

Cooperatively updated knowledge bases as an optimal medium to learn, publish, evaluate and collaborate

[Dr Philippe Martin](#), [Dr Jun Jo](#), [Ms Vicki Jones](#)

School of I.C.T.

Griffith University - PMB 50 Gold Coast MC, QLD 9726 Australia

We argue that cooperatively updated formal or semi-formal knowledge bases (semantic networks) can and should be used as a shared medium for the tasks of researching, publishing, teaching, learning, evaluating and collaborating, and that students, teachers, researchers, decision makers, employees and other information providers or consumers, teams or individuals, could and should be performing very similar tasks and use very similar tools. This does *not* mean that traditional methods (e.g., face-to-face teaching and document publishing) should be *replaced* by knowledge-based methods *but* that they should be *complemented* by them whenever this is possible. To support our claims we have designed a knowledge-based server (WebKB-2) implementing the ideas that we propose and we have begun applying them to our own research domains and to several courses in our university. Although we have implemented this server as a Web server for practical reasons, the same techniques could be applied to a semantic grid, a learning grid or a peer-to-peer network.

Keywords: knowledge representation/sharing/retrieval/learning/evaluation

Introduction

Nowadays, writing a course material, a research article, or a technical documentation, most often involves writing sentences in a (static) document. This is a lengthy process which implies summarising or describing ideas or facts that have already been summarised or described by countless other persons and also implies making rather arbitrary choices and compromises about which information to describe, at which level of detail, in which order, etc. Because information in a base of documents is not semantically organised and includes lots of redundancies, *retrieving or comparing precise information* (e.g., finding and comparing all the particular tools or techniques that are meant to so solve a particular problem) relies heavily on memory, reading and cross-checking. *Learning* from a base of documents has related problems: the lack of details may cause interpretation or understanding problems (this is typical with mathematical books), limit the depth of learning, and a lot of the information may already been known by the reader. To conclude, using documents is a very sub-optimal process for the writers and the readers. Yet, whatever the field of study, there is *currently no well structured semantic networks of techniques or ideas* (even semi-formal ones) that a Web user could (i) navigate to get a synthetic view of a subject or, as in a decision tree, quickly find its path to relevant information, and (ii) easily update to publish a new idea (or the explanation of an idea at a new level of detail) and link it to other ideas via semantic relations. Some information repository projects use formal knowledge bases (KBs), e.g., the [Open GALEN](#) project which has created a KB of medical knowledge, the [QED Project](#) which aims to build a "formal KB of all important, established mathematical knowledge", and the [Halo project](#) (or "Digital Aristotle" project) which has for (very) long term goal a system capable of teaching much of the world's scientific knowledge by (i) [adapting to its students' knowledge and preferences](#) (Hillis, 2004), and (ii) preparing and answering (with explanations) test questions for them (this implies the [encoding of the knowledge in a formal way and meta-reasonings on the problem-solving strategies](#)). Designing such KBs supporting problem-solving is difficult even for teams of trained knowledge engineers (the 6-month pilot phase of Project Halo was restricted to 70 pages of a chemistry book and had encouraging but far-from-ideal results). At the other extreme, [Wikipedia](#) is a great help for students and researchers in many subjects because each of its pages *centralises* and organises the most important information about one particular object (e.g., technique, language or person) and relates it other such objects, thus permitting and enticing the user to delve into details. However, because Wikipedia is not a network of concepts and statements related by semantic relations (e.g., specialization, partOf and argumentation relations), with a record of who authored them, it cannot support an efficient or precise information retrieval, understanding or comparison, it *cannot organise* all the information from teaching materials, research articles and e-mails

between researchers, teachers and students, and it cannot support knowledge update protocols and knowledge filtering mechanisms to permit such an organization to take place and be exploited. This article shows that a semantic network (which does not have to support problem-solving and hence can contain informal nodes) is possible and advantageous to support learning and research. The next section reviews some problems of *less structured approaches*. Then, we present some elements of solutions (i.e., notations, ontologies, cooperation protocols) supported by our knowledge server [WebKB-2](#) (Martin, 2003a) to support the cooperative building of a semantic network by researchers, teachers and students. Finally, the conclusion will summarize our answers to the questions related to the theme of this conference according to the vision of that particular network, however long-term that vision might be.

Issues of some information sharing technologies for learning or research

The smaller the units of information and the more inter-connected via metadata (concept types or semantic relations), the better for information retrieval and automatic exploitation. Ideally, there is no difference between content and metadata. Yet, in order to exploit legacy data (e.g., research articles or learning materials) or because of the difficulties related to knowledge representation, sharing and management for tool designers and users, most [Semantic Web](#) related researches and all the [Learning objects](#) (LOs) related [standards](#) (e.g., AICC, SCORM, ISM, [IEEE WG12](#)) or current projects (e.g., [CANDLE](#), GEODE, MERLOT, VLORN) essentially focus on associating *simple meta-data* to whole documents or big parts of them (e.g., author, owner, terms of distribution, presentation format, and pedagogical attributes such as teaching or interaction style, grade level, mastery level, and prerequisites).

For example, [currently existing LOs](#) are almost never about *one un-decomposable statement only* (i.e., typically, one relation between two concepts: a current LO does not describe one relation between one object and another, such as for example a relation between the "cosine" function and one definition for it or one theorem using that function, or a relation between the "Java" language and one of its features); instead, current LOs are about a set of objects (for example, a typical LO about Java is an "introduction to Java" describing the main features of Java and giving an example of code). This goes against the theoretical goal of LOs: obtaining modules of information as small as possible in order to increase the choice and possibilities of re-using and combining these LOs. There is actually no standard for the granularity of a learning object and, according to the IEEE LTSC (2001), a LO should consist of 5 to 15 minutes of learning material. This seems a failure to distinguish between what individual LOs should ideally be and what convenient "*packages* of such LOs" should be. A package is a set of objects (e.g., concepts, assessment objects) related to a certain curriculum and that a student might be advised to explore in a certain order according to a certain pedagogical strategy. Packages are useful for pedagogical purposes and ease the task of most course designers (precisely because they are ready-made packages). However, current LOs are black-box packages: their content is not explicitly represented and poor metadata is associated to them. They hardly support a semantic search of LOs (and thus an efficient retrieval since for example there may be thousands of LOs on Java) and even less the comparison of these LOs (the more objects implicitly included by the LOs, the more difficult the comparison of these LOs; comparing two concepts or two ideas is possible because one of them can for example be a specialization, an argument or a sub-task of the other but it is extremely rare that two articles or two documentations can be connected by such relations). On the other hand, if LOs were about only one statement and packages were sets of such LOs, they would be much easier to retrieve, compare, modify and combine, especially if the underlying concepts belonged to a well-organised semantic network.

Similarly, the modellings of the preferences and knowledge of students are often very poor (whether the modelling occurs within one pedagogical tool or a learning grid), typically a keyword for each known LO (e.g., "Java") and a learning level for it (e.g., "advanced"). This does not permit a software to know precisely what a student actually knows, as opposed a more fine-grained approach in which all the statements for which a student has been successfully tested on are recorded. This not at all a criticism of the many good softwares related to the Semantic Learning Web (Stutt & Motta, 2004), to the Educational Semantic Web (Devedzic, 2004), or more generally to the Semantic Web, which are almost all based on metadata rather than on precise representations of the content but this is a reminder that the knowledge sharing and exploitation that they achieve and that they theoretically can achieve is necessarily limited compared to what can be achieved when a semantic network is exploited.

Despite the advantages of fine-grained semantic networks (with formal or informal terms and statements as nodes), such networks have rarely been used by *projects aimed to represent lecture materials or relate research ideas*, although [concept maps](#) (or their ISO version, [topic maps](#)) have often been used to represent various topics for teaching purposes (for example in biology by [Leung \(2005\)](#))¹. The hypertext and argumentation system [SYNVIEW](#) (Lowe, 1985) was designed for organising ideas or debates from various articles and hence also index them. Each unit of information was an informal sentence directly or indirectly valued by users and connected to other sentences by predefined or user-invented argumentation relations. In [Robert Horn's knowledge maps about debates such as "Can Computers Think?"](#) (Horn, 2003), the units are single informal sentences as well as one or several paragraphs and images, and these units are related by argumentation relations. In the [Text Outline project](#), which aims to have scholars index the main ideas of classic philosophy books, the units are paragraphs and headings for these paragraphs and are often only implicitly and loosely related. ScholOnto (Buckingham-Shum & al., 1999), which is supposed to make explicit the intellectual lineage and impact of "ideas" as well as their convergences or inconsistencies, allows some conceptual or argumentation relations between concepts, statements or whole research articles. Finally, some [semantic wikis](#) (that is, wikis systems such as Wikimedia but allowing the use of some semantic relations such as `subprocessOf`, `agentOf`, `resultOf`, `correctionOf` and `argumentOf`) begin to appear and be used.

We share the goals of these projects but do not believe these goals can be achieved (at least in a useful and scalable way) without concepts and single statements as main units of information, without a precise and generic argumentation structure, and without knowledge management protocols. If the main units are paragraphs and documents and hence contain several ideas and (counter-)arguments, 1) these ideas and arguments cannot be merged or related to each other into an argumentation structure and hence it is not possible to access the pros and cons for these ideas (and so on, recursively, for the ideas of the pros and cons), 2) it is not possible to associate precise concepts to these sets of ideas or to relate and organise them in principled (non-arbitrary, semantically guided) ways, and to retrieve (or know where to add) a particular idea or an argument (this is well illustrated by the current headings of the Text Outline pilot project), and hence 3) people (researchers, pedagogues, students, experts, employees, directors, product users, etc.) cannot add information (e.g., a statement about a particular software, or the fact that a certain step of a certain proof lack details for understanding or checking purposes) at the "right place" in a semantic network: they can only create new documents and thus add to the amount of "data" that later users will have to deal with, instead of (for the same effort) adding to the "knowledge" of a semantic network. To sum up, the usual problems of finding the outcomes of debates in a mailing list archive still apply to the above cited projects. These outcomes (i.e., the lists of ideas and their related argumentation trees or graphs) *cannot be derived automatically* from the texts: even for people it is often difficult to find the concepts, assumptions and argumentation relations that are implicit in natural language texts or exchanges. Furthermore, such analysis often reveal a surprisingly large proportion of un-argued assertions and actually very few statements serving (counter-)argument functions. Directly updating a precise argumentation tree avoids having to repeat already stored arguments and leads to more carefully written and useful statements.

There are many reasons why researchers, companies, or people in general, choose not to develop or use more knowledge-oriented solutions (Shipman & Marshall (1999) give an interesting survey for hypermedia related tools). We mentioned that there is an inherent difficulty in representing things explicitly and using semantic relations, and that people are not used to that nor therefore to the use of notations (such use therefore requires some training). Most people (even post-graduate students and researchers) have problems understanding the need for explicit structures, and the mere sight of structures as simple as attribute-value pairs can put them off because they find them "ugly and unreadable", especially when textual notations are used to express them; graphical interfaces are more easily accepted but they are more bulky and less practical to use in many situations. Furthermore, many notations and graphical interfaces have strong expressivity limitations that force the users to make premature and arbitrary choices and lead to biased and hard to re-use knowledge representations. This is for example the

¹ However, concept maps are overly permissive and hence do not guide the user into creating a principled and exploitable semantic network (e.g., they can use relations such as "of" and nodes such as "other substances" instead of semantic relations such as "agent" and "subtask", and concept names such as "non_essential_food_nutrient"). Thus, concept maps are often more difficult to understand, retrieve and exploit than when regular informal sentences are used ([Sowa \(2006\)](#) gives commented examples).

case for most argumentation systems or typed hypermedia systems since they only allow to use a few predefined relations or structures (e.g., Toulmin's structures), force nodes to be typed in certain ways, and do not allow to put relations on relations (more generally, they often do not have meta-statement capabilities). The limitations of these systems were reasons why they failed to have a large set of users. On the other hand, the authors of many of these systems (e.g., AAA (Schuler & Smith, 1990)) reported to have made such restrictions to guide the users and avoid scaring them off (this "promoting adoption by a large number of people" reason has also been used by the authors of SYNVIEW, ScholOnto and Text Outline project for explaining why they did not provide more formal or fine-grained features). We agree with the conclusion of Shipman & Marshall (1999) that tools should provide users with generic and expressive structuring features but also convenient default options, and the users should be allowed to describe their knowledge at various levels of details, from totally informal to totally formal so that they can invest time in knowledge representation when and only when they feel the benefits outweigh the costs (and indeed, as detailed by Shipman & Marshall (1999), for many applications, more formalization can be avoided). MacWeb (Nanard & al., 1993) was an example of user-friendly *and* quite expressive knowledge-based hypertext system. We also acknowledge that two more serious problems are that (i) formalization requires an important critical mass of existing knowledge in order to be useful, and (ii) the work of knowledge providers often do not directly benefit them.

Thus, is the particular semantic network based approach that we propose (and that is supported by the solutions described in the following sections) a realistic long-term vision for the teaching and research domains? We believe it is because of the following particularities of these domains: 1) the need for small LOs has been recognised, 2) the representation of research outputs could be directly re-used for learning purposes (by students, industrials or researchers), 3) adding to domain dependent semi-formal knowledge base (KBs) would be an efficient way for researchers to advertise, retrieve, compare and discuss their research ideas and outputs, and as explained in the "Supporting cooperation" section, additions to such KBs could be used for valuating research outputs and researchers, thus offering a better and more rational complement to the current peer-reviewed system which, in many research fields and at least in the Semantic Web related areas, leads to value "easy-to-read research descriptions repackaging old ideas under new names" more than "original and technical research outputs" (since reviewers have to evaluate informal articles instead of semantic networks, since reviewers rarely happen to be experts in the exact domain treated by the article, and since reviewers and authors cannot engage in "structured discussions" to clear misinterpretations and allow or force both of them to provide further rationale), 4) it is the role of teachers and researchers to represent things in explicit and detailed ways (the semantic network approach permits them to do that, without space restrictions nor all the constraints related to information ordering and summarization), and 5) students can complement the semantic network and this provides a way to evaluate their analytic skills.

Notations and ontologies

To represent knowledge, graphical or textual notations are needed and they must be *expressive* (in order not to lead to biased and poorly re-usable representations), *normalising* (to increase the comparability of the representations and hence their retrievability and use for inferencing), *readable* and *concise* (to permit the visualization of big chunks of knowledge without having to browse or scroll, an essential characteristic for the understanding and design of realistically sized KBs). As previously noted, graphical interfaces are more readily accepted by beginners but textual notations are less bulky and often more practical to use: textual knowledge representations or queries can easily be copy-pasted, generated, edited, included in emails or HTML documents, and therefore hyperlinked or used as "commands" by knowledge servers such as WebKB-1 and WebKB-2 (see (Martin & Eklund, 2000) for details on such uses). It should be noted that restricting the expressivity of a notation when its uses or domains of application are unknown is a very poor strategy to avoid completeness, decidability and efficiency issues since 1) such issues can happen even with a relatively low expressivity, 2) the handling of these issues is application-dependant (e.g., for some knowledge retrieval or filtering purposes, efficient graph-matching procedures that ignore the detailed semantics of certain elements can be used (Martin, 2003a), while for other purposes exploiting all the details is essential and tractability is not an issue), 3) the level of complexity of a set of representations can be specified (RDF+OWL has three levels but even its uppermost level, RDF+OWL-Full, cannot be used for representing many uses/forms of disjunction, quantification, meta-statement and collection, and thus for representing most common natural language sentences).

Martin (2002) introduces three notations accepted by WebKB-2 - FL (For-links), FCG (Frame-CG) and Formalised English (FE) - derived from the Conceptual Graph linear form (Sowa, 1984) to improve on its readability, expressivity and "normalising" characteristics (their combination is what made Conceptual Graphs famous). FE looks like some pidgin English but is structurally equivalent to FCG which is a very concise notation that includes constructs for extended quantifiers, meta-statements, functions and various interpretations of collections (hence, it is as expressive as [KIF](#) but is higher level). However, since FCG and FE are too complex for most researchers, teachers and students to use, only FL is introduced here and it seems sufficient for most of what teachers and researchers would like to represent in a formal or semi-formal way. Because FL is restricted to representing simple relationships between concepts or statements (no complex use of quantifications, collections and meta-statements), relations can be presented aggregated onto a same concept instead of having to be written into different statements. Here are examples of English (E) sentences translated in FL, FCG and KIF (the RDF+OWL translation is too bulky to give here). The second example shows the source of each concept and relation.

E: Any human_body is a body and has at most 2 arms, 2 legs and 1 head.
 Any arm, leg and head belongs to at most 1 human body.
 Male_body and female_body are exclusive subtypes of human_body and so are juvenile_body and adult_body.
 FL: human_body subtype of: body, part: arm [0..1,0..2] leg [0..1,0..2] head [1,1],
 subtype: {male_body female_body} {juvenile_body adult_body};
 E: According to Jun Jo (who has for user id "jj"), a body (as understood in WordNet 2.0) has for
 part (as defined by "ph") at least 1 leg (as defined by "fg") and exactly 1 head (as understood by oc").
 FL: wn#body pm#part: fg#leg [0..1](jj) oc#head [1](jj);
 FL: wn#body pm#part: at least 1 fg#leg (jj) 1 oc#head (jj);
 FCG: [wn#body, pm#part: at least 1 fg#leg, pm#part: 1 oc#head](jj);
 KIF: (believer '(forall ((?b wn#body)) (atLeastN 1 '?l fg#leg (pm#part '?b ?l))) jj)
 (believer '(forall ((?b wn#body)) (exists1 '?h oc#head (pm#part '?b ?h))) jj)

The first example uses informal terms. The second example uses formal terms, e.g., "wn#body" is one of the *identifiers* (in the KB of WebKB-2 accessible at www.webkb.org) for the Wordnet concept that has for *names* "body", "organic_structure" and "physical_structure". Thus, another identifier for this concept is "wn#body__organic_structure__physical_structure". Since a name (i.e., an informal term) can have many meanings, it can be shared by many categories (concepts or relations). We created the KB of WebKB-2 by transforming WordNet 1.7 into a genuine lexical ontology and extending it with several top-level ontologies and domain-related ontologies (Martin, 2003b). This work is important for knowledge representation and sharing since people cannot be expected to create an ontology that define all the terms they use and hence connect them to all related terms (e.g., subtypes, supertypes, partOf, etc.) from all other existing ontologies on the Web: one or several natural language ontologies of English such as the one in WebKB-2 are necessary to provide guidance, permit semantic checkings and allow people to simply pick the right category or, as WebKB-2 allows, add it to the ontology if that category does not yet exist (knowledge sharing issues within the KBs and between KBs are discussed in the next section). Regarding the last example, it should be noted that in the KB of WebKB-2 there is currently only one category with name "organic_structure", so using this name instead of "wn#body" would currently not be ambiguous (assuming that the KB of WebKB-2 is the default namespace used for resolving terms). More complex expressions such as "body__physical_structure" would also currently be unambiguous. If, as the W3C proposes, knowledge representations are not directly stored in large KBs but stored in small Web files (e.g., RDF files), the more information are connected to each term, the more the categories (and then, the representations) of a file can be safely connected to or identified with the categories in other files or KBs, and thus re-used. Using a category identifier from a large KB is the most concise way to provide a lot of information. However, looking for identifiers of each category used in a statement may not be something that knowledge providers are prepared to do. A quick alternative is to provide many synonyms (e.g., using the "__" shortcut in FL) and then let knowledge parsers propose category identifiers to the users or let them make automatic guesses (a very unsafe option but sufficient for some applications). This is analogous to the use of "folksonomies" (instead of "ontologies") where people use words (instead of terms from an ontology) as simple metadata for a file (text, image, video, etc.), except that a category is about one thing only whereas people index many different things about a file.

Below is a small portion of [a semantic network which we prepared in 2005 for a post-graduate course in Workflow Management](#) and asked the students to extend (as a replacement for an informal learning journal) using the information from their other learning materials, especially [Aalst & Hee \(2002\)](#). This

source is left implicit in the example below for concepts but it is represented for relations: "b" refers to this book and then a page number is given. The general schema is "CONCEPT1 RELATION1: CONCEPT2 (SOURCE1 INTERPRETER1)" which can be read "According to SOURCE1, interpreted and represented by INTERPRETER1, any CONCEPT1 may have for RELATION1 one or many CONCEPT2". The example below shows only one interpreter: "pm". The information about the terms and their relationships is scattered all over the book. It is also expressed at various levels of details, and sometimes in inconsistent ways (in such cases, different meanings must be represented or choices must be made). Thus, a precise understanding of these terms and their relationships was difficult: we had to create this semantic network in order to gather the relationships and thus better understand the terms.

procedure informal_definition: "a generic piece of work: can be seen as the description of activities"(b/p3 pm),
 subtype: WF_process (b/p33 pm);

WF_process informal_definition: "a procedure for a particular case type"(b/p33 pm)
 "a collection of cases, resources and triggers representing a particular process"(b/glossary pm),
 synonym: workflow (b/p22 pm) WF (b/intro pm) network_of_task (b/p22 pm)
 subtype: ad-hoc_workflow (glossary pm) project (b/p9 pm) task (b/p32 pm),
 description: at least 1 condition (b/p15 pm) process_diagram (b/p15 pm),
 agent: resource (b/p11 pm), part: WF_process (b/p34 pm),
 object: at least 1 case (b/p33 pm), characteristic: complexity (b/p18 pm);

task synonym: atomic_process (b/p32 pm) logical_unit_of_work (b/p32 pm),
 informal_definition: "a process considered indivisible by an actor is a 'task'
 in the eye of this actor"(b/p32 pm),
 subtype: {manual_task automatic_task}(b/p33 pm) semi-automatic_task (b/p32 pm)
 work_item (b/p38 pm),
 example: typing_a_letter (b/p32 pm) stamping_a_document (b/p32 pm),
 responsible_agent: 1 resource (b/p35 pm), parameter: knowledge (b/p7 pm);

work_item informal_definition: "a task for a particular case"(b/p38 pm);

The whole network was stored in a wiki and the students were asked to modify it via the wiki. The absence of syntactic and semantic checking resulted in a large proportion of badly formed or semantically meaningless representations, especially in the training phases (e.g., "part" relations were used between processes and physical entities, instead of between processes or between physical entities). [All the details \(including the final network\) are available at http://www.phmartin.info/wf/](http://www.phmartin.info/wf/). This semester, three other courses will be tested ("Knowledge Representation" "Introduction to Multimedia Development" and "System Analysis and Design") and WebKB-2 will be used by the students. The students' understanding of the content of the course and analysis skills will be tested both via their additions of concepts but also via their additions of informal statements. Indeed, FL has recently been extended to include all the necessary constructs for "structured discussions", which makes it a good (and often more expressive) alternative to the notations used in argumentation systems (e.g., AAA and gIBIS). More details on some ways to evaluate the students' contributions are given in the next section.

Below is an excerpt from a ["structured discussion" about the use of XML for knowledge representation](#), a topic which leads to recurrent debates on many knowledge related mailing lists. The parenthesis are used for two purposes: (i) allowing the direct representation of links from the destination of a link, and (ii) representing meta-information on a link, such as its creator (e.g., "pm" or "fg") or a link on this link (e.g., an objection by "pm" on the use of an objection link by "fg", without stating anything about the destination of this link). The content of the sentences and the *indentation* in the example below should permit the understanding of these two different uses. (Note that in this example the creators of the statements are left implicit but that prefixes such as "pm#" could be used exactly as in the first example). The use of dashes to list joint arguments/objections (e.g., a rule and its premise) should also be self-explanatory. The use of specialization links between informal statements may seem odd but such links are used in several argumentation systems: they are essential for modularising purposes and for checking the updates of argumentation structures, and hence guiding or exploiting these updates (e.g., the (counter-)arguments for a statement also apply to its specializations and the (counter-)arguments of the specializations are (counter-)examples for their generalizations). Few argumentation systems allow links on links ([ArguMed](#) is one of the exceptions) and hence most of them force incorrect representations of discussions. Even fewer provide a textual notation that is not XML-based, hence a notation readable and usable without an XML editor or a graphical interface.

"XML is useless for knowledge representation, exchange or storage"
 argument: ("using XML tools for KBSs is a useless additional task"
 argument: "KBSs do not use XML internally" (pm,
 objection: "XML can be used for knowledge exchange or storage" (fg,
 objection: "it is as easy to use other formats for knowledge exchange or storage" (pm),
 objection: "a KBS (also) has to use other formats for knowledge exchange or storage" (pm)))
)(pm);

"XML can be used for knowledge exchange or storage"
 argument: - "an XML notation permits classic XML tools (parsers, XSLT, ...) to be re-used" (pm)
 - "classic XML tools are usable even if a graph-based model is used" (pm),
 argument of: ("a KRL should (also) have an XML notation",
 specialization: "the Semantic Web KRL should have an XML notation" (pm),
 specialization of: "a KRL (Knowledge Representation Language) can have an XML notation" (pm),
)(pm);

For each domain field, an initial domain ontology must be created to incite people to enter knowledge and guide this entering for knowledge sharing purposes. We have begun representing the features of ontology editor related tools and this led us to partially model other related domains. This required modularising information into several files to support readability, searches, checking and systematic input. In order to be generic, we created six files (see <http://www.webkb.org/kb/classif/> for details): [Fields of study](#), [Systems of logic](#), [Information Sciences](#), [Knowledge Management](#), [Conceptual Graph](#) and [Formal Concept Analysis](#). The last three files specialize the others. Each of the last four files is divided into sections, the uppermost ones being "Domains and Theories", "Tasks and Methodologies", "Structures and Languages", "Tools", "Journals, Conferences, Publishers and Mailing Lists", "Articles, Books and other Documents" and "People: Researchers, Specialists, Teams/Projects, ...". This is a work in progress: the content and number of files will increase but the sections seem stable. As soon as we feel that our representation of major ontology editor related tools is sufficient to guide the classification of other such tools, we will invite the authors of these tools to complete the classification. There is currently a demand for a comparison of the dozens existing ontology editing tools and this demand cannot be satisfied by the few informal surveys that have been done, e.g., ([Denny 2004](#)). The result may then serve as a platform for organising research ideas related to these tools or for categorising other tools.

Supporting cooperation between knowledge providers

Here, only asynchronous cooperation is considered since it both underlies and is more scalable than exchanges of information between co-temporal users of a system. The most decentralized knowledge sharing strategy is the one envisaged by the W3C for the "[Semantic Web](#)": many very small KBs, more or less independently developed and thus partially redundant, competing and very loosely interconnected. There now are many tools to align concepts from different ontologies; these tools are necessarily far from perfect although they can be sufficient for certain applications; Euzenat & al. (2005) give an evaluation. Thus, despite these tools, small ontologies have problems similar to those we listed for documents: (i) finding the relevant ontologies, choosing between them and combining them require commonsense (and hence is difficult and sub-optimal even for a knowledge engineer, let alone for softwares), (ii) a knowledge provider cannot simply add one concept or statement "at the right place" and is not guided by a large ontology into providing precise objects that complement existing objects and are more easily re-used, and (iii) the result is more or less lost to others and increases the amount of "data" to search.

A more knowledge-oriented strategy is to have a knowledge server permitting registered users to update a single large KB on a domain. We know of only two knowledge servers having special protocols to support cooperation between users: [Co4](#) (Euzenat, 1996) and [WebKB-2](#). (Note: most servers support concurrency control and many servers support users' permissions on files/KBs but "cooperation support" is not so basic: it is about helping knowledge re-use, preventing most conflicts and solving those detected by the system or users). The approach of Co4 is based on peer reviewing; the result is a hierarchy of KBs, the uppermost ones containing the most consensual knowledge while the lowermost ones are the private KBs of the contributing users. The approach of WebKB-2, which is based on a KB shared by all its users, leads to more relations between categories or statements from the different users and may be easier to handle (by the system and the users) for a large amount of knowledge and a large number of users. Details can be found in (Martin, 2003a) but the next paragraph summarizes its principles.

Each category identifier is prefixed by a short identifier for the category creator (who is also represented by a category and thus may have associated statements). Each statement also has an associated creator and hence, if it is not a definition, may be considered as a belief. Any object (category or statement) may be re-used by any user within her statements. The removal of an object can only be done by its creator but a user may "correct" a belief by connecting it to another belief via a "corrective relation" (e.g., pm#corrective_specialization). (Definitions cannot be corrected since they cannot be false; for example, the user "fg" is perfectly entitled to define fg#cat as a subtype of wn#chair; there is no inconsistency as long as the ways fg#cat is further defined or used respect the constraints associated to wn#chair). If entering a new belief introduces a redundancy or an inconsistency that is detected by the system, it is rejected. The user may then either modify her statement or enter it again but connected by a "corrective relation" to each belief it is redundant or inconsistent with: this allows and makes explicit the disagreement of one user with (her interpretation of) the belief of another user. This also technically removes the cause of the problem: a proposition A may be inconsistent with a proposition B but the belief that "A is a correction of B" is not technically inconsistent with a belief in B. (Definitions from *different* users cannot be inconsistent with each other, they simply define different categories/meanings). If choices between beliefs have to be made by people re-using the KB for an application, they can exploit the explicit relations between beliefs, for example by always selecting the most specialized ones. The query engine of WebKB-2 always returns a statement *with* its meta-statements, hence with the associated corrective relations. Finally, in order to avoid seeing the objects of certain creators during browsing or within query results, a user may set filters on these creators, based on their identifiers, types or descriptions. Since inconsistency detection and "corrections" can be done by the users, this approach to encourage re-use, precision and object connectivity also works when the KB contain informal terms or statements.

The above described knowledge sharing mechanism of records and exploits annotations by individual users on statements but does not record and exploit any measure of the "usefulness" of each statement, a value representing its "global interest", acceptance, popularity, originality, etc. Yet, this seems interesting, especially for semi-formal discussions: statements that are obvious, un-argued, or for which each argument has been counter-argued, should be marked as such (e.g., via darker colours or smaller fonts) in order to make them less visible or invisible and discourage the entering of such statements. More generally, the presentation of the combined efforts from the various contributors may take into account the usefulness of each statement. Furthermore, given that the creator of each statement is recorded, (i) a value of usefulness may also be calculated for each creator (and displayed), and (ii) in return, this value may be taken into account to calculate the usefulness of the creators' contributions; these are two additional refinements to both detect and encourage the production of argued and interesting contributions. Ideally, the system would accept user-defined measures of usefulness for a statement or a creator, and adapt its display of the repository accordingly. Below, we present the *default* measure implemented in WebKB-2. For now, we only consider beliefs: we have not yet defined the usefulness of a definition. To calculate the usefulness of a belief, we first associate two more basic attributes to the belief: its "state of confirmation" and its "global interest".

- 1 The "state of confirmation" is equal to 0 if the belief has no argument nor counter-argument (examples of counter-argument relation names: "counter-example", "counter-fact", "corrective-specialization"). It is equal to 1 (i.e. the belief is "confirmed") if (i) each argument has a state of confirmation of 0 or 1, and (ii) there exists no confirmed counter-argument. It is equal to -1 if the belief has at least one confirmed counter-argument. It is also equal to 0 in the remaining case: no confirmed counter-argument but each of the argument has a state of confirmation of -1.
- 2 Each user may give a value to the interest of a belief between -5 and 5 (the maximum value that the creator of the belief may give is 2). Multiplied by the usefulness of the valuating user, this gives an "individual interest" (this may be seen as a particular multi-valued vote). The "global interest" of a belief is defined as the average of its individual interests (thus, this is a voting system where more competent people in the domain of interest are given more weight). A belief that does not deserve to be visible (e.g., because it is a particular case of a more general belief) is likely to receive a negative global interest. We let each user give an interest value; we do not take into account the way the belief is generalized by or connected to (or included in) other beliefs because interpreting an interest from such relations is difficult. For example, a belief used as a counter-example may be a particular case of another belief but is nevertheless very interesting as a counter-example.

Finally, the *usefulness of a belief* is equal to the value of the global interest if the state of confirmation is equal to 1, and otherwise is equal to the value of the state of confirmation (i.e. -1 or 0: a belief without argument has no usefulness, whether it is itself an argument or not).

Our formula for a user's usefulness is: sum of the usefulness of the beliefs from the user + square root (number of times the user voted on the interest of beliefs). The second part of this formula acknowledges the participation of the user in votes while decreasing its weight as the number of votes increases. (Functions decreasing more rapidly than square root may perhaps better balance originality and participation effort). These various measures are simple but should incite the users to be careful and precise in their contributions and give arguments for them: unlike in traditional discussions or anonymous reviews, careless statements here penalise their authors. Thus, this should lead users not to make statements outside their domain of expertise or without verifying their facts. (Using a different pseudo when providing low quality statements does not seem to be an helpful strategy to escape the above approach since this reduces the number of authored statements for the first pseudo). On the other hand, the above measures should hopefully not lead "correct but outside-the-main-stream contributions" to be under-rated since counter-arguments must be justified. Finally, when a belief is counter-argued, the usefulness of its author decreases, and hence she is incited to deepen the discussion or remove the faulty belief.

The above supports do not solve the problem caused by the fact that one piece of information can be of interest in many domains and the fact that one knowledge server clearly cannot support the knowledge sharing of all Web users; this problem is "which knowledge server should a person choose to query or update?". A server has to be specialized or act as a broker for more specialized servers. However, if each server periodically checked *related servers* (more general servers, competing servers and slightly more specialized servers), imported the knowledge relevant to its domain and, for the rest, stored pointers to those servers, it would not matter much which server a Web user attempts to query first. For example, a Web user could try to query or update any general server and, if necessary, be redirected to use a more specialized server, and so on recursively (at each level, only one of the competing servers would have to be tried since they would mirror each other). If a Web user directly tried a specialized server, it could redirect her to use a more appropriate server or indicate which other servers may provide more information for her query (or directly forward this query to these other servers). Integrating knowledge from other servers is certainly not obvious but this is a more scalable and exploitable approach than letting people and softwares select and re-use or integrate dozens/hundreds of (semi-)independently designed small KBs/ontologies. A more fundamental obstacle to the widespread use of this approach is that many industry-related servers are likely to make it difficult or illegal to mirror their KBs. However, other approaches would likely suffer from that too. Instead of replicating knowledge between servers on the Web, the above approach could be used as a knowledge replication mechanism between the machines of a particular peer-to-peer network or learning/semantic grid, with the assumption that each machine or user has its own knowledge server. Knowledge sharing is simpler in this case since knowledge replication mechanisms can be systematically performed according to the goals and particularities of that grid, and the above replication mechanism can be seen as a backbone for integrating into each user's KB the updates made in other user's KBs on objects shared by these KBs (more precisely, the updates/additions of statements which for inference purposes are important for these shared objects).

Conclusion

We argued that cooperatively updated formal or semi-formal knowledge bases (semantic networks) can and should be used as a shared medium for the tasks of researching, teaching, learning, evaluating and collaborating since all these tasks rely on the same information retrieval, comparison and sharing subtasks which are better supported by large KBs. We believe that KBs created by researchers can be directly re-used for teaching ("package of LOs" would be selected sets of statements and assessment objects, possibly navigated according to pedagogical strategies), that all interested users (researchers, teachers, students, product users, etc.) can and should contribute to those shared same KBs (directly, or indirectly via the replication mechanisms amongst KBs; the more contributions, the better), and that all contributions and contributors should be evaluated (via argumentation relations, votes and possibly the assessment objects associated to statements) to support this sharing by encouraging and rewarding "good" contributions/contributors and permitting knowledge filtering. We have recognised the need to allow and support the representation of information at various level of details (the smallest and the more formal the units of information the better, but structuring can be incremental and should fit the knowledge providers' constraints). Reading and correctly using formal notations may be difficult but less than programming languages or musical notations; like the Web, the creation of a web of knowledge will be a social process.

References

- van der Aalst, W.M.P., & van Hee, K.M. (2002). *Workflow Management: Models, Methods, and Systems*. MIT press, Cambridge, MA, 2002.
- Buckingham-Shum, S., Motta, E. & Domingue, J. (1999). *Representing Scholarly Claims in Internet Digital Libraries: A Knowledge Modelling Approach*. Proceedings of ECDL 1999 (pp. 423-442), 3rd European Conf. Research and Advanced Technology for Digital Libraries, Paris, France, Sept. 1999.
- Denny, M. (2004). *Ontology Tools Survey, Revisited*. <http://www.xml.com/pub/a/2004/07/14/onto.html>, July 14, 2004.
- Devedzic, V. (2004). *Education and the Semantic Web*. International Journal of Artificial Intelligence in Education, 14, pp. 39-65, 2004.
- Downes, S. (2001). *Learning Objects: Resources For Distance Education Worldwide* International Review of Research in Open and Distance Learning, Vol. 2, No.1, Oct. 1st 2001.
- Euzenat, J. (1996). *Corporate memory through cooperative creation of knowledge bases and hyper-documents*. Proceedings of 10th KAW, (36)1-18, Banff, Canada, Nov. 1996.
- Euzenat, J., Stuckenschmidt, H., & Yatskevich, M. (2005). *Introduction to the Ontology Alignment Evaluation 2005* Proceedings of K-Cap 2005 (pp. 61-71), workshop on Integrating ontology, Banff, Canada, 2005.
- IEEE LTSC (2001). *IEEE Learning Technology Standards Committee Glossary*. IEEE P1484.3 Glossary Working Group, draft standard 2001.
- Hillis, W.D. (2004). *"Aristotle" (The Knowledge Web)*. Edge Foundation, Inc., No 138, May 6, 2004.
- Horn, R. (2003). *Mapping Great Debates: Can Computers Think?*. <http://www.macrovu.com/CCTGeneralInfo.html>
- Leung, J. (2005). *Concept Maps On Various Topics*. http://www.fed.cuhk.edu.hk/~johnson/misconceptions/concept_map/concept_maps.html
- Lowe, D. (1985). *Co-operative Structuring of Information: The Representation of reasoning and debate*. International Journal of Man-Machine Studies, Volume 23, Number 2, pp. 97-111, August 1985.
- Martin, P., & Eklund, P. (2000). *Knowledge Indexation and Retrieval and the Word Wide Web*. [IEEE Intelligent Systems](#), special issue "Knowledge Management and Knowledge Distribution over the Internet", May/June 2000.
- Martin, P. (2002). *Knowledge representation in CGLF, CGIF, KIF, Frame-CG and Formalized-English*. Proceedings of [ICCS 2002](#), 10th International Conference on Conceptual Structures (Springer Verlag, LNAI 2393, pp. 77-91), Borovets, Bulgaria, July 15-19, 2002.
- Martin, P. (2003a). *Knowledge Representation, Sharing and Retrieval on the Web*. Chapter of a [book titled "Web Intelligence"](#), (Eds.: N. Zhong, J. Liu, Y. Yao; Springer-Verlag, pp. 263-297), Jan. 2003.
- Martin, P. (2003b). *Correction and Extension of WordNet 1.7*. Proceedings of [ICCS 2003](#) (Springer Verlag, LNAI 2746, pp. 160-173), Dresden, Germany, July 2003.
- Nanard, J., Nanard, M., Massotte, A., Djemaa, A., Joubert, A., Betaille, H., & Chauché, J. (1993). *Integrating Knowledge-based Hypertext and Database for Task-oriented Access to Documents*. Proceedings of DEXA 1993 (Springer Verlag, LNCS Vol. 720, pp. 721-732), Prague, 1993.
- Newman, S., & Marshall, C. (1992). *Pushing Toulmin Too Far: Learning From an Argument Representation Scheme*. Technical Report SSL-92-45, Xerox Palo Alto Research Center, Palo Alto, CA, 1992.
- Schuler, W., & Smith, J.B. (1990). *Author's Argumentation Assistant (AAA): A Hypertext-Based Authoring Tool for Argumentative Texts*. Proceedings of ECHT'90 (Cambridge University Press, pp. 137-151), INRIA, France, Nov. 1990.
- Shipman, F.M., & Marshall, C.C. (1999). *ormality considered harmful: experiences, emerging themes, and directions on the use of formal representations in interactive systems*. Computer Supported Cooperative Work, 8, pp. 333-52, 1999.
- Sowa, J.F. (1984). *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, Reading, MA, 1984.
- Sowa, J.F. (2006). *Concept Mapping*. <http://www.jfsowa.com/talks/cmapping.pdf>
- Stutt, A. & Motta, E. (2004). *Semantic Learning Webs*. Journal of Interactive Media in Education, Special Issue on the Educational Semantic Web, 10, 2004.