

It's your power  
over information



## Corporate Search Uncovered

---

© SearchInform Technologies Inc.



**Information technologies have long settled in the corporate sector. It's a rare thing for the company not to have a well-organized local network and various specialized software providing a proper control of information flow, document storing and information structuring.**

**This document will deal with the potential problems enterprises may face when working with large amounts of data and discuss a possible solution.**

### **Information Diversity**

Any company's information can be roughly divided into three types depending on its "virtual" or physical location and its use in the work process. Starting with files from the user disk (plus electronic mail and logs of various instant messaging programs, like ICQ or MSN Messenger) and on to the corporate information, documents of different formats and electronic mail (MS Exchange, for instance), or a file information archive on the company server. And finally — the data in various information systems: DMS, PDM, CRM, etc. This may include everything from system objects found in file archive or database like MS SQL to "external" electronic messages and documents used in the work of the system.

### **Search?**

Considering such a vast variety of information, demand for perfect search and information processing tools keeps growing. Some problems such tools should to deal with are:

- physical data volume
- lack of proper organization of data, i.e. data scattered all over the local network
- vast variety of file formats containing the needed data

Besides search management (in file archive, corporate electronic mail or document management system), there is quite a number of other requirements that corporate software has to comply with:

- **Working with local networks**

Ability to work with the local network implies client-server software architecture, flexible network policy settings, support of different operating systems, etc. One of the latest trends is having a web-interface for the client part of the corporate software — it gets rid of the problems of additional workstations when extending the information structure. But since the number of workstations is unlimited in that case, this solution may be rather costly to the company's budget. The choice between web-interface and an independent client program depends solely on the needs and problems to be solved by the software being purchased.

- **Compliance with information security policies and user access management**

The next critical factor of search software's work is compliance with information security and access management policies set in the company. Any information system should present a structure with clearly defined channels of information exchange both between users and with the "outside world." Thus any corporate software must meet strict information security requirements: user access differentiation, multi-level access to different sorts of information, authorization system and a flexible structure of security policies adjustment depending on the client's query.

- **Working with an existing system without violating previously set business processes**



Another factor is the feature of corporate software that lets you work with company's previously installed software products of various types. As it has already been mentioned, the information in any organization can be stored in files both on disc or in DBMS and in various information systems (whether it's PDM, CRM or some accounting program). That is exactly why the third feature of any information system is the ability to function not instead of an already existing in the company software, but rather simultaneously with it. It's even more crucial for the corporate search engine because organization of search in all company's information resources is the main goal of any search software application.

### **Search Functions**

Besides the listed requirements, which put various search systems on the same level with corporate software, functional part of such systems responsible for quick and efficient search should also meet certain criteria.

First of all the old generation of "straight search" (simple blind search) and search strictly by document attributes is replaced by the full-text search with prior indexing. It's more than convenient as it's faster even when the search process is a dozen times more complicated.

Second of all it's the support of different file formats (both widely used and specialized ones) as well as flawless work with various types of DBMS, information systems etc. This list shouldn't neglect irreplaceable means of electronic mail (TheBat! or MS Exchange, for instance) and instant messaging programs like ICQ or MSN Messenger. Another must-have attribute of a quality program is a set of search features: a variety of search options (by phrase or by separate words), search with due consideration to stemming and synonyms and so on and so forth. And, of course, specifically for the corporate sector with its gigantic volumes of information high performance speed (both in data indexing and in the search itself) is essential.

### **The Root Of All Search**

Modern search technologies are based on two root processes: indexing of available information and query processing followed by display of results. What concerns the former, any program creates its own area of search. That is, it processes documents and creates an index of those documents (an organized structure that contains information on the processed data). Later on this created index is used by the program for quickly getting a list of documents related to the query.

### **SearchInform?**

SearchInform is a corporate solution used for data search both on the user computer and in the local network. SearchInform, a program that has not so long ago entered the corporate market, has already managed to acquire a high reputation among the end-users, although its merits show more in the corporate use. The integration of this solution into various types of information networks is evidence to that. SearchInform search modules already work in products of such companies as Hummingbird DM (document-management of Western grand developer Hummingbird), PDM Lotsia Plus from the Russian developer of project documentation management systems Lotsia Soft, and in HelpDesk system of IPI company. All this only proves previously voiced thesis on the need for full-text search in various information processing and management applications.

### **SearchInform Complies With All Requirements**

The corporate version of SearchInform offers client-server architecture, indexing files from all accessible (and if there's administrator's permission - in all) folders on all computers in the

It's your power  
over information



local network as well as indexing and subsequent search on all the connected network disks, and user access management system based on NTFS authentication. The user can only search in the network resources that he has permission to access.

SearchInform works **not only with the information on discs (electronic mail — SearchInform supports TheBat!, Outlook, Eudora), but also with other sources of data**, including individually designed ones. The procedure of connecting new data sources is just as simple as, for instance, working with the electronic letters or a mail client: you only need to select the data source (in this case it would be MS Access database and show the program which fields should be indexed or simply leave that to the program and it'll index all the fields automatically).

SearchInform supports practically any information structure as data source for indexing, which only proves that it's more than enough for any enterprise to organize search in all of its information under the management of one program.

### **Search Capacities & Functions**

First of all it's the number of supported file formats: SearchInform has over 60 of them, including most popular formats like TXT, DOC, RTF, PDF, HTML etc., as well as CHM, MHT, Open Office, etc. Besides that SearchInform can conduct search in multimedia files (audio and video), various specialized "programmer's" formats, dozen types of archives and logs of instant messaging programs (MSN Messenger, ICQ, Trillian).

SearchInform standard phrase search includes a huge number of options: search with due consideration to stemming and synonyms, fuzzy search (with mistakes), phrase search and search by separate words that the phrase contains, search by attributes, etc.

### **Something No One Else Has**

Besides the standard phrase search, SearchInform also has another unique type of function – search for documents similar in their content to the query. This type of search is much more so to say 'intellectual' than simple phrase search. In actual practice it helps save the time you'd normally waste on continuously having to pick new keywords, looking over and over and comparing all the documents already existing in the company's database to see whether there are duplications, etc. The practice shows that combining simple phrase search and "similar document" search allows you to successfully and with a greater benefit apply the full-text search software in information systems from DMS to ERP and PDM.

All in all, SearchInform is a product that has duly deserved the attention it is given. Its "mix" of a large number of functions (similar search leading the chain), high speed and quality performance only proves it. On top of all that, SearchInform lives up to and satisfies at least the nominal demands of the corporate user. Considering the product's break-neck development (no longer than one year on the market) and its transformation from an average desktop search engine into a presentable corporate system, SearchInform can become quite a demanded part of any enterprise's information system.