

Semantic marketing – product knowledge base

Dariusz Daćko

October 26, 2008

updated on August 12, 2010

www.consensualknowledge.net

1. Product purchase

In 2001 I wanted to buy an inkjet printer for everyday personal use. Having read the information provided in a ranking article in one of computer magazines, I selected one of the models. My choice resulted from the fact that the model had been produced by a fairly well-known company and, according to the article, the printer printed quickly and cheaply. Moreover the printout itself was of good quality. The inkjet heads were an integral part of the printer and not of the cartridge with the ink. Thanks to it, the purchased ink was cheaper and the print itself was economical. I checked the information in the producer advertising leaflet. The described printer model turned out to be the best, so I bought it. Having read the printer manual I found out that the printer indeed printed 15 pages per minute, provided that an additional head was bought. Unfortunately, the head cost almost as much as the printer itself. Of course, the purchase of the head was not an option since I wanted to print economically. In this way, I and also the editors of the magazine were misled by producer, and what is more, I became the owner of a printer that could print only two black-and-white pages per minute. Printing 120-page materials took one hour. Because I preferred reading the materials on paper to reading them on the computer screen, I often had to wait until they were printed. One day I went on a two-month holidays. When I returned, it turned out that the printer nozzles had dried up. Unfortunately, there was no way to unclog them. On contacting customer service, I found out that in order for the nozzles to not dry up, one should print some pages from time to time. I was not aware of this, as it was not described in the instruction manual. Despite this, the nozzles should not clog irreversibly during a two-month period. Unfortunately, I found out that my model of printer was especially vulnerable to nozzle clogging.

I am convinced that the producer deliberately misled us providing information about quick printing and leaving out the information that it could be achieved only when an expensive head was bought. It is obvious that he acted in this way to avoid discouraging the customer from making the purchase. Thus the manufacturer took care of his own interests and not the interests of his customers.

2. Needed information and available sources

The process of choosing the appropriate product is difficult. The product should be the best for us, and at the same time, it should not be harmful for the environment. It should be useful, healthy (for instance foodstuff should not contain harmful preservatives) and cheap. The product as well as its packaging should be biodegradable, and the process of its production should be ecological. We may also want to support, for instance, local manufacturers. We have to make a multi-criteria decision based on a large amount of uncertain and often contradictory or redundant information. More over, we may have no information about some features of the products at all. Most often, the advertising materials contain only the information favourable to the manufacturer or the seller. For example, on a leaflet about a vacuum cleaner equipped with power adjustment, we will find a piece of information that it does have such a function. However, if the vacuum cleaner has no power

regulator, most often it will not be stated overtly. Since such a piece of information is copied to many websites, we have to deal with great redundancy. Therefore, most often, we can find product descriptions that do not contain the information, which would suggest that the product is not the best. To sum up, we must evaluate the product taking into consideration many criteria, and at the same time we need to face the situation where access to information about it is difficult. In most cases, this causes the decision made by us to not be the best.

Later in the article, a project of a computer system that enables consumer to obtain objective and full information about products has been described. Thanks to this, the customer will be able to make more conscious decisions and buy the products that are best for him. On the other hand, changes in the demand should enhance the quality of the manufactured goods, contributing to positive changes in the whole market.

3. Web 2.0 is a better source of information

Most of the products are concepts less diverse than other terms of natural language. For example, trees of a specific variety may have coronas shaped differently and the number of their branches may vary too. But products most often have the same characteristics just after they are manufactured. Certain model of vacuum cleaners may have some feature (for example power regulation) or not, but all vacuum cleaners of this model are the same. Of course used vacuum cleaners may differ from each other since something could be broken in one of copies. However, despite this, the vacuum cleaners of a given model are not too different. Thanks to it, each product has strictly specified characteristics. Information about the products is not ambiguous. It is just enough to get to it and write it down, and it will be available to all consumers. As has been shown before, the manufacturers and the sellers are not a good source of information about the products. The consumers are a much better source.

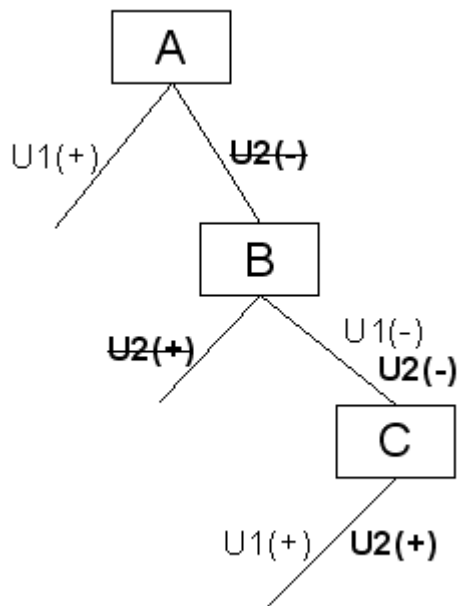
Not for long we can observe in Internet the Web 2.0 trend consisting in the active creation of Internet content by Internet users. It could be possible to create a Web 2.0 service enabling each Internet user to add information about products. Thanks to it, a very large and objective database about all the products could be created.

4. Determining of the information truthfulness

As has been shown before, it is possible to verify most information about the products and to determine their correctness. This is important, since contrary to the manufacturers, the Internet users may be mistaken more frequently. Some of them will devote less time to verifying the information being added, and therefore, it may be false. Also, it often happens that users consciously add false information (vandalism). The fact that people add information of various quality to the Internet can best be seen on the currently existing Internet.

If one of the users (let us mark him as U1) has added a certain piece of information to the system (let us mark it as information A, see picture 1), then everyone, to whom it seems false (e.g. user U2) may vote against the correctness of this information. However, additionally, he or she must provide an argument against the correctness of the information (e.g. argument B). The primary author of the information (U1) and all the users who have marked information A as correct must refer to argument B which has just been provided. Each argument is an ordinary piece of information. Additionally, it is related to the original piece of information (A) as an argument against it. User U1 opting for the correctness of information A should prove the falseness of the information of argument (B) or that a correct argument B does not show the correctness of information A. If he chooses the first possibility, he will have to give an argument against information B (e.g. argument C). User U2, having become familiar with argument C, may agree with user U1 and mark information B as incorrect. Thanks to it, all users interested in information A and B reach a consensus. In their opinion,

information A finally turned out to be correct, and information B incorrect. The consensus has been reached owing to a discussion based on arguments.



Picture 1. Argumentation tree. The information added to the system is presented as rectangles. The branch, which goes downward from the rectangle means that one of the users opts for or against a piece of information. The user identifier and the fact that he votes for or against it are shown in label placed next to the branch. The first vote for the information correctness may have an argument, but it does not have to. The arguments are treated as ordinary information. If a branch has several labels, it means that several users cast their votes, justifying it using the same argument. The user who came to the conclusion that the piece of information added by him is false, may change the voting value, deleting his vote “for” its correctness (the crossed out labels in the picture) and possibly adding a vote “against”.

The more pieces of correct information and the less pieces of incorrect information that are added by the user, the greater credibility he will have. If some of the users opt for the correctness of the information and some of them against, it means that the correctness of this information is not established. In such a case, the probability of its correctness will be evaluated taking into account credibility of users voting for and against it. Thanks to it, each vote of the users will be adequately considered in the system.

If vandalism i.e. intentional misleading is proved against some user, it will be possible to remove him from the system for a long period of time (for example two years). In order to avoid the repeated registration using a different login, the users should be verified, for instance, on the basis of the place of residence, as it is done in the case of Allegro - a Polish online auction website.

Let us come back for a moment to the example discussed at the beginning of the text. Let us assume that the manufacturer added a piece of information that his printer, costing a given amount of money, prints 20 pages per minute. After purchasing the printer and reading the instruction manual, one of the customers evaluates this piece of information as false, quoting a paragraph from the instruction manual which states that the standard equipment of the printer (included in the price specified by the manufacturer) makes it possible to print only 2 pages per minute. The manufacturer would be forced to admit that he made a mistake and correct the false piece of information. If he does not do it, Internet users will evaluate this piece of information by themselves, casting many votes that confirm its incorrectness. This way, the manufacturer will lose his credibility as he put in false information. Additionally, the consumers would not be misled anymore. Obviously, the manufacturer, being aware of the consequences, probably would never decide to add false information to the system.

Most information about the products may be presented naturally as the values of their features. For instance, the weight of the product can be presented as a number. Owing to the fact that the information is fine-grained, its correctness can be easier determined. The determination of the truthfulness of a comment consisting of a few sentences would be much more difficult, as the comment could contain partly correct and partly incorrect information.

The evaluation could not be expressed as completely true or completely false. Therefore, it seems that fine-graining is a key feature to enable people to reach a consensus. It is also a feature that distinguishes the proposed system from existing systems which only enable users to add comments about chosen products.

5. Characteristics of the database

The proposed system can contain lots of information. In order to facilitate its search, it should be ordered in a hierarchical knowledge organization system called a semantic net. In such a system we define objects, their properties and relations between the objects, including inheritance relations. For example, following three terms may be defined: “Printer”, “Inkjet printer” and “LG model A”. The terms are connected by inheritance relations. For the term “Printer” we define, among other things, a property “Number of black-and-white pages per minute”. Now this property does not have to be defined for the term “Inkjet printer”. But we define another property, for example: “Use of ink per 100 pages” or “Number of ink colours”. Thanks to it, while adding the concept “LG model A” to the system, probably we do not have to define any new properties – it is just enough to fill in the missing data that are not inherited from parent terms. Owing to the hierarchical organization of terms, it would be possible to find a given piece of information quickly. Moreover, the information being added would not duplicate existing one. If it is difficult to present some information as a value of the property, it may be added as a sentence in natural language. However, it must be related to objects which it concerns, in order to facilitate the search.

The organisation of information has one more additional advantage. If we want to find out everything about the product in which we are interested, it is just enough to localize it in the hierarchy. All information stored in the system, which concerns this product, will be connected to the found object by implicit relations. It is no longer necessary to browse through excessive amounts of redundant web pages on the Internet. If we do not find the information on the web page of a given term, it means that this information is not stored in the system – it is not necessary to look in other places in the system.

The users will have specific credibility and would care about it. Even if only one user added some piece of information to the system, it can be considered credible provided that the user is credible. The given piece of information does not have to be checked by anyone, in order to be credible. This means that the growth of information contained in the system may be very fast. Because of this, it is possible to add also regional products and not only the ones available on the global market.

6. Applications of the knowledge base

So far only one application of the product knowledge base has been mentioned: the base may be used for finding the specific information that one is searching for. After some time, we may expect that the knowledge about the most popular products would become complete, that is, almost all properties will have specified values. What is more, the knowledge would be certain, that is, there will be no incorrect information in the database. Knowledge of so high quality may be used for the automatic determination of the best product meeting the user expectations. It would no longer be necessary to compare the values of the features of all products by one’s own. It will be sufficient that user specifies his preferences, and the system will find the best product for the user by itself. In order to use an automatic search of the products, user should complete his preferences in the profile. For example, he can provide the following general information:

- the user is not wealthy (it is necessary to provide the amount of money specifying what is meant by this),
- the user cares about environmental protection,

- the place of residence.

Additionally, during the search for a printer, he can provide the following information:

- the user is looking for a printer for everyday household use,
- the user wants to print a lot of black-and-white pages, so he is anxious about printing such pages quickly.

The appropriate plug-in for the system, containing domain knowledge (in this case knowledge related to printers) will take into account the preferences of the user and will select the best products on the basis of data contained in the knowledge base. The user is not wealthy, so price is a criterion with significant weight. The consequence of the fact that user cares about the environment is that the product itself and its packaging should be biodegradable, and the production process should be ecological. Of course the printer should generally be good (quality, speed etc.) – all these rules would be included in the plug-in. Additionally, the plug-in would pay particular attention to the speed of printing. The products which will show up as results of the search may be presented for comparison on the resulting webpage. It would also be possible to explain there the meanings of the respective product properties and the consequences resulting from the differences between them. On the resulting page it will also be possible to buy the selected product. The plug-ins can be as complex as it is need. The least complex ones will take into consideration only basic rules. The most complex ones may approximate the knowledge of an expert in a given field. This will be a chance for the programmers to display their skills.

The database could also be used to display information about the product being sold on the website of a shop, auction service etc. Salesmen or the vendors in the auction will not have to write the product description on theirs own or look for a description in the web. It will be enough that they provide an identifier of the product in the knowledge base, and the data will be displayed automatically on their web pages.

The introduction of the additional concept of a “used product” may be helpful for those who buy at auctions. If the product is used, it may be significantly different from the unused one. The features of the used products may be common for all the products, for example: the degree of wear, how long has the product been used, is the vendor the first owner of the product? etc. The features may also be specific for a selected type of the products. For instance, in the case of cars, this will be: mileage, information as to whether a car has been in an accident; and in the case of a TV set: usage of the screen’s luminophore. If the vendor does not fill in a certain field inherited from “used product” concept in given domain, it will be a signal for the buyer that the vendor could be trying to conceal this piece of information. In this way, we will also avoid unclear expressions camouflaging the bad features of a sold product copy.

The system may also be used in mobile devices. Let us assume that we have a miniature barcode reader equipped with a Bluetooth interface and a mobile phone with Internet connection. The barcode reader could read the code of the product which we are interested in and send it to the mobile phone. Optionally, this could be performed independently by a mobile phone equipped with a camera. Then, the mobile phone would check whether the product in which we are interested meets the criteria stated by us before in the profile. If the product does not meet the criteria, it could be signalled by a red indicator on the Bluetooth module or in the mobile phone. At the same time, a better product could be proposed. Meeting of the criteria could be signalled by a green indicator. In such a case, we could put the product in the shopping cart. Of course we would use the database only during the first purchase of a product unknown to us.

Sellers will be also able to add information on which products are available in which shops. If a shop does not have a product stock management system, it may add only the categories of the products available at the shop to the database. Also, the information on the

opening hours of the respective shops may be added. Thanks to this data, the mobile system may provide us with information as to which shop located near us we should go to in order to buy the product selected by us cheaply, or where there is a cheap and nearby pharmacy, open on Saturdays at 8 p.m. At present, the information about the respective shops is often scattered over many websites in order to enhance the chance of finding it by a potential customer. But what will happen if one of the shops is closed down or if the opening hours are changed. Nobody will update tens of entries. The updating of the information about any shop will be simple in a centralised product database. If the time that a shop closes changes, only one number will have to be updated and everyone will see the updated piece of information.

7. Consequences

As has been mentioned before, the building of this product knowledge base may have both local and global results. Locally, it may bring profits to the consumers, who, owing to it, will select better products more appropriate to their needs. Demand controls supply. Therefore, if the database is used by many people, it may cause global economic effects on the whole market. The manufacturers will adapt to the needs of the consumers, manufacturing better, healthier and more ecological products. The competition in satisfying these needs will have a positive impact on the further direction of the economy, the health of people and environmental protection. People's welfare should increase owing to a more effective satisfaction of our needs at a lower cost.

Besides the described benefits, the database may also bring an additional profit to its creator. After choosing the appropriate product, the customer has the possibility to make an order. Due to the global scope of the system, profits from the sale of the ordered products may be quite considerable.

8. All-thematic knowledge base

The idea described above was obtained after narrowing down the idea of construction of an all-thematic knowledge base to concepts describing products. The first of these databases has been described in [1]. Apart from the difference in scope, the ideas are the same. Therefore, in order to become more familiar with the knowledge management method, one can read a more general text. The idea was described again from the products point of view, because:

- 1) As has been written before, the product features have exactly specified values. Therefore, it will be easier to convince somebody that consensus about the database content is possible.
- 2) Information about products can be structured in a natural way. Product features are often presented as a feature-value list. As product features are specified clearly, it is easy to determine the hierarchical relations between different product classes.
- 3) Building the knowledge base is economically justified.
- 4) The project self-finances itself as the costs incurred during its implementation will be returned during its use.

The experience obtained during the process of building the product knowledge base may be used to create a general knowledge base, which would have a lot of additional advantages and applications in comparison with the first one.

9. Example

Example concepts concerning vacuum cleaner will be presented below. The structure of concepts has been simplified since a presentation of the full structure is not necessary to explain the idea. The following four concepts have been added to the product database:

- a) a "Product" concept with the following properties (features) defined:

- series of product,
 - colour - of type “Colour”,
 - manufacturer - of type “Company”,
- b) an “Electrical product” concept which is a hyponym (that is, an inheriting concept) of the concept “Product” with the following additional properties:
- power - of numerical type and Watt unit,
- c) a “Vacuum cleaner” concept which is a hyponym of the concept “Product” with the following additional properties:
- “equipped with a HEPA filter” - of logical type,
 - “equipped with power regulation” - of logical type,
- d) a “V-CP753NBB” concept which is a hyponym of the concepts: “Electrical product” and “Vacuum cleaner”, with the following values of the previously defined properties:
- series: Turbo Zet (a concept)
 - manufacturer: LG (a concept)
 - equipped with a HEPA filter: no
 - equipped with power regulation: no
 - power: 1500 W
 - colour: blue (a concept)

After selecting the “Turbo Zet” concept, the values common for the whole series, that is information about the manufacturer, HEPA filter and power regulation could be copied. If some value is a concept, it means that it is selectable from a list of the defined concepts of a given type.

References

1. Dacko D.: *A method for building a scalable centralized ontology*. Available at: www.consensuallknowledge.net/files/centralized_ontology.pdf.

Changes in the text

- August 12, 2010 – two small corrections