

MBD Search Engine - the searching technology for the legal informational trade in the Internet

By Valeriy Zakharchenko

The only restriction to the electronic editions access in the Internet must be the balance of your account

What do I lack in the Internet as a user?

I lack freedom. The freedom of access to the needed and valuable information. If I search in the Internet where to buy a vacuum cleaner, the searching system dumps me thousands of advertisements of different value. If I search for data, which usually may be found in the library collection, then it finds students reports ore books annotations. This search evokes nothing except of irritation. Something can be found in the paid servers, but the searching possibilities, informational base, access system and payment order are not well implemented.

I want the informational space in the Internet, which possesses everything: any books, artistic and scientific, any newspapers and magazines, popular data bases. You may call it anything you like: library, searching system, infomarket. I want to freely enter the searching system of such library, to search for the information, read found fragments, output needed pages, look illustrations, read needed articles, skim pages, load editions in my own archive, to sign for the needed periodical editions. I want much more else that can be provided by the additional services. I want the free information market. I am ready to pay as a customer and I want to get the real access to the needed product for my money without burdensome procedures. I want the balance of my account to be the only restriction. All the rest is the salesman's problems, not mine.

All the mentioned that I want and hundred millions of users of the Internet like me is the goal setting. The task of the informational technology development and the program supplement for the electronic editions market and information on the whole. Of course, there are already the projects of the worldwide library, and there are existing electronic libraries. They exist, but can I now receive the access in the Internet to the needed information for my money? No, I can't. For me as a user, the task is not solved. Judging by the legal claims to the searching systems, it's also not solved for the authors and the editors. It means that these systems lack something substantial to solve this task in complex. What do they lack? I'll try to formulate the demands to the informational technology that can help to solve the problem of the electronic editions spread.

Can you imagine the Internet without the searching systems? Personally I can't. Exactly the same with the global or just big electronic library which is impossible without a searching system. It is possible to search for a famous book there, but it's impossible to search for the information in it. It's even impossible to find a detective to read, like we, standing near a bookshelf, skim and choose the book. It means that the searching system for the electronic libraries is necessary. The searching system not as a single searching mechanism for the electronic library, but as the united program informational area, completely managing all the library's functions.

The searching system must take into account the search specification in the books and magazines. The main thing is that it must find not the whole books, as in the majority of the known searching machines, but the single needed pages in the books and skim them in any sequence. It makes the system work ten and hundred times faster. A book or a magazine may be of a great volume and it's hard to work with them. That's why you often need only the distinct pages with the information and there is no need at all in the whole book, for example in the encyclopedia.

The pages comprise illustrations. That's why the system needs to output not only the separate text pages but also the illustrations connected with them. That's why the

system should content in its index or inner base not only the texts of the books but also all the graphical images. There should be context image search and its output to look with the attached texts.

The searching system should be compatible with the existing edition format, used in the editorial and archival process. Now, to get the electronic version of the book, magazine, newspaper, available for the search, they are, as a rule, structured and divided into dozens of small pages- files of HTML format. That's why a newspaper or magazine electronic version is a bit of another product. Such technology doesn't fit the mass edition insert in the searching system. The system should index automatically the files of .pdf, .doc format of any size together with the illustrations.

Important information is often comprised in the data bases, which can have a size of tens and hundreds of megabytes. For example the phone bases, legal data on the manufactures and so on. Taking into account the importance of this information, the searching system should index and include these data in the general informational base.

The system should provide with a very fast search and information output, distinguished by the channel capacity. It should search so by the whole document texts as by the formal distinctions: date, name, authors, size. It should do the search by the complex messages with the logic operators. Should allow to set the limits upon the distance between the messages, found in the text. To range found documents due to the several criteria.

The most needed information for the user, with which he must work quickly and effectively, using the specialized programs for the adaptation, may be placed not in the network, but in his personal base. The same as there are social library, and there are the personal one with the bookshelf and the books in it. That's why the searching library system should have its local version, installed in the user's computer.

And finally, the most difficult problem, the problem of authors and editors. Evidently, that the authors and the editors of the books, newspapers, magazines must receive money for their labor, unless the writers stop writing and the editors – editing. How to compete the freedom of the users' access to the informational resources with the guaranteed labor compensation of the authors and editors?

Firstly, the fulfillment of the searching system is very heterogeneous. One documents are open for the free access, another are followed by the editors. That's why the authors and the editors must have the complete control under their editions, inserted in the Internet. For example in our system "MBD Search Engine" every document, either a book, an article or a photograph from the many millions, kept in the base, may have its own individual settings. The author may define on his own the size of the output textual fragment, to block the output of the full-sized images, to forbid the skimming of the full pages of the document, to block the downloading of the initial document, to establish the cost of the reading page of the document, the cost of the whole production or its part. The complete control under his production, output in the Internet!

Another, even more difficult problem, is the protection of the productions, legally received from the network, from unapproved copywriting. If the problem isn't solved, the whole technological chain loses its sense. In our system the problem is solved the following way. All the information in the searching system, including texts, graphics, files with the initial documents, is encoded and it may be decoded with the key, consisting of two parts. One part is the specific computer code, for example the processor serial number, the other one is the difference between the basic key and the first part. You can get the basic key, which is able to decode any information from the searching system, by the adding of the two parts the specific part for every computer and the difference key. In fact it looks much more difficult mathematically, but the idea is exactly such. The information is the same, but the key is different for every computer. To copy any data with such encryption is useless.

Electronic library subscriber sets the decoding program on his computer. It may be the program of the local version of the searching system MBD Search Engine. While registering the subscriber enter his computer code. On the base of this code the difference key is calculated and is said to the user to enter the program. Now any information from the electronic library decodes invisibly for the user and is laid in the browser, personal searching system or main program for the initial document file.

Having registered in the system and opened the personal current account, the user gets the free access to all the library materials: books, magazines, daily newspapers, data bases. Every user's claim, every read information page, an article or a book are counted according to the prices set by the author. The price for the user lights on when the information outputs. The whole sum is withdrawn from the subscriber's account and transferred to the rightholder's account. The searching systems and its service programs take on their own all the calculating problems between the salesmen and the customers, the authors and the subscribers. The system is flexible enough. For example as the subscriber's resource in the Internet may serve the local network of the University library. For them another price may be set by the authors. Under this system not only the electronic editions may be spread but also the encoding files with any information, for example with the films or music.

Our company MBD Soft offers such technology. Someone offers another. In any case, it's necessary to understand that the insert of the technologies of the commercial spread of the electronic editions is unavoidable. In the close time these technologies must change basically the market of books, periodical editions, audio and video products, library business.

December 24, 2007
Moscow