

The comparison with most similar systems and the connections with ideas from various fields. An example from economy.

An appendix to the text: "A method for building a scalable centralized ontology - An idea"

Dariusz Daćko

April 26, 2008

updated on July 27, 2009

www.consensualknowledge.net

1. Comparisons

Below I will describe briefly the references to existing ideas and computer systems which were not described in the main text. I hope that they will let you place the presented idea among existing ones more easily.

Philosophy – consensus theory of truth

(http://en.wikipedia.org/wiki/Consensus_theory_of_truth)

The theory was invented by a German philosopher Jürgen Habermas (http://en.wikipedia.org/wiki/J%C3%BCrgen_Habermas). According to him when there is a general consensus about something then it is the truth. After a long enough conversation and after exchanging all arguments people should reach a consensus. He has been calling this situation as "ideal conversation". In the theory he has used terms and rules similar to these used in the presented idea:

- 1) consensus concept (I think it is the most important part of this theory),
- 2) considering that the whole mankind can constitute the group involved in a discussion,
- 3) "communicative action" concept (http://en.wikipedia.org/wiki/Communicative_action).

People can perform various actions: communicate with each other, "advertise" themselves, ideas, products; do something for themselves etc. Only action of communication is entirely selfless. Currently there are few selfless actions in the Internet. Most websites are created in order to serve their owners. Selfishness spoils the quality of information being communicated. Wikipedia is quite selfless and this is why it is good. On Internet forums people have a lot of freedom, and that is why they often exchange emotions instead of information. In the proposed system, thanks to organization of information into an ontology, there will be no place for exchanging negative emotions. The discussion will be more matter-of-fact and constructive.

The philosophers negating the consensus theory of truth say that a consensus among all people or among any of their subgroups is unfeasible, because there are too many people and there is no technical possibility to conduct a constructive discussion. But in my opinion, the idea presented above enables such a discussion. I do not think that philosophers have ever considered such an implementation of a discussion since even computer scientists have not done it so far.

You can easily notice that computer scientists have not considered this solution from Open Mind approach from MIT, which is described further in the article. This system lacks consensus which results in vandalism and a lot of work during the process of adding knowledge to the system.

Most computer scientists that inquire into ontologies think that it is impossible to reach a consensus because of cultural diversity etc. However, one of the most prominent philosophers does not agree with them. It means that it is worth checking the correctness of the presented idea. It can be said that the presented idea is a computer implementation of the consensus theory of truth and a “Public sphere” concept. The public sphere (http://en.wikipedia.org/wiki/Public_sphere) is “an area in social life where people can get together, discuss freely, identify social problems and through that discussion influence political action.”

Management – collective intelligence (http://en.wikipedia.org/wiki/Collective_intelligence). The term is described in the main text. The presented idea could be a potentially interesting example of collective intelligence.

Open Space Technology (http://en.wikipedia.org/wiki/Open_Space_Technology) is a way of conducting meetings enabling effective exchange of people’s views. It can be used for example as an alternative to conferences. The main rule is that there is no prearranged agenda for the meeting. The agenda is established at its beginning and everybody can introduce any interesting problem, describing it shortly. Then, each participant of the meeting enrolls on different subjects that he is interested in. Afterwards, the participants go to separate rooms to discuss them. Thanks to it, everyone can express his opinion and potentially reach a common conclusion.

Also in proposed system everyone who is interested can express his opinion. It can contribute to speeding up the process of exchanging knowledge among people.

Cyc ontology (<http://en.wikipedia.org/wiki/Cyc>) is the biggest ontology built so far. Its weak point is that it is built only by a few tens of people (as far as I remember by about 20-30). Until now about one million pieces of information have been stored in it in this way. It is still very little in comparison with a number of commonsense pieces of information, which is known to each human and is estimated to be a half billion. I think that a number of pieces of information that can be useful for every human, including specialists, can be even as big as few or few tens of billions. It is obvious that collecting such a big amount of information requires joint work based on the idea of Web 2.0 (http://en.wikipedia.org/wiki/Web_2). Wikipedia, which is the most prominent example of Web 2.0 service, is a good argument for scalability. Within a short period of time, it surpassed the Encyclopedia Britannica edited by a narrow group of specialists.

Citizendium project (<http://en.wikipedia.org/wiki/Citizendium>) is a system similar to Wikipedia which is implemented by its co-author. Citizendium is going to be a credible version of Wikipedia. Credibility can be archived, among other things, by authentication of users on the basis of their personal data. A similar solution is used in Allegro, a Polish online auction website, which requires authentication on the basis of the home address. In order to be authenticated, the user should provide an address at which he can receive a letter. Then Allegro sends a letter with a generated key to the given address. When the user receives the letter he is to read the key and pass it to Allegro website. If the keys are the same, Allegro assumes that address provided by the user belongs to him. Such a solution was introduced to increase credibility of users. In Citizendium, it was introduced to increase credibility of users and to eliminate vandalism. A similar solution is also proposed in the idea presented in the main text. In Open Mind there is no authentication which results in vandalism.

Both Wikipedia and Citizendium are encyclopaedias. It means that if we want to add some information to them we must add it to some article in such a way that it will constitute an

integrated whole with the currently existing information. Because of this, content is added to encyclopaedias in top-bottom manner. One cannot add to Wikipedia a single piece of information without writing an entire article about a certain concept. It slows down the process of collecting information since many people willing to add information resign from writing time-consuming articles.

Global brain (http://en.wikipedia.org/wiki/Global_brain) is another concept presenting an idea similar to the “Public sphere”, however, this time, not from the point of view of politics and society but rather from the point of view of cognitive science and Semantic Web (http://en.wikipedia.org/wiki/Semantic_Web). The “Global brain” is a “metaphor for intelligent network formed by humans together with the knowledge and communication technologies that connect them.” Three aspects of this concept were described on the above-mentioned page in Wikipedia. One of them, the encyclopaedic one, puts emphasis on the development of universal knowledge network. Wikipedia approximates this vision in the best way from all currently existing systems. However, Wikipedia is only an encyclopaedia, that is, a resource of general knowledge. Therefore, the detailed information will never be included there. The Semantic Web could be considered a “Global brain”, however, at present, the scientists are sceptical about it due to the lack of ideas for its implementation.

Open Mind (http://en.wikipedia.org/wiki/Open_mind), implemented by MIT, is a system with a goal, to some extent, analogous to the Cyc system. Cyc is a formalised ontology. The aim of Open Mind is also to create a database, however, this time, less formalised and based more on the common sense, for example of following kind: “a bed is used for sleeping”. The authors of Open Mind gave up formalisation (and thus, the possibility to understand this knowledge by computer) in order that the knowledge can be introduced by a large number of Internet users. In this way, over a million pieces of information were collected. Unfortunately, the lack of formalisation and the lack of control over the users lead to redundancy of a large number of entries, and there are numerous cases of vandalism. For example, several percent of the users claim that “a bed cannot be used for sleeping”, what causes that the statement “a bed can be used for sleeping” is marked as uncertain. Because there is no control over the users, a large number of votes are necessary for each piece of information, in order to make it credible. About one hundred persons had to make an effort of voting on the piece of information “a bed can be used for sleeping”, nevertheless this piece of information is still uncertain. In the presented system, one vote would be enough to add a credible piece of information. This means that within the same period of time several tens times as many information as in Open Mind should be collected. Additionally, using such a system would be much more pleasant. We will know that our contribution is very important, since it does not duplicate with others and there is no need for its verification. What is more, the system would have a practical application, so many people would be interested in its development. We can conclude that at the time needed for the development of the Open Mind system, it would be possible to introduce into the proposed system the common sense knowledge of size comparable with the man’s knowledge. The specialist knowledge, unknown to laymen, is more formalised and less fuzzy, so its collecting would be easier. The system would be a credible source of information, therefore the scientists would be willing to use it in contrast with the present Wikipedia and Open Mind. This also creates a chance for recording the specialist knowledge in the system, such as, for example: properties of the various sorting algorithms, properties of the various programs for file comparison, or mathematical and astronomical knowledge.

Mindpixel (<http://en.wikipedia.org/wiki/Mindpixel>) was an initiative similar to Open Mind. Unlike Open Mind, it had some control over the users. However, the control was connected with a cost, as each piece of information had to be added by at least twenty users.

To sum up there are various ideas for building a large knowledge base and a few attempts of their implementation. There were attempts to create a database by various methods and using various rules. It seems that it would be worth building it, using the best principles applied in the respective systems. From the Cyc system we could use the structuring of knowledge, so that the knowledge added to the system would not be redundant and we would be able to find and use it easily. From the Citizendium project we could use the principle of representing the users by real persons. Thanks to it, we will obtain the certainty of the stored information without the cost related to this as is now in the Mindpixel. As you can read in the main text, this will probably not lead to identification of author of the information. It is also possible that thanks to cryptographic techniques, in the case of breaking into the system, we could avoid stealing all entries written by a given person.

It seems that: avoidance of redundancy of introduced information, the certainty of the introduced information and the speed of their introduction thanks to the credibility of the users, could make it possible to build a system storing much more information than the presently existing ones.

HITS algorithm (http://en.wikipedia.org/wiki/HITS_algorithm)

The Hyperlink-Induced Topic Search algorithm was a hyperlink analysis algorithm used by some search engines to rate popularity of web pages in days when there were not too many web pages in the Internet. Algorithm was defining two kinds of pages: hubs and authorities. The hubs were web pages with many links to other pages (for example web directories http://en.wikipedia.org/wiki/Web_directory). The latter were called authorities. Popularity of ordinary web page (i.e. authority) depended on number and quality of hubs linking to it. It was proportional to sum of qualities of hubs linking to it. On the other hand the quality of hub was evaluated on the basis of number and evaluation of authorities it was linking to. It was proportional to sum of popularities of hubs it was linking to. The more good hubs were linking to a page, the greater was the popularity of the page. The more good pages were included in a hub, the better was its quality. As it can be simply noticed, the quality of hubs and the popularity of authorities were evaluated using so called mutual recursion (http://en.wikipedia.org/wiki/Mutual_recursion). In the Internet the HITS algorithm was replaced by simpler PageRank algorithm.

The presented method of evaluating the credibility of information is analogous to HITS algorithm, where the analogous concepts are as follows:

- what is evaluated: *fine-grained information instead of web pages (authorities)*
- who does evaluate: *people instead of hubs*
- what quantity is evaluated: *credibility of information instead of popularity of web pages and credibility of people instead of quality of hubs*

The presented method of evaluating information credibility has two additional elements:

- people can link to information positively or negatively while hubs can link to authorities only positively,
- if some information has positive and negative incoming links then the argumentation and discussion is required, potentially leading to only one type of links.

Below is a table showing the analogy more precisely. Elements marked with grey colour are not directly related with evaluation of quality. Concepts that differ in source and destination of analogy are marked in italics.

	HITS Hyperlink-Induced Topic Search	presented method “Argumentation-Induced Information Credibility”
what is evaluated	web pages → popularity/authority (or hubness)	<i>fine-grained</i> information → <i>credibility</i>
method of information searching/localizing	TF-IDF (in PageRank)	self-created ontology
quality of piece is evaluated on the basis of	evaluation of <i>hubs</i> that link to it	evaluation of <i>people</i> (credibility) <i>voting for or against</i> the information, maybe also based on this if it is fresh information and on the basis of “credibility of evaluation of credibility”
additional process	–	<i>discussion through arguments</i> aiming at reaching consensus in votes
quality of evaluator is evaluated on the basis of	evaluation of pieces it links to	evaluation of information (credibility) he voted <i>for or</i> <i>against</i> (only for information that is not fresh, also take into account “the credibility of evaluation of information credibility”), <i>decrease evaluation if required argument was not given</i>

2. The application in the economy

In main text I described implications for economy resulting in particular, from the semantic description of the products. Below I will describe one more aspect related to obtaining the automatic evaluation of the product.

Each product is described by the EAN number, consisting of: the country identifier (e.g. 590 for Poland), producer’s identifier (e.g. 0500 for Hortex¹), a five-digit product identifier of a given producer (e.g. 02427 for 1-litre black currant nectar produced by Hortex) and one control digit. The examples of codes are as follows:

Hortex Black Currant Nectar 1l	590 0500 02427 6
Hortex Ruby Grapefruit Nectar 1l	590 0500 02431 3
Hortex Multivitamin Nectar 1l	590 0500 02432 0
Hortex Multifruit Nectar 1l	590 0500 02438 2
Hortex Peach Nectar 1l	590 0500 02440 5
Hortex Banana Nectar 1l	590 0500 02441 2

The specification of the Black Currant Nectar provided on its packaging includes, among other thing, its ingredients:

- black currant juice (25%),
- water,
- sugar,
- natural flavour from black currants.

If a name of Polish product includes the word “nectar” or “juice”, then we have a legally guaranteed certainty that it really contains natural juice and not chemical compounds of unnatural origin. However the ingredient called “natural flavour from black currants” may make us wonder. Why is the juice itself not enough, that it needs to be additionally flavoured? How is the flavour obtained? There will be a possibility to find out about it in the proposed

¹ A company in Poland that produces frozen foods and fruit juices, <http://en.wikipedia.org/wiki/Hortex>, www.hortex.com.pl/en/

system. In the field “components” of the concept “Hortex Black Currant Nectar”, we will select the fourth component and check its meaning. Of course, in this case, such precision seems to be quite an exaggeration, however what if we take, for instance, cured meats? How many preservatives and colourings does the cured meat chosen by us contain on the basis of its nice appearance? If a lot, then perhaps we could choose such cured meat which does not contain them? Maybe a different producer does not use so many preservatives and colourings? But which one? Coming back to the nectar example, it may turn out that the nectar itself is healthy, but its packaging is not biodegradable or the producer uses non-ecological technologies in the production process.

The EAN code will facilitate finding a piece of information, which is of interest to us. Instead of searching for an object representing the product interesting us in the presented system, it is enough to find it by the EAN code. We can imagine a Bluetooth module or software for mobile phone camera automatically reading the barcode from the packaging, and sending the EAN code to the mobile phone (equipped with the Internet connection) or computer. Instead of displaying all the product properties, it is possible to display the collective information whether the product is: a) healthy, b) its production is ecological, c) whether it meets any other criteria previously set by us etc. Thanks to the use of ontology, the determination of these criteria will be easy. It will be enough that the consumer specifies that he wants to buy “health food”, and on the basis of ontology, a decision will be made on what is healthy. For example, an Ennn preservative is healthy and Emmm is not, or at least less healthy than the other one. If our criteria are not fulfilled, it could be signalled by a red indicator on the Bluetooth module, mobile phone or in the computer. At the same time, a better product could be proposed. At the time of industry and marketing development, such a solution can have a significant impact on producing better, healthier and more ecological products by the producers, and that the consumers will buy more consciously and from better producers. This may influence both the people’s health and the environmental protection. The idea could be applied for products from all the industries: food industry, electronic industry etc.

Changes in the text

- July 27, 2009 – the section about HITS was added