A method for building a scalable centralized ontology
An idea

Dariusz Daćko
January 26, 2008
www.consensualknowledge.net

Abstract

In this document an idea for building a centralized and scalable ontology will be presented. In contrast to ontologies built by small teams of experts, the presented ontology will be built by whole community. Scalability will be achieved by treating all the elements of the ontology homogenously as pieces of information. Information correctness will be evaluated by reaching consensus as a result of constructive and structured discussion. This approach is consistent with the idea of collective intelligence that emerges in the scientific community. The proposed ontology will be coherent, complete, credible and easy to use, contrary to a set of small ontologies proposed for the Semantic Web. Any information can be stored in the ontology, not only this stored in currently existing ontologies. Thus, the proposed ontology should be treated as an integrated system for knowledge acquisition, creation, storage and delivery, not only as a knowledge organization system. Therefore, the proposed ontology can be considered as a centralized version of the Semantic Web. It will have all the features of the Semantic Web and more and it seems that it could be easily implemented, contrary to the traditional approach to the Semantic Web. Thanks to the form of stored knowledge it can be used both by people and computers contributing to the development of society and artificial intelligence. This document should be treated only as an idea and not as a scholarly paper, because it should be reviewed by professionals who have competence in professional ontology building and logic.

Keywords: building a centralized scalable ontology, community-driven ontology building, collective intelligence ontology

Description of the method

Building a big centralized ontology is not currently planned[1], because if the size of a central ontology increases, the control over it becomes more and more difficult [1, 3]. Apart of it while building a central ontology it is difficult to reach a consensus among all interested people [2]. Therefore, in the Semantic Web, it is planned to use many ontologies built separately. But decentralized approach has one disadvantage, because concepts from different ontologies do not have to be coherent. A way of using many noncoherent ontologies in the Semantic Web is not currently known [1, 3]. In this text a method for building a centralized and scalable to large size ontology will be presented.

The idea is to build an ontology, not by knowledge engineers or any other small group of people, but by whole community as it is in Wikipedia. In ontology, in contrast to Wikipedia, knowledge is stored in fine-grained form i.e. not as articles but as concepts and relations between them. Keeping Wikipedia articles in good quality is difficult. But it seems that it is possible to build an ontology editor, which can ensure that Internet users using it will add to an ontology only correct relations and they will define concepts correctly i.e. they will add only true information. The editor, due to its features, will enable to reach consensus among all users adding information. Thus the control over ontology will be passed from a

[1] After writing this text it turned out that myOntology project also tries to build such an ontology.

central institution to Internet users. Due to potentially huge number of users, the editor will enable building very large ontology in a short time. The ontology will not have the disadvantages of a distributed approach. The ontology could contain all concepts used by humankind and huge amounts of information about these concepts. The rest of the text will present features of the editor which will ensure the self-control of the users as well as reaching consensus among them.

The possibility of discussion is used to reach consensus. This method of solution is used because of belief that if a big group of people have the possibility of a long enough discussion about some statement, they can reach consensus after some time (shorter or longer if statement is not obvious) about this, whether the statement is true or false or that it is not possible to logically conclude the correctness of the sentence on the basis of the knowledge they have. In the latter case we can determine the probability of the correctness on the basis of the votes of each person. Such a situation can be observed in the scientific community. In science there is a consensus about most of sentences i.e. it is possible to say if they are correct or false. Consensus is a result of debates in scientific communities. The correctness of some sentences is not known because there is a lack of some premises. But in these cases most scientists agree with this situation i.e. that it is not possible to determine correctness. In such cases some scientists opt for true and some opt for false. We can derive from this statistical conclusion that it is not known whether a sentence is true but 70% of scientists say that it probably is. The probability of correctness can be calculated on the basis of the prestige of each scientist i.e. statements of scientists with a higher prestige (in some domain) will have a higher weight.

In contrast to science, in everyday life discussion rarely takes place, so reaching consensus is difficult. If one does not discuss with his opponent one cannot convince him to his point of view. The ontology editor should change this.

The example of the scientific society shows that reaching consensus is possible. The ontology editor should make this process easier. There is no central institution controlling scholarly publications to ensure the development of knowledge. The scientific community is able to come to correct conclusions. Similarly, central management of ontology is not required to ensure its high quality – the community should control itself as the scientific community does. It seems that to obtain this, only a few rules described below should be obeyed in the ontology editor.

Fine-graining in ontology causes that discussion about the statements it contains will be easier. The ontology editor will be used in traditional way i.e. as other ontology editors in case when only one person will use it. Editing ontology by more people will be a natural extension of edition by only one person. At the beginning, a person who wants to add information to the ontology should add the concepts included in the statement being added. Let us assume that someone wants to add information that the concept/object “Bill Clinton” was president of (some) state in 1992-2000. To add this information one should check if following concepts are defined in the ontology: “Bill Clinton”, “president of a state”, “years 1992-2000”. If one of these concepts is not present it should be added and defined by relations and features uniquely identifying it on the basis of other concepts. Alternatively the definitions can be asserted in a natural language but they must uniquely identify the concept. When all the concepts of the statement are present in the ontology one can add information by joining “Bill Clinton” concept (subject) by relation “is” with “president of a state” concept (object) and “1992-2000 years” concept (adverbial). In the same way one can add other information. Alternatively, information can be asserted in more natural language by connecting object “Bill Clinton” and “president of a state” by relation “is” with following description in a natural language: “[was] in 1992-2000”, however this should not be done for

such simple cases. Let us assume that the definitions of these three concepts and information connecting them were added by a user registered with U1 identifier.

Other users can use the content of the ontology at once, for example they can read it. The content of the ontology was also read by user U2. User U2 noticed that not all information was correct. The definition of object "Bill Clinton" as a person is correct but in U2's opinion, Clinton was president of a state only in the years 1993-2000. It means that in U2's opinion the aforementioned relation is not correct, because it states that Clinton was president also in 1992. U2 does not agree with it, so he adds his evaluation of correctness of the relation i.e. he evaluates it as false. It can be thought that other users can also add their evaluations thus they will vote on the correctness of the information. After casting of some votes we will know more or less if the information is considered by people to be true or false. We will probably come to some interesting results. But the editor should work in a different way because of its philosophy. In the case of a difference of opinion between two users they should discuss it. The user who does not agree with information existing in the ontology and marking it in the system has to give a convincing argument. User U1, i.e. user that added the information which was absent, can also give a justification/argument for the information he added, but it is not required. A reliable source of information can be an example of the argument. By a reliable source of information I mean a source that is not questionable. For example Wikipedia is not a reliable source of information[2] because its content can be falsified without any problem (but as a last resort it can also be used). Examples of credible sources of information are: articles, books, other scholarly publications, documents, official home pages of people under consideration etc. For example user U2 can add following sentence as an argument: "Information is false because Clinton was not the president in 1992. Source: biography on Bill Clinton's official website. Reasoning: In the biography it is only written that Clinton was sworn into office in 1993. From assumption about completeness of the biography I conclude that Clinton was not the president in 1992." If U1 adds arguments for added information, U2 would have to add an opinion about these arguments as well, i.e. show that they are false or they do not necessarily indicate that information added by U1 is true. So if information added is not obvious it is worth to add arguments confirming it. User U1 is obliged to add an opinion about user U2's arguments within a certain fixed time (not too long). U1 can be notified about this for example by SMS or by e-mail. U1 can agree with U2's arguments and change the evaluation of the inserted information to false. U1 can also disagree with U2's arguments. Arguments should rather be written (and they rather will be) in the same way as all other information. So U2's arguments should rather be written as two pieces of information:

1) Information "Bill Clinton was not a president in 1992" that will be an argument against U1's original information.

2) Information "The biography on Bill Clinton's official website does not contain information that Clinton was president in 1992", this will be an argument for the above-mentioned information with additional comment about rule of conclusion: "on the basis of completeness of biography".

Such list of pieces of information can be named in general as a debate/argumentation tree. It is named as a tree because one piece of information can have more than one argument. In argumentation, information should be logically proved to be true or false. Fine graining of information makes debating easier for statements difficult to prove. For such statements we can add arguments for and against original information. Every argument will be a piece of information whose correctness should also be proved in case it is not obvious. So for every new piece of information a new tree of argumentation can be built. It can be imagined that argumentation trees can be very large for statements difficult to prove.

[2] Citizendium is a project aiming at the creation of a reliable Internet encyclopedia.

To defend his original statement, U1 should prove falsity of one of information in U2's argumentation tree or he should prove falsity of U2's reasoning. Thus arguments used, as well as reasoning used, should be difficult to call into question. For example U1 can say that in his opinion the information: "Bill Clinton was not a president in 1992" is false because in The Times in 1992 there was an article describing Clinton as the new president of the USA. Furthermore U1 says that concluding information 1) from information 2) is not correct. U1 adds to the ontology information about the article in The Times, votes against correctness of information 1) and sets the aforementioned information as an argument against information 1). Apart of it he votes against correctness of reasoning information 1) from information 2). He can also add source of original information he added at the beginning.

In this case U2 cannot give any other arguments against U1's arguments and he must accept them. Thus he should delete entry from the ontology that describes information 2) as an argument for information 1) and mark information 1) as false. Thanks to discussion, the users have reached consensus. Other users will add their information and finally they will discover the truth.

It could seem that argumentation trees are a necessary evil but in my opinion they are not. Trees themselves can be a very interesting source of information especially in the case of unobvious statements.

Users of the editor will be evaluated on the basis of the number of correct and incorrect pieces of information they have added to the ontology. Statements finally evaluated as false will decrease the user's credibility, and statements finally evaluated as true will increase the credibility (but by a smaller value than false statements). If a user is obliged to complete arguments for information he previously added, but he does not complete them within certain time, his evaluation will be decreased too. User registered in the system will be permanently associated with physical person. Therefore people will avoid adding false information to the system. Association will be used only for administrative purposes. It will not be visible to other users.

Statement credibility will be also evaluated on the basis of credibility of users voting for its correctness. Votes against statement correctness will decrease the statement credibility, especially if users voting in this way will be credible (in some domain). During a debate about information, i.e. when votes for and against exist, information credibility is quite low since the system should base on consensus of all credible users (or their large majority). It is also possible to use two measures of information credibility: a probability of information correctness and a credibility of evaluation of this probability.

After some time, the correctness of any information should be evaluated. Of course discussion about some statements can take longer.

It is worth noticing that the system could have up to tens of millions users. Adding of information will be easy because of its fine-grained form. Because of this, building of the ontology can be very fast. The ontology can store not only concepts but also their instances. For example one can add to the ontology object representing himself and define it by describing himself uniquely. Next he can add information about himself for example that he graduated at Cambridge University as lawyer. Such information can be available for every user or only to specified groups, for example to users being hyponyms of "lawyer" concept. Thus his prospective employer can find information about him including his CV.

Another example of fine-grained class is the "Nokia 6021" concept being a hyponym of cellular phone. In this case we do not even have to add a definition of this concept, because it is uniquely defined by its name, i.e. its name is unambiguous. Next we can add information about this concept, for example that it is a hyponym of "modem" concept (i.e. the phone can be used to a connect computer to the Internet), set values of "GPRS data transfer support" and "EDGE data transfer support" properties to "true" and set "EDGE class" property to 6 (EDGE

is faster than GPRS and class is related to transmission speed). All these properties are inherited from “cellular phone” concept (and were defined in “cellular station” concept). Defining (adding) a property for a concept can be seen as adding information that a concept has this property. This information can be treated like other regular information, so others can disagree with it. Information about concepts and instances can be used in auction services such as eBay. Thanks to the ontology it will not be required to describe all the features of a cellular phone we want to sell. It will be enough to choose the phone model in the ontology and give only the features of phone instance we own. Thanks to this, buyer will be able to see complete and credible information about all products in all auctions. Thus both selling and buying products will be simplified. The instance of the phone that we want to sell can be also added to the ontology. This will simplify selling because seller will automatically see all the features that he can fill, for example: condition (new, good, poor), year of purchase etc. As we can see using the ontology could be very beneficial.

If there is information about particular phone models in the ontology we will be able to find out for example which “receiving cellular stations” both support EDGE data transfer and are modems. They could be cellular phones, smartphones or dedicated modems. Dependence finding algorithm, on the basis of prices of devices being currently sold, can make us aware of the fact that if we search for a device to use as a GSM modem then we should buy a cellular phone with EDGE support and not a dedicated modem, because the latter is much more expensive. Phone models that are Pareto optimal in multicriteria decision support process can be presented to buyers. All the features essential to buyers will be taken into consideration, for example: phone price, transmission speed etc. (this was written in vague manner but it seems that it would be easy to implement it).

It is also possible to add statements and features of well-known people to the ontology, for example politicians. Thanks to this political debates would be more constructive.

If a user is unsure of the information that he adds he can set the probability of information correctness to a value lower than 100%.

User reading content of the ontology can decide whether to see all information or only that with high credibility.

To sum up: presented method solves two problems of building a centralized ontology. Problem with control of ontology building can be solved by passing control to community. Consensus between all users can be reached by requirement of logical and structured discussion. A centralized ontology will have advantages described in [1] and [3]. Moreover, due to the number of users, it seems that the ontology size could grow very fast – even one hundred thousand to a few million concepts per day (including instances of concepts). The system would enable a more efficient exchange of information among people, and searching for information would obviously be much more efficient. The proposed ontology will have not only the basic features of the Semantic Web but additionally its content would be credible. Because of homogeneity, the use of the ontology by people as well the use of it for reasoning and analyses will be simplified. Using the ontology will be so simple and intuitive that everybody could use it. The ontology can gobble up many currently existing services including auction sites, FOAF sites etc.

A few elements remain unclear. What should be done if someone adds a concept to the ontology but defines it incorrectly? Defining a concept means to describe its features and/or relations with other concepts, and adding this concept to the ontology can be treated as information “this object exists”. If definition defines an empty concept, i.e. concept does not exist, then the information about existence of the concept is false and after reaching consensus the concept can be deleted. This is one of solutions in this case. However, if the concept defined by definition really exists, but the definition defines concept different than this pointed by its name, then information about its name is false. Such a name should be deleted

after reaching consensus and new name should be added. Concept can have more than one name in one language. Names in languages other than English can also be given. After reaching consensus any other concept can be deleted, for example because it is useless in the ontology.

Definition i.e. set of relations and properties can be seen also as information saying that the concept can be uniquely defined by this set. There can be more than one definition. Each of them can be treated as regular information and anybody can agree or disagree with them. One of the definitions should be marked as main definition.

It seems that all elements of an ontology i.e. concepts, definitions, names of concepts, properties and relations can be homogeneously treated in the described way as information. The main rule is to determine correctness of information through debate.

Collective intelligence

Presented method of ontology building is an attempt at computer implementation of so-called collective intelligence (of people). The concept will be described below.

Collective intelligence [4] is a form of intelligence that emerges from a collaboration of many individuals, most often as a result of consensus-based decision making. It can resemble swarm intelligence but I think that the latter rather applies to less intelligent animals. Collective intelligence emerges, for example, among people and leads to development of knowledge of all civilization. An example of collective intelligence is decision-making process in the scientific community. It is hard to describe collective intelligence among people because it emerges as a result of verbal communication and it is stored inside particular people's minds. There are attempts at reading of collective intelligence from human activity in the Internet. Two most famous attempts are:

- PageRank – Goggle created commonly known method allowing reading of collective knowledge about the quality of individual web pages. The information read concerns only popularity of pages. Voting and not consensus is used to determine it.
- Web 3.0 – Collective intelligence will emerge in Web 3.0, but I think that it can be possible only due to artificial intelligence not human intelligence.

In the scientific community, consensus is used to obtain collective intelligence, however the rest of the people rarely use it, because they most often do not discuss in exhaustive and constructive way. The proposed method of ontology building can be an attempt at implementation of people's collective intelligence in a form of an ontology. As was shown, the proposed method uses intensively two of main collective intelligence features: cooperation of many people and consensus-based decision making. In my opinion and also in the opinion presented in [5], ontologies built by knowledge engineers, not by society, will be too small to apply them in real world. On the other hand merging many small ontologies is difficult and nobody knows how to do this [1, 3]. Thus, it seems that community-driven ontology building is currently the only solution.

If the presented method is correct it:

- will cause expanding methods of collective intelligence from the scientific community to all humanity and could cause an evolutionary jump in civilization development,
- will cause arise of an explicit computer representation of collective intelligence which exists now among people. Thanks to this, computers will have access to vast knowledge. Maybe it would be enough to apply logic to this data for artificial intelligence arise.
- will have all the advantages of common intelligence, including that which follows from consensus,
- will be a combination of common intelligence idea with an ontology.

myOntology

One or two years ago so-called myOntology project was started [5, 6]. Community-driven ontology building is also a goal of this project. But if I am right, in myOntology, consensus can be difficult to reach, because in this project one can discuss only by plain text and not in a structured way as was described in this text. In myOntology, only classes, properties and relations will be stored so it will not be a store of any information. The method described in this text enables to store information because its credibility can be determined during structured discussion. In the described approach everything is treated as information. Credibility of information is determined from credibility of users and vice versa. Because the ontology created in described way can contain any information, it can be considered as a centralized version of the Semantic Web, while myOntology will be only an ontology. Using myOntology as any other ontology to build the Semantic Web will cause problems with credibility of information stored on web pages. I do not know if it would be possible to add to myOntology such fine-grained instances of concepts as "my instance of Nokia 6021". I also do not know if, in myOntology, users will be associated with real physical people.

It is obvious that there should be only one central ontology. So in my opinion, the proper ontology should be chosen before making it available to the community. Otherwise the community can be discouraged from inserting billions of pieces of information next time.

Applications

The proposed ontology can have many applications and implications. Some of them are:

- Semantic Web
- The creation of a centralized version of the Semantic Web which will be additionally credible and complete. Credibility means that any information stored in the ontology will be true. This feature is not necessarily associated with the Semantic Web. Completeness means that any information stored in the ontology will be complete in the sense that we do not have to search for other sources of information to find out about other views. The ontology will also be non-redundant i.e. there will not be two the same statements.
- The proposed ontology can also be used as an ontology to describe web pages in the process of creation of the Semantic Web, if one wants to store information on his web page and not in a central repository.
- Implications for people
- New kind of information medium.
- Improvement of information exchange among people.
- Development of people's awareness, thanks to improved access to information. Development of people's awareness could solve huge amount of problems.
- The ontology will allow access to knowledge about individual people and thus, because of semantic searching possibilities, it can decrease the level of social isolation.
- Sometimes one well matched piece of information can radically change the way man thinks. The described ontology may also be a good source of this kind of knowledge. Due to credibility, completeness, uniqueness and easy searching, it can be treated as a very good source of knowledge on every question we have.
- Implications for society
- The system will be a platform for matter-of-fact discussion and debates.
- The system will cause people to be more aware of the situations of other people, for example the poor.
- The system will prevent people from lying and corruption because it can easily come to light in a centralized, credible and non-redundant system.

- Discussion and awareness can prevent conflicts, terrorism and wars.
- The system can leverage differences between communities, for example financial differences, thanks to an improvement of people's awareness.
- A return to the idea of direct exchange of knowledge between people. Some time ago people exchanged knowledge by direct verbal communication. Now much information is exchanged by mass media which are susceptible to manipulation.

• Implications for economy

- Opposition to disinformation resulting from the rules of capitalism which can be observed for example in advertising. This can be achieved thanks to semantics, for example by evaluation of each product by community in presented way. The company goal will be the satisfaction of all customers and not only profits as it is currently.
- People will have possibility to simply choose the products best for them thanks to knowledge stored in the ontology, as well as intelligent multicriteria decision support systems using the above-mentioned knowledge. Let us suppose that a customer has filled in his profile information required for these systems, such as: geographical localization (for example address), wealth, description of owned computer(s). In such case, buying a product, for example a GSM modem, can be as simple as accepting the proposition of buying instance of "Nokia 6021" cellular phone from a particular seller on eBay or in some store depending on the cost of operation etc.

  If the buyer is poor then "Nokia 6021" will probably be proposed. Otherwise "Nokia 6230[i]" would be proposed because it has the 10^{th} class of EDGE which means that it is faster. The speed of the Internet connection over EDGE depends on the distance from the base station. Thus if it is known that in user location the signal could be bad, then for each phone it will be mentioned that its data transfer speed will probably be 2 times less than the theoretical. In such a case, the system could automatically propose "Nokia 6230[i] if the user wants to connect to the Internet with speed not slower than 20 kB/s (160 kb/s)".

  A particular phone "Nokia 6021" can be chosen depending on wide range of information. For example in our location the EDGE signal of one of the GSM operators is stronger, so the cellular phone bought should have a SIM lock of this operator or have no SIM lock. In the second case cellular phone is more valuable, for example, by $10 if its unlocking is not free.

  I did not describe exactly the way that intelligent multicriteria decision support systems could work, but it can be imagined that they will think in the same way as experts solve problems. Such systems should think logically and use any important information to optimize the satisfaction of a person. Decision support systems can be more or less intelligent. But it is obvious that even the simplest system will be more powerful than systems now existing. Most intelligent systems can even be much more useful than decisions of experts, because they will have much more information to base on.
- Possibility of geolocalisation of the best shops close to us that have a profile being a close hyponym or hypernym of the profile we are interested in. For example we can connect to the system by a cellular phone with GPS and ask for a cheap drugstore close to us. Thanks to the proposed ontology, we will be able to go to a really cheap drugstore in opinion of many people, and not only cheap in the opinion of one or two people (maybe being owners of this drugstore).

  Currently information about a particular shop is often distributed across many websites to increase the chance of being read by potential customers. But what if one store is closed or its information has changed? Nobody will update dozens of entries. Updating of information about one's shop will be easy in a centralized ontology. One would

have to update only one number if he has changed closing time and everybody will see the updated information. Let us suppose it is Saturday 8 p.m. and we want to find a close and cheap drugstore. We can do it because in a centralized ontology there are up-to-date opening hours.

- Implications for science
- A tool for scientists improving the exchange of knowledge and ideas.
- A database which can be used in all kind of semantic analysis.
- The proposed ontology will be a good tool for exchanging ideas and thoughts. Currently, thoughts published on the Internet are not accessible to all interested people because they are hard to find. The proposed ontology could change this. For example we can add to the ontology the currently discussed idea by adding an instance of "ontology" concept with "scalable" property set to true and "status" property set to "idea". There will be only two instances with these properties: currently discussed and myOntology. So we will know that only these two ideas were proposed until now.
- Implications for computer science
- The possibility of building intelligent multicriteria decision support systems.
- The proposed ontology can be considered as a knowledge database because information stored there would be correct, complete and everybody would have very fast access to any information that they are searching for (thanks to completeness, uniqueness and organization of knowledge).
- The ontology can be used as platform for development of artificial intelligence. Maybe it will be easier to build AI if we have a structured human knowledge base. If people are able to use the ontology as a source of knowledge then maybe computers will also be able to do the same.

Additional information (unsorted)

- I think that it is also worth mentioning how information about instances of concepts is currently published on the Internet. Now there are many possibilities to publish it but published information is difficult to find. For example one can publish information on Usenet, a second user will publish it on a forum and a third – on another forum. If other users think that one of these pieces of information is false they can reply to the original message but nobody will correct the original information! So somebody who wants to find an answer to his question must firstly search all different media. Next he must filter inadequate search results. Next he must read through whole threads, also containing messages with false information. What is more he will often find threads consisting of only one message which will be the same question as he has – without any answer! In my opinion it would be much easier if all information is in one place and false information will be corrected to true. It will cause that people will be more willing to publish on the Internet the information that they have. For example somebody wants to add information about a particular song. What can he do now? He must find an appropriate forum? Write it on his home website? In the described editor it will be easy. One would only have to find the proper object describing the song and then add any information, if this information is currently absent. What must a person who wants to read this information do? He only must find the proper object and read this information.
- If not all users agree with some information it will be possible to calculate statistics. For example it will be possible to calculate what percentage of Nokia 6021 phones is defective from the "was broken" property of all Nokia 6021 instances. The credibility of each user can be used in the calculation.

- Thanks to notification by SMS about new argument or information added to a hyponym of a subscribed concept anyone can read this information immediately. If the newly added information is a question (for example it can be information with an unspecified probability of correctness) then within a few minutes an answer to this question should be added to the ontology i.e. the probability of correctness of this information should be determined by some people. So thanks to the system, the user will have possibility to obtain an answer from experts within a few minutes. Within one hour the answer should be exhaustive. It seems that distributed Semantic Web will not have this possibility because computationally it would be too expensive.
- Because of the discussion structure it will be easier to discuss difficult statements.
- If information with credibility less than 100% is used as an argument for another information then credibility of the latter is also less than 100%.
- Information is stored in a fine-grained form to allow easy determination of its correctness also by other users. If information is stored as article (as it is in Wikipedia) it would be more difficult to reach consensus in determining its correctness.
- Thanks to fine-graining of stored knowledge and its partial formalization it seems that it would be possible to build a natural language interface to the ontology. For example after sending a question by SMS we could obtain an answer in the same way.
- Creation of the proposed ontology will cause separation of information layer from information presentation layer. Currently almost all commonly used information is stored as longer texts, for example in articles, books and web pages. In the proposed ontology, information will be stored as indivisible pieces of information. This will make knowledge management and access easier. Information can be presented in a raw way or for example by construction of a sentence in a natural language as a response for a question. Apart from that, in the future it may be possible to generate longer texts from information in the ontology.
- Users can connect concepts by any relations that are true. Thus in the ontology, superfluous relations that follow from the transitivity of other relations can exist. For example, at the beginning, somebody added information stating that “Pluto” concept is a hyponym of “Astronomical object” concept. Then another user, after adding “Planet” concept to the ontology, added information that “Planet” is a hyponym of “Astronomical object” and a hypernym of “Pluto” concept. Thus the first relation became superfluous, but there is no reason to delete it, since it is still true. But because it is superfluous, it will decrease the clarity of the semantic network so it can be omitted during displaying concepts. But after it turned out that Pluto is not a planet the hypernymy connecting “Pluto” with “Planet” concept was deleted. Now the original relation between “Astronomical object” and “Pluto” should be displayed, because it is not superfluous.

  Set of “Animal”, “Mammal” and “Cat” concepts and hypernymies joining them is another analogous example. This example can be presented to younger children without displaying “Mammal” concept, because they do not know it. So in this case only the superfluous relation connecting “Animal” with “Cat” will be displayed, because “Mammal” concept does not exist for information recipients. Similarly Latin and popular names of drugs can be presented to doctors, pharmacists and chemists, but for other people only popular names would be presented. These examples illustrate the separation of information layer from information presentation layer very well. Every concept can have a property describing on what level of education people become familiarized with it. For example for “Mammal” concept the property can be set to 4^{th} year of primary school. The decision as to whether a user is familiar with a given concept can be taken on the basis of his or her current education.

- Information is stored in a formalized form to allow its automatic processing including semantic search. The information that "Bill Clinton was president of (some) state in 1992-2000" should be connected with all concepts appearing in this information. Thanks to this, we are able to find this information while reading the content of one of concepts (or its hyponyms). For every concept appearing in information its role should be given. For example the "years 1992-2000" concept is used as a chronological adverbial and it limits range of the relation to a given time. Such a form enables automatic reasoning and semantic searching, as opposed to storing the adverbial in text form i.e. "was in 1992-2000".
- As everybody knows, Wikipedia is not fully reliable. In Citizendium project, it is proposed to obtain reliability thanks to experts' supervision. Such a solution is not required in the proposed ontology, because adding false information to the ontology cannot worsen its quality significantly, in contrast to an Internet encyclopaedia. It can even be said that every performed action, also adding false information, leads to an increase in information amount and quality. False information will be most probably quickly corrected (because correcting is very easy). Corrected information and more correct evaluation of user credibility (i.e. the user that inserted false information will be less credible) will be the only durable results. Thanks to more correct evaluation of user's credibility we will be able to evaluate more correctly credibility of information that he contributed to.
- Thanks to the reliability of the ontology, many scientists will use it more willingly than Wikipedia.
- It is also possible to add to the ontology information that is not in fine-grained form. For example photos, audios and videos can be added. They can be treated for example as information that they represent given concept.
- It can be said that Google returns results on the basis of word similarity and information popularity. The presented ontology will return information about hyponyms of exactly these concepts that we are looking for. What is more, the result will be sorted from the best hyponyms according to features important to us and not from the most popular ones.
- Thanks to the ontology it would be possible to notify immediately anyone interested in newly added information. For example if one wants to buy the 8^{th} edition of book "Core Java 2", he can set a trigger on the "published" field of this book concept. If the field changes its value to "true", then the user will be notified about this. It is also possible to subscribe to new important information added to one of the hyponyms of some concept.

Origin of the idea

Below I will describe the reasons for the presented idea. If it turns out that the idea is not correct then maybe they will help to transform it.

The need of access to knowledge is currently more and more evident, so a very big database of knowledge will be useful for people. The database should store not only ontology, but also other information.

I was wondering how consensus about content of the database can be reached. I cannot imagine how it can be done without artificial intelligence in a distributed Semantic Web consisting of many ontologies. I also cannot imagine that editing of one article in Wikipedia by a million people would be possible. Everybody would want to change something, for example to add something. Such a solution is hardly scalable. Scalability can be achieved if we divide information into very small indivisible parts. Thanks to this, there will be no need to change information because it will be so small that if it turns out to be false it can be simply

deleted (or marked as false). Evaluation of information correctness can be done by any number of people, also by millions or even billions of people! Additionally we can force people to reach consensus about the correctness of information by requirement of discussion if they do not agree with each other. Large amounts of information have specified and unquestioned correctness, we must only discover the truth. For example New York undoubtedly lies in the USA. I think that most often people cannot achieve consensus because they do not discuss. This can be observed at every turn.

Information in such a system would be organized and fine-grained so that access to it would be very fast! What is more, it will be correct and complete so we can treat this information as knowledge and not as information of questionable value. It can be treated as an extension of the knowledge stored in our minds.

If people are able to use it as their own knowledge (or good notes in a notebook) in process of reasoning, then maybe also computers will be able to do this. I think that current approaches to artificial intelligence do not work because they do not have access to such knowledge. Maybe if they will have access to this knowledge they could solve problems that humans have. As we saw, even decision support systems without any logic can be very useful. But because a lot of information would be stored in quite formal form, maybe it would be possible to apply reasoning to them. Maybe artificial intelligence only needs knowledge to arise?

Evaluation in the real world

Around January 15, 2008 in Italy, academics and students of one of universities protested against a visit of pope Benedict XVI, because they claimed that in 1990 the pope called the Galileo trial reasonable and just. Did the pope really say this? I do not know. On the next day, on a radio station, I heard that he did not. Who is right? I think that debate is more fruitful than emotional protest, isn't it?

References

1. Berners-Lee T., Hendler J., Lassila O.: *The Semantic Web*, *Scientific American*, May 2001, p. 34–43. Available at: http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21.

2. Ding L., Kolari P., Ding Z., Avancha S., Finin T., Joshi A.: *Using Ontologies in the Semantic Web: A Survey*. Available at: http://ebiquity.umbc.edu/get/a/publication/190.pdf.

3. Shadbolt N., Hall W., Berners-Lee T.: *The Semantic Web Revisited*, *IEEE Intelligent Systems*, May/June 2006, p. 96-101. Available at: http://eprints.ecs.soton.ac.uk/12614/1/Semantic_Web_Revisted.pdf.

4. Wikipedia: Collective intelligence: http://en.wikipedia.org/wiki/Collective_intelligence.

5. Siorpaes K., Hepp M.: *myOntology: The Marriage of Ontology Engineering and Collective Intelligence*, *Proceedings of the Workshop Bridging the Gap between Semantic Web and Web 2.0 at the ESWC 2007*. LNCS Springer: Innsbruck, Austria, June 7, 2007. Available at: http://members.deri.at/~katharinas/files/publications/SiorpaesHepp-myOntology-eswc07-camready-final.pdf.

6. myOntology - www.myontology.org.