

As an example, given the input text:
Moody's Investors Service Inc said it assigned an A3 rating to the Province of Saskatchewan's C$115 million bond offering that was priced today.

The text analyzer recognizes concepts such as rating_assignment and bond_issue, while categories assigned are: RATING_ON_LOCAL_GOVERNMENT, MOODY_S, BOND_ISSUE_LOCAL_GOVERNMENT.

The de-coupling between recognition of concepts, which mainly requires domain and linguistic knowledge, and assignment of categories, which involves judgement about relevance and knowledge about the context, allows to split the purely linguistic information extraction part from the knowledge driven strategies needed to interpret the real messages conveyed by the text, and to properly place the results in a knowledge structured system. Analysis aspects strictly related to linguistic processing are insulated and masked to the final user, while domain knowledge representation and organization can be left visible and can be handled by trained users.

Figure 5: knowledge-based classification

A sketch of the idea beyond knowledge-based classification is presented in figure 5. Here knowledge about basic domain entities, in the example relationships between stock markets and market indexes (MIB Stock_market_index RELATED_TO Milan_stock_market) enables to understand that a piece of text (reported in Italian as this has been taken by a real case from the web) is dealing with Milan stock market (not mentioned in the text). The same representation layer, other than being used by the classification system to analyze texts, also supports knowledge aware browsing and it is used to relate textual information with structured data, recorded in static or dynamic databases (e.g. the. values of the MIB_index reported in figure 5), on the basis of conceptual references to database schemas.

The described classification system, that has been developed as a joint effort by research institutes providing linguistic technology and user organizations providing business competencies, is currently used by Quinary to support Knowledge Management projects. It has already been deployed for production use in a major Italian financial news agency, where it is used to classify news according to a well defined category structure and to provide integration with structured data, generating dynamic links between stories, financial data, real time quotes, historical data and articles picked from one of the richest proprietary editorial databases in the field. Other than representing concrete savings in terms of human effort, automatic classification carries over a major benefit in uniformity of judgements, allowing integration of heterogeneous material coming from different sources within a common conceptual structure, easing retrieval and sharing.

Fully automated analysis is however not always possible or it could be possible in theory, but requiring both linguistic resources and computational power still out of reach in practice. While fine grained concepts identification, as well as identification of relations, could be made with an accuracy topping 90% in both precision and recall over a fairly complex classification hierarchy in almost real time, identification of relation roles (e.g.
identification of who is the buyer and who is the seller in the description of a financial transaction) requires full NLP, up to discourse analysis. Within the Facile project we have proved feasibility of integrating NLP techniques for information extraction to cope with these issues, but here we still face serious scalability problems. Support of a rich representation is however anyway a prerequisite, whether in trying to provide specific partial solutions, generalised ones as in CYC [Lenat 95], or exploiting mixed approaches.

The concept of a shared and browsable knowledge map is also behind current research activities we are carrying on within the framework of the Concerto project, where we rely on technology deployed within Facile to support human interactions in the task of textual annotation and to provide query capabilities over a database of annotations. In this framework, the human knowledge editor interacts with the system to correct and validate automatically proposed conceptual annotations so that knowledge can be more effectively shared and understood, while new insights gained can be integrated in a common repository [McNaught et all 2000].

6 Conclusions.

In this paper we sustained that effective knowledge sharing and growing inside organization must be supported by an explicit representation of the enterprise knowledge asset. To motivate this opinion, we described how we can focus on well-founded, user and domain-oriented Knowledge Maps to support semantic navigation and knowledge sharing, and also to integrated the “two worlds” of textual and structured information. To describe the Knowledge Map idea in practice, we reported a couple of experiences, that focus on the combination quadrant of the Knowledge cycle, providing knowledge aware tools to support integration of knowledge and information from different sources.

We are currently trying to go beyond the reported experiences, as we believe that a structured representation of the domain can also be the unifying glue of a comprehensive suite of tools, that can benefit from the results of research in different areas to offer support to the whole knowledge cycle.

Given a representation of the domain, online discussions could be driven by domain aware agents with NLP capabilities, that could act as facilitator of the discussions, suggesting new topics and encouraging diverging and converging concepts associations [Van Dyke 99]. Externalization can be helped by automatic extraction of representation structures from text and by intelligent authoring tools that help the author proposing part of existing documents that contain related concepts, suggesting analogies and metaphors. Internalization itself could be enhanced by graphical tools that show each explicit piece of information in the context of the corporate repository.

The integration of domain knowledge with knowledge about the individual actors also paves the way for new tools that could be available in the very near future. According to World Bank [World Bank, 98] interesting opportunities to support externalization could be offered by (still not well developed) electronic workspaces, enabling participation, across time and distance, in project design and repository development. Other on-going research, mainly relevant to the Socialization process, is about Cognitive Social Structures (networks of “who knows who knows who”), based on collaborative filtering, and Cognitive Knowledge Networks (“who knows who knows what”) [Contractor, 99].

References

[van der Spek 97] R. van der Spek, A. Spijker, Knowledge Management: Dealing Intelligently With Knowledge, Kenniscentrum CIIBM, may 1997