

A Network of Semantically Structured Wikipedia to Bind Information

[Dr Philippe Martin](#), [Dr Michel Eboueya](#), [Dr Michael Blumenstein](#) and [A.Prof. Peter Deer](#)

School of I.C.T. - Griffith University - PMB 50 Gold Coast MC, QLD 9726

Australia

pm at phmartin dot info

Abstract: In this article we show how a network of cooperatively updated semi-formal knowledge bases with adequate knowledge valuation, organization and filtering mechanisms can solve the numerous problems of Wikipedia (lack of structure and evaluations of the information, limitation to overviews, edit wars, etc.) and be a good support to learning, research and more generally information sharing and retrieval.

Introduction

When researchers, lecturers, students, decision makers, or people in general, search information on the Web or in libraries, their goal is rarely to find documents, it is to find the various existing possible answers (definitions, facts, techniques, products, people, etc.) to a problem or information need, the respective advantages and drawbacks of these answers for a given goal, and their relations to each other (e.g., a statement or technique may be represented as superseding, specialising, correcting, proving or illustrating another statement or technique from another author). However, no search engine or Web site currently provides a comprehensive and well organised semantic network of information about a particular subject, from general concepts/techniques/statements understandable by anyone to very precise or technical concepts/statements/techniques generated by experts in the subjects.

Some information repository projects use or intend to 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](#) which has for very long term goal the design of a “[Digital Aristotle](#)” which would be capable of teaching much of the world's scientific knowledge and using it to solve classic exercises. However, even when not aimed to support problem solving, designing completely formal KBs is inherently a difficult and very time-consuming exercise (even for trained knowledge engineers) that current generic KB systems (KBSs) still do not guide well. Furthermore, formal KBs only designed for problem-solving are difficult to understand and are far from ideal medium to browse for learning purposes.

Wikipedia is currently the only Web site that provides good overviews about a large number of subjects. It is a great help for students and researchers because each of its pages *centralises* the most important information about one particular subject (e.g., technique, language or person), and relates it to other particular subjects thus permitting and enticing the user to delve into details. For example, reading the [page on Java \(Sun\)](#) is helpful to Java novices because it lists the various components of the Java platform and their various names and abbreviations; such information is very difficult to extract and synthesize by reading the many documents listed on the [Sun's Java Web site](#). However, Wikipedia is extremely informal and loosely structured: it is not a network of *objects (concepts or statements)* linked by semantic relations (e.g., specialization, partOf and argumentation relations), with a record of who authored these objects. This leads to several problems. *First*, it is sometimes difficult to understand precisely how the implicitly or explicitly referred objects are related and how they compare to each other (for example, from the informal sentences of the [Wikipedia articles about logic](#) it is difficult to understand which theory is a part or a refinement of other theories). *Second*, Wikipedia is limited to storing overviews: it cannot scale to permit the organisation and retrieval of all the information contained in teaching materials and research articles or e-mails. Partitioning into different repositories is clearly a poor alternative to a (semantically organised) larger repository and there are expectedly a lot of redundancies and relatively few cross-references between the pages of Wikipedia and those of its sister-projects (Wiktionary, Wikiquote, Wikibooks, etc.). *Third*, it cannot support knowledge update protocols, voting mechanisms and knowledge filtering mechanisms based on the relationships between the objects and on who authored these objects or voted on their originality, veracity or “usefulness”. Since it does not store such

meta-information, and thereby does not lead the users to be precise and permit each of them to retrieve and see what she wishes according to her current goals and applications, Wikipedia simply allows anyone to delete anything that she disagree with. This leads to “edit wars”, does not permit to trust the information, and makes many experts reluctant to add information. It should be noted that the more classic strategy of letting a committee of experts in each subject decide what should or should not be included in a repository is as limiting and nearly as arbitrary since the content of the repository will only reflect the current goals and beliefs of the members of that committee.

In this article we show how a network of cooperatively updated semi-formal KBs can solve the problems of Wikipedia and support learning, research and more generally information sharing and retrieval (ISR). After a comparison of our approach with other ones, examples of simple textual notations will be given, followed by a summary of our solutions to support knowledge organization, evaluation and filtering within a shared KB. Our approach is supported by our KB server [WebKB-2](#) (Martin, 2003).

Rationale For The Proposed Semi-formal Approach

In an ideal information repository (ideal for ISR purposes), any conceivable object (e.g., the driver seat of my car, a sentence, a type of tools) should be referable in an unambiguous way, each statement about that object should unambiguously refer to it, and statements should be related to each other (e.g., if one rephrases, specializes or arguments another one, a relation between them should explicit that point). Thence, all statements directly or indirectly associated to an object can be easily found and compared. The problem is that such a repository has to be a formal KB, that no current technology would permit to build it automatically from Web documents (if only because they are not precise enough) and that a formal KB is difficult for people to update.

One problem is to have people explicit relations between terms (formal or not, quantified or not) and between statements, and therefore to follow a notation that makes these relations explicit. The mere sight of textual notations as simple as attribute-value pairs can put many persons off because they find them “ugly and unreadable”; graphical notations are more easily accepted but they are more bulky and less practical to use in many situations. However, simple statements (e.g., the use of a relation with no associated meta-statements) can be expressed using relatively simple notations (we give examples in the next section) that are less difficult to learn and use than musical notations, programming languages, or most XML-based languages. Furthermore, their adoption can be incremental: someone can first use nodes composed of many sentences, and then, when the need to merge or compare the content of various nodes emerge, those nodes can be decomposed. Nonetheless, two very classic mistakes should be avoided: 1) allowing relation names to be any linguistic expression, 2) restricting the expressivity of all the notations that a system accepts and the number of concept types and relation types that can be used.

The first mistake can be found in [concept maps](#) (or their ISO version, [topic maps](#)). They are so permissive that they do not guide the user into creating a well-defined and exploitable semantic network and are often more difficult to understand, retrieve and exploit than regular informal sentences. For example, as in the concept maps used by [Leung \(2005\)](#) to teach biology, they can use relation names such as “of” (instead of semantic relations such as “agent” and “subtask”) and node names such as “other substances” (instead of concept names such as “non_essential_nutrient”). [Sowa \(2006\)](#) gives other commented examples. One of the *minimal conventions* listed by Martin (2000) for knowledge representation and sharing (KRS) is to use singular nouns for concept and relation names. This is an important convention to follow, especially for relations, even if informal terms are used.

The second mistake (“restricting what can be expressed”) is common in hypermedia or argumentation systems (as noted by Shipman & Marshall (1999)) and in knowledge representations languages that aims to be general (e.g., RDF), as noted by Patel-Schneider (2005). Indeed, it leads to biased and hard to re-use knowledge representations. The completeness, decidability and efficiency issues, or how to handle elements such as sets and modalities, are *application dependant issues* (e.g., for some knowledge retrieval purposes, efficient graph-matching procedures that ignore the detailed semantics of certain elements can be used, while for other purposes exploiting all the details is essential and tractability is not an issue). Many argumentation system authors (e.g., Schuler & Smith (1990)) made such restrictions to “guide the users”, “avoid scaring them” and hence “promote adoption”, but that strategy proved counter-productive. On the other hand, two extremely unstructured hypermedia systems have been really successful: Wikipedia and the Web itself. This is why we agree with the conclusion of Shipman & Marshall (1999): the adoption of notations should be allowed to be progressive, various notations should be accepted (some simple, some very expressive), and the users should be allowed to define new concept types and relation types. MacWeb (Nanard & al., 1993) was an example of user-friendly and quite expressive knowledge-based hypertext system.

A bigger problem is to have people use formal terms for relation types and concept types/individuals, since this is a time consuming task even when, as with [CYC](#) or WebKB-2, the used KB server proposes a large lexical ontology that can be browsed or queried in various ways and that provides URIs or short unique identifiers for each category expressing the meaning of a word. The W3C still envisages people building a “[Semantic Web](#)” (Shadbolt & al., 2006) by creating their own small private KBs and defining their categories with respect to some categories in other persons’ KBs (or ontologies). This approach means that for most terms each knowledge provider has to find, understand and combine ontologies on the Web (amongst a large set of more or less independently developed and thus partially redundant and very loosely interconnected ontologies) and has to use URIs to refer to relevant formal terms (category identifiers) from some of these ontologies. That process is far less optimal in time, precision, reliability and re-usability than if a large KB server is used since (i) a server can quickly give access to a large choice of precise and well-organised categories for what the user wishes to represent, (ii) the large ontology of a server permits it to do many cross-checks on the definitions and uses of all the terms, and to guide the insertion of new terms, (iii) in a large KB, a newly added category or statement is added “at the right place” in the KB and thus is easily accessible and re-usable by other persons. Like creating Web documents, creating private KBs increases the amount of redundant, and unconnected *data* to search. There now are many tools to align concepts from different ontologies but they necessarily have very poor results compared to cooperatively built ontologies (although we acknowledge that for some applications those results can be sufficient; Euzenat & al. (2005) give an evaluation). To sum up, using a large KB server is a minimum but we still cannot expect people to look up formal terms for each word in their statements. However, a KB server can provide formal term suggestions (based on other used terms or on provided synonyms), especially explicit relations with known names are used between the terms (for KRS, a small set of relation types can and should be used over and over; thus, for example, the ontology of WebKB-2 has currently more than a hundred thousand formal terms for concept types/individuals but only one thousand relation types and less than fifty of them are actually sufficient for representing most sentences; most of the other relations are for organizational purposes or come from various integrated ontologies but we discourage their uses).

Several other problems derive from the fact that many users edit the same KB. This requires protocols and mechanisms to valuate, filter and organise knowledge. The fourth section presents our solutions. However, these mechanisms do not solve the problem caused by the fact that a piece of information can be of interest in different domains and that one knowledge server 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 that has a general KB (e.g., a semi-formal version of Wikipedia) or is domain dependant but not specialised (as Open GALEN is) would have to point to more specialized servers in the results of searches by browsing or querying. However, if each server periodically checks *related servers* (more general servers, competing servers or slightly more specialized servers) to import the knowledge relevant to its domain/scope and, for the rest, stores pointers to these servers, it does not matter much which server a Web user attempts to query or update first. For example, a Web user can query any general server and, if needed, 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 tries a specialized server, she can also be redirected to use a more appropriate server. Automatically integrating knowledge from other servers is certainly not obvious but this is eased by the organisation and large size of the KBs and their similarities (since they import knowledge from each other). It is thus much easier than automatically integrating dozens/hundreds of (semi-)independently designed small KBs/ontologies.

In order to exploit legacy data (e.g., research articles or learning materials) or because of the difficulties related to KRS, most [Semantic Web](#) related projects 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, distribution terms, presentation format, and pedagogical attributes such as teaching or interaction style, grade level, mastery level and prerequisites). Most of the *projects aiming to relate ideas*, e.g., [SYNVIEW](#) (Lowe, 1985) or argumentation systems in general, [Robert Horn's knowledge maps](#) (Horn, 2003) and the [Text Outline project](#), use informal relations or a restricted set of predefined relations between sentences or whole paragraphs. ScholOnto (Buckingham-Shum & al., 1999), which aims 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, in most Semantic Learning Web projects (Stutt & Motta, 2004) and Educational Semantic Web projects (Devedzic, 2004), the modellings of the preferences and knowledge of students are often very poor (whether that 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 a criticism of these projects (they are important as “first steps” or to exploit legacy data) but a reminder that the KRS that they achieve or that they theoretically can achieve is necessarily limited compared to what can be achieved with a much more fine-grained approach. We believe our approach is a realistic long-term vision (that is, can become popular), at least for teaching and research purposes, because 1) Wikipedia, its advantages, its drawbacks, are now well known by many persons, 2) the users of Wikipedia can be invited to use our system to do what they cannot do in Wikipedia (i.e., organise information, be technical, solve edit wars, etc.) and we have acquired the domain name *stwikipedia.org* so that adding “st” in the URL of any page in Wikipedia will point to the corresponding object in our system, 3) the need for small LOs has been recognised, 4) the representation of research outputs could be directly re-used for learning purposes (by students, industrials or researchers) and, conversely, semantically organised LOs could be used as a basis for organising research outputs, 5) adding to domain dependent semi-formal KBs would be a much more efficient way for researchers to advertise, retrieve, compare, value and discuss their research ideas and outputs than (only) using the current methods based on reading and writing articles and messages in mailing lists (e.g., recurring or superficial discussions are strongly reduced with our approach; this is already true with most argumentation systems), 6) 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 the information ordering and summarization tasks that have to be done when linear texts have to be written), and 7) students can complement the semantic and this provides a way to evaluate their analytic skills.

Examples of Simple Textual Notations

Our system accepts three notations - FL, FCG and FE (Martin, 2000) - derived from the Conceptual Graph (CG) linear form (Sowa, 1984) to improve on its readability, expressivity and “normalising” characteristics (their combination is what made CGs famous). FE looks like some pidgin English (it is a [controlled language](#) with an LALR(1) grammar) but is structurally similar to FCG which is a concise notation that includes constructs for extended quantifiers, meta-statements, functions and various interpretations of collections (hence, FCG is as expressive as [KIF](#) but is higher level). However, since even FCG and FE are too complex for most researchers, teachers and students to use, only FL is really introduced below. FL seems sufficient for most of what teachers and researchers would like to represent in a formal or semi-formal way. Since FL is restricted to representing simple relations (*links*) 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, FE and KIF. The second example shows the source of each relation, statement and category (and hence uses formal terms for categories).

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 J.B. (who has for user id "jb"), a body (as understood in WordNet 2.0) has for
 part (as defined by "pm") at least 1 leg (as understood by "fg") and exactly 1 head (as understood by "oc").

FL: wn#body pm#part: fg#leg [0..1](jb) oc#head [1](jb);
 FL: wn#body pm#part: at least 1 fg#leg (jb) 1 oc#head (jb);
 FCG: [wn#body, pm#part: at least 1 fg#leg, pm#part: 1 oc#head](jb);
 FE: `A wn#body has for pm#part at least 1 fg#leg and for pm#part 1 oc#head'(jb).
 KIF: (believer '(forall ((?b wn#body)) (atLeastN 1 '?l fg#leg (pm#part '?b ?l))) jb)
 (believer '(forall ((?b wn#body)) (exists1 '?h oc#head (pm#part '?b ?h))) jb

The next example is a small portion of a semantic network that we prepared for a post-graduate course in Workflow Management and asked the students to extend as a replacement for an informal learning journal. This FL example uses informal terms but shows the source and the creator of each relation. For example, “b/p18” refers to the page 18 of the used textbook (Aalst & Hee, 2002) and “pm” is here the only creator of the relations (they are *his interpretation* of the sources). Creating this semantic network helped us (and then the students) better understand the terms and their relationships because that information was scattered all over the book, expressed at various levels of details, and sometimes in inconsistent ways (in such cases, different concepts had to be distinguished).

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)
"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);

The next FL example is a small excerpt from a "structured discussion" about the use of XML for knowledge representation, a topic leading to recurrent debates on many knowledge related mailing lists. It should be noted that 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. The creators of the statements are left implicit. 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 modularisation purposes and for checking updates on 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);

Supporting Knowledge Organization, Evaluation and Filtering Within a Shared KB

We know of only two knowledge servers having special protocols to support asynchronous cooperation between users: [Co4](#) (Euzenat, 1996) and [WebKB-2](#). (Most servers support concurrency control and many servers support users' permissions on files 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 is based on a KB shared by all its users. Details can be found in Martin (2003) but the next paragraph summarizes the principles.

Each category identifier is prefixed by a short identifier for the category creator (who is represented by a category and thus may be described). Each statement also has an associated creator and, if it is not a definition, is considered as a belief. Any object (category or statement) may be re-used by any user within her statements. The removal of an object may 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 belief 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: an assertion A may be inconsistent with an assertion 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). 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 beliefs. The query engine of WebKB-2 always returns a statement *with* its meta-statements, hence with the associated corrective relations. 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. To sum up, contradictories beliefs from different users or statements for inconsistent theories can be put into a same consistency KB (by prefixing the category identifiers with their source identifiers and using corrective relations between beliefs) and various modules (e.g., the private KBs of Co4 or a lattice of theories (Sowa, 2000)) can be generated according to various criteria but our approach leads to the entering of more relations between the categories or statements from different sources and is easier to handle (by the system and the users) for a large amount of knowledge and a large number of users.

The above described knowledge sharing mechanism 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 is interesting for a KB and 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 smaller fonts) in order to make them invisible or less visible and thus discourage the entering of such statements. More generally, the presentation of the combined efforts of 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 should accept user-defined measures of usefulness for a statement or a creator, and adapt its display of the repository accordingly. We will propose various options and combinable functions to allow this but the general template will be based on the *default measure* currently implemented in WebKB-2 and that we now present. 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 (hence no “counter-example” or “corrective_specialization” relation). 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 since using the way a belief is connected to other beliefs seems too risky. For example, a belief used as a counter-example may be a particular case of another belief but may nevertheless be 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 default formula for a user's usefulness is: sum of the usefulness of the beliefs from the user + square root of the number of times the user voted on the interest of beliefs. The second part of this formula values 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. Although independently developed, the above approach appears to be an extension of the version of [SYNVIEW](#) (Lowe, 1985). In this hypertext system, statements had to be connected by (predefined or user-invented) relations and each statement was valued by users (this value and another one calculated from the value of arguments and counter-arguments for the statement was simply displayed near the statement to “summarize the strengths assigned to the various items of evidence”).

In addition to the previous features, for the users to cope with large amounts of information, the system should guide the entering and structuring of knowledge and present it in a structured way. A lengthy list of query results may be automatically structured into specialization/partOf hierarchies. If this does not provide enough structure, additional schemes should be used: for example, the results can be grouped according to common characteristics; this is a categorization task which in the general case may be difficult to solve optimally and efficiently but important concepts (such as “process”, “physical object”, “person”, “civil status”, “recreational activity”, etc.) provide cues for natural groupings. In the ontology of WebKB-2 records a number of categories are marked as important but this is not yet used for automatically structuring results. However, the schemas associated to high-level or some medium-level categories are used for generating combinable knowledge entering menus (i.e., a menu can be called from another to build a bigger assertion or query). Any schema added by any user is exploitable. Such menus may ease and help to normalise knowledge entering. The organisation of knowledge into modules (seen as an additional system of indexation) is also helpful for coping with large amounts of information. Furthermore, for each domain field, an initial domain ontology must be created to incite people to enter knowledge and also guide this entering for knowledge sharing purposes. Our on-going experiment in organising the features of ontology editors led us to partially model other related domains and create the following modules to support readability, searches, checking and systematic input: Fields of study, Systems of logic, Information Sciences, Knowledge Management, Conceptual Graph, and Formal Concept Analysis. The last three modules specialize the others and, when presented as Web documents, each of them is divided into sections such as “Domains”, “Properties, Measures, Models and Laws”, “Processes, Tasks, Techniques and Methodologies”, “Process agents”, “Data structures and Formats”, “Interfaces and Languages”, “Software”, “Hardware”, “Journals, Conferences, Publishers and Mailing Lists”, “Articles, Books and other Documents” and “People: Researchers, Specialists, Teams/Projects, etc.”.

Conclusion

In his description of a “Digital Aristotle”, Hillis (2004) describes a “Knowledge Web” in which researchers could add ideas or explanations of ideas “at the right place”, and suggests that this Knowledge Web should “include the mechanisms for credit assignment, usage tracking, and annotation that the Web lacks”, thus supporting a much better re-use and evaluation of the work of a researcher than via the system of article publishing and reviewing. Hillis did not give any indication about such mechanisms but our approach seems to provide a template for them. Ideally, each user should be allowed to adapt the system’s “default measure” in order to specify what she wants to see according to her own combination of criteria about (i) the authors of the information (e.g., a user may want to discard information that has not been approved by a certain committee or has been authored by a person who does not have a PhD or has authored something that she thinks inappropriate) and (ii) the information itself (e.g., a user may want to see a piece of information only if it has been given arguments that attracted no counter-arguments or if it has been voted as “relevant” and “very original” by persons who, according to that user, have not so far demonstrated bad taste). This approach relies on explicit and fined-grained but possibly informal representations of concepts and statements and their relationships. It is in-line with the constructivist and argumentation theories and can be seen as a particular implementation and support of the [Critical Thinking](#) approaches and [Brandom's model of discursive practice](#) (Brandom, 1998). The use of whole paragraphs for statements should be allowed as *first steps*

towards more explicit modellings, better comparisons of arguments and deeper debates. WebKB-2 cannot detect redundancies, refinements and inconsistencies amongst a set of informal statements but people can and our approach exploits this. This approach can therefore be used as a solution to the problems of Wikipedia and to support very useful complements, replacements or indexes to research articles, emails, blogs and other independently written documents. It can also be used to develop teaching materials collaboratively and let students complement them as part of assessments or to indicate parts that they found unclear. Pedagogical strategies could still be independently associated by teachers to subsets of the KB (as they currently are to large LOs) in order to indicate to students what they should explore and remember, which exercises they should do, and in which order. If each user's success to exercises associated to a statement is also stored in the KB (in a public or anonymous way), this may be used as a precise tool to compare and select people on their knowledge about a certain set of concepts or statements, and this may also be used in user-defined functions to evaluate knowledge providers and thus also for filtering information.

References

- van der Aalst, W.M.P., & van Hee, K.M. (2002). *Workflow Management: Models, Methods, and Systems*. MIT press, Cambridge, MA, 2002.
- Brandom, R. (1998). Action, Norms, and Practical Reasoning. *Noûs*, Volume 32, Supplement 2, pp. 127-139(13), October 1998.
- Buckingham-Shum, S., Motta, E. & Domingue, J. (1999). Representing Scholarly Claims in Internet Digital Libraries: A Knowledge Modelling Approach. *ECDL 1999* (pp. 423-442), 3rd European Conf. Research and Advanced Technology for Digital Libraries, Paris, France, September 1999.
- Devedzic, V. (2004). Education and the Semantic Web. *International Journal of Artificial Intelligence in Education*, 14, pp. 39-65, 2004.
- Euzenat, J. (1996). Corporate memory through cooperative creation of knowledge bases and hyper-documents. *10th KAW*, (36)1-18, Banff, Canada, Nov. 1996.
- Euzenat, J., Stuckenschmidt, H., & Yatskevich, M. (2005). Introduction to the Ontology Alignment Evaluation 2005. *K-Cap 2005* (pp. 61-71), workshop on Integrating ontology, Banff, Canada, 2005.
- 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. (2000). Conventions and Notations for Knowledge Representation and Retrieval. *ICCS 2000* (Springer, LNAI 1867, pp. 41-54), Darmstadt, Germany, August 14-18, 2000.
- Martin, P. (2003). Knowledge Representation, Sharing and Retrieval on the Web. *Web Intelligence* (Eds.: N. Zhong, J. Liu, Y. Yao; Springer-Verlag), Chapter 12, pp. 263-297, January 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. *DEXA 1993* (Springer Verlag, LNCS Vol. 720, pp. 721-732), Prague, 1993.
- Patel-Schneider, P.F. (2005). A Revised Architecture for Semantic Web Reasoning. *PPSWR 2005* (LNCS 3703, pp. 32-36), Principles and Practice of Semantic Web Reasoning: Third International Workshop, Dagstuhl Castle, Germany, September 11-16, 2005.
- Shadbolt, N., Berners-Lee, T. & Hall, W. (2006). The Semantic Web Revisited. *IEEE Intelligent Systems* 21(3) pp. 96-101, May/June 2006.
- Shipman, F. & Marshall, C.C. (1999). Formality considered harmful: experiences, emerging themes, and directions on the use of formal representations in interactive systems. *CSCW*, 8, pp. 333-52, 1999.
- Schuler, W. & Smith, J.B. (1990). Author's Argumentation Assistant (AAA): A Hypertext-Based Authoring Tool for Argumentative Texts. *ECHT'90* (Cambridge University Press, pp. 137-151), INRIA, France, Nov. 1990.
- Sowa, J.F. (1984). *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, Reading, MA, 1984.
- Sowa, J.F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks/Cole, Pacific Grove, CA.
- 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.