

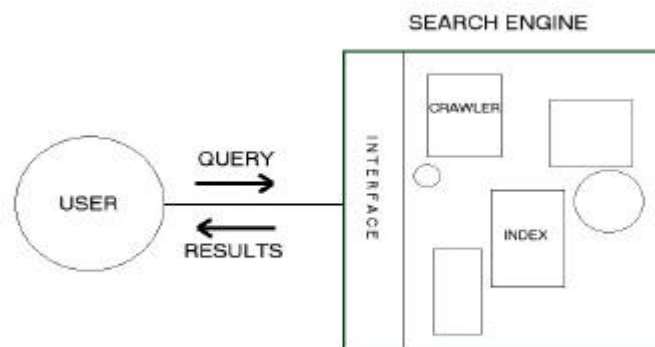
Search engines: do you speak their language?

A search for ideal interface

Petar Perkovic
Faculty of Philosophy
Department of Information science

Introduction

The Internet is an enormous source of information, but because that information is not globally classified and easily available, it makes it hard to find it and access it. This makes a search for any type of information rather complicated, and it is necessary to have a “mediator” in the form of a search engine (SE), and a “mediator of a mediator” in the form of a SE interface. The SE interface connects the user with the rest of the structure that we call a search engine. In this way, the access to information is made drastically easier for every user, especially for the beginners.



A simplified view of a search engine

However, a fact of the matter is that we are still far from the day when a person who has little or no experience with a computer will in a few seconds be able to find an airplane line for the quickest and the most practical way to his destination. Since there is a larger percentage of Internet users who access the Net relatively rarely and to find specific information, SE interface is moving more and more towards plainer and simpler form. The ideal solution would be an absolutely simplified SE interface, which would consist of a keyword entry box, and a “search/find” button. This would be ideal solution for most of today’s users and, from a technological point of view, very easy to achieve. But, while outlining the interface some

important factors are being overlooked: a simplified interface usually results in larger amount of found information in a search based on a user's query, and thus more time is needed to refine and find useful information. Since one of prerequisites of a good search is the smallest possible time spent doing it, it is also important that this information should be retrieved in as smaller time frame as possible. Seeing as that which a standard SE is offering, search results are very often far from what the user has set as a goal of a query. The quickest possible access to any amount of information requested without wasting time and wading through unwanted data is still a goal to be achieved. Considering the speed in which the Internet is growing and changing, it is obvious that any kind of processing enormous amounts of data accessible on the Net is not going to be an easy task.

A simple search interface consisting of a rectangular text input box on the left and a rectangular button labeled "Submit" on the right.

A typical simple search interface consisting of a keyword box and a button

Of query and its results

Every query that results in a huge number of hits impossible to fit on one screen cannot be called a successful query. That is why it's necessary to create an interface with a larger configurability for user to use, giving it a certain level of freedom to modify it by its own needs and wishes. Because of the fact that creating a user-accessible crawler would be unpractical, and allowing user access to main SE database could be contraproductive, interface remains as the only SE element that can be offered to user as a tool. On the other hand, seeing that a large number of users are beginners, the SE interface should have its simple form as well. Consequently – two interface types, both simple and advanced, with as larger level of configurability from user's side as possible, are required for fast and quality searching to satisfy the users of every skill. There is only one person who knows what the search engine user wishes to find – the user himself.

A great amount of SE resources is invested in manipulation of results obtained by the user's query. By sorting the results, the corresponding data is able to surface. But if a part of these resources could be directed towards the process of creating a better query, the final number of results would be substantially smaller and therefore easier to manage. This can be accomplished by creating more power features inside the user's interface, which would give him more control over the search area and would allow him more detailed definition of his search. For example, to my knowledge there isn't a search engine that allows a partial search of its database. This type of search would allow usage of more common concepts as keywords and at the same time keep the number of results at levels user will be able to analyze. This way a smaller amount of not wanted results would be returned, and the search itself would last less.

With the additional options inside the pre-search interface, it is necessary to enable additional refining of search results by different elements that can be found in every web page.

Comparing the search engines

The comparison table consists of five whom, according to publicly available statistics, are among the 10 best and biggest global search engines. Concrete statistical facts like number of pages contained in indexes, the number of users and similar, were not at my disposal, since they are considered a sort of strategic information and therefore kept secret. However, the table consists of a comparison of basic and advanced features those 5 SE have (or don't have). Some of data is inconclusive, but the basics were available via SE online help. The sixth column contains characteristics and basic features of CROSS, Croatian search engine. CROSS is designed to search and find information contained inside Croatian web space, and primarily uses Croatian language. This engine has no ambition to compete with SE portals like Yahoo! or Lycos, but if the needed data can be found on the web pages inside .hr domain, CROSS is without a doubt a much better choice than a global SE. CROSS has a simple/advanced user interface and uses primarily Croatian language, which makes it easy to use and ideal for beginners.

Observing the table, the conclusion can be made that most of the current large SE agree on the necessary options for today's user, considering the present state of Internet and the profile of a typical user.

| | Altavista | Ask Jeeves | Google | Hotbot | Yahoo | CROSS |
|----------------------------------|-----------|------------|--------|--------|-------|-------|
| simple/advanced search | + | - | + | + | + | + |
| RESULTS RANKING | | | | | | |
| Link popularity ranking | + | ? | + | + | ? | - |
| Direct hit ranking | - | ? | - | + | ? | - |
| META ranking? | - | - | - | ? | ? | + |
| QUERY TYPES | | | | | | |
| Special symbols (+ - *) | + | + | + | + | + | + |
| Boolean operators (AND, OR, NOT) | + | - | - | + | + | + |
| Advanced operators (NEAR i ()) | + | + | + | + | + | + |
| Free query | + | + | + | + | + | - |
| Form (GUI) | + | - | + | + | - | + |
| Supports tag search | + | - | + | + | + | + |
| Search by title | + | - | ? | + | + | + |
| Search by author | + | - | ? | + | - | + |
| Search by language | + | - | + | + | - | - |
| Search by date | + | - | - | + | - | - |
| Search by URL | + | - | + | + | + | + |
| Search by domain | + | - | + | + | - | - |
| OTHER FEATURES | | | | | | |
| Find similar | + | + | + | + | + | - |
| Case sensitive | + | - | - | + | - | + |
| Stemming | + | + | + | + | + | + |
| Clustering | + | - | + | + | - | - |

What must an “ideal” interface posses?

A rather small percentage of users are actually satisfied with the effect of a SE and the relevancy of given results. Can this be changed or at least improved? The basic idea is that the interface should be simple, but at the same time user-configurable to the certain level. Although this seems contradictory, it is possible to achieve a certain level of simplicity, while sustaining the limited level of user features. One of more popular methods to achieve this is to create more graphics as a substitute for textual options. Graphical solutions allow faster manipulation of the interface menus seeing that it is easier to handle icons and pictures than reading lines of text. By using graphics, one of the larger problems is very elegantly bypassed – the usage of (English) language is reduced to minimum. Even though English language is accepted as the official language of the Internet, a large number of people in the world still do not speak it, nor do they want to learn it. As those persons are also potential information seekers a graphic-based SE would intuitively lead them through actions of writing the keyword, commencing the search and receiving the required information in a language familiar to them.



A simple version of a user-friendly search engine interface

A string of powerful post-search features is also an important element of the “ideal interface”. The possibility of sorting by various elements of web page (e.g. meta-tags, author) is an element that must not be overlooked.

Conclusion

Strictly speaking, there are not enough resources spent on the quality of the results themselves. Seeing as the process of classifying Internet resources has yet to become a global project, the only way of getting better results is by educating an individual in combination with creating as ideal interface as possible. The ideal interface: one that will be simple to use but with enough pre and post-search options to enable a high-grade query processing and allow retrieval of results which will match users query.

References:

1. Search Engine Shoot-out
<http://home.cnet.com/category/topic/0,10000,0-3817-7-276910,00.html>
2. Search Engine Failures
<http://www.zdnet.com/pcmag/insites/dvorak/jd981020.htm>
3. The Anatomy of a Large-Scale Hypertextual Web Search Engine
<http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm>
4. Direct Hit
<http://www.directhit.com/>
5. Search engine watch
<http://www.searchengineworld.com>
6. Recent developments in Indexing, Searching and Information Retrieval Technologies.
<http://www.terena.nl/projects/reis/isir/reisnews9908sendex.html>
7. Marti A. Hearst. "Interfaces for Searching the Web", Special Report Article in *Scientific American*, #3 1997.
<http://www.sciam.com/0397issue/0397hearst.html>