DIALOGUES WITH SPECIALISTS

Scientific article
UDC 513.2; DOI: 10.61260/2304-0130-2024-4-69-72

INTERNET SEARCH ENGINES

✉**Labinsky Alexander Yu.**
Saint-Petersburg university of State fire service of EMERCOM of Russia, Saint-Petersburg, Russia
✉***labinskyi.a@igps.ru***

Abstract. The features of search engines used to find information by keyword input in various information repositories including the Internet are reseached. Various search engines were investigated, including popular search engines Google and Yandex. The features of search engines are shown, including the main components and stages of information search related to the components of the search engine. The main types of search engines are considered in detail. Examples of practical implementation are provided for each type of search engine. The article focuses on the example of a search engine using fuzzy search. The advantages of fuzzy search in comparison with clear information search are given, in particular, taking into account such features of the query text as linguistic and semantic relationships, grammatical forms of query words, as well as possible errors and typos. For a fuzzy search engine implemented as a computer program, a flowchart of an algorithm for comparing a string with a query sample and a flowchart of information retrieval program interface events are presented.

Keywords: search engine, fuzzy search engine, keywords, types of search engines, components of search engines, string comparison algorithm, computer program

For citation: Labinsky A.Yu. Internet search engines // Supervisory activities and forensic examination in the security system. 2024. № 4. С. 69–72. DOI: 10.61260/2304-0130-2024-4-69-72.

Introduction

The search engine is a set of computer programs designed to search documents by keywords in various information repositories, including on the Internet. The search result can be web pages, images, audio files, etc. The quality of the search is assessed by the number of documents that are relevant (most relevant) to the user's query.

Currently, the most popular search engine in the world, including in Russia, is Google, whose algorithm was developed in the early 2000s by Sergey Brin and Larry Page, the founders of Google. The approximate number of Google servers in the mid-2000s was 1 million, the total capacity of Google data centers was estimated at 250 MW, and the cost of these data centers was about $2.5 billion per year.

Let us formulate the problem statement. It is necessary to consider the features of information search systems, including the features of the search engines used in the Internet. The article is relevant, as the amount of information posted on the Internet is constantly increasing. Therefore, a significant number of works are devoted to search engines [1–11].

Features of search engine operation

The main components of a search engine are a search robot, an indexer, and a search engine. Each component of the search engine implements a specific stage of information search.

At the first stage, the search robot downloads the data corresponding to the user's query.

The search index allows you to speed up the search for information as much as possible. As a result, an index database is created.

Yandex search engine uses the «Metasearch» search engine (program).

Some search engines allow for a fuzzy (approximate) search that takes into account the distance to keywords..

An example of a search engine using fuzzy search [10]

Fuzzy information retrieval during the search process takes into account the editing distance, which determines the number of editing operations that need to be applied to the search query string in order to account for all possible distortions, errors and typos.

The editing distance characterizes the degree of «similarity» of the lines.

In 1965, the Soviet mathematician V.I. Levenshtein [6] solved the problem of line similarity and proposed the currently widely used editing distance, which is calculated using the following algorithm: the first character of the text is compared with the query sample, if the characters match, the transition to the next character is performed, etc.

The comparison process stops if all characters match (the search stops) or there are mismatched characters.

A block diagram of the algorithm for comparing a string with a sample query is shown in fig. 1.

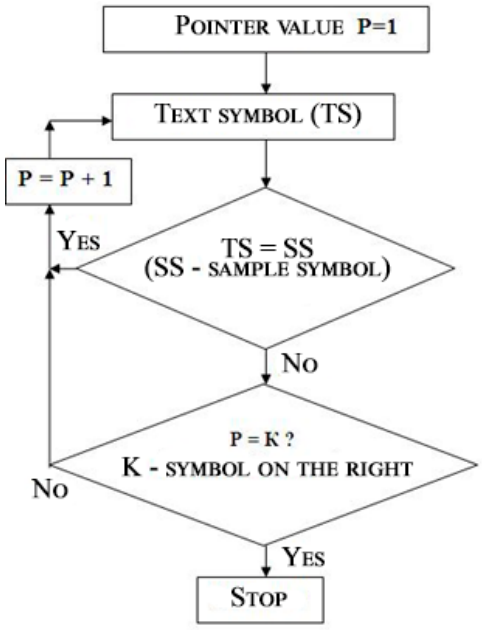

Fig. 1. **A block diagram of an algorithm for comparing a line with a sample query**

Computer model of fuzzy search

As an example of a fuzzy search system, a computer model is presented, borrowed in [11] and implemented as a computer program (fig. 2). The program allows for a clear and fuzzy search for test substrings in Internet files (file types *.htm, *.xml, *.txt). The text substring is searched in the files located in the specified directory.

In the text input line «Find substring:», the desired substring is entered and by clicking the «Search» button, the process of searching for files containing the specified substring is started. The names of the found files are placed in the «File to view» list.

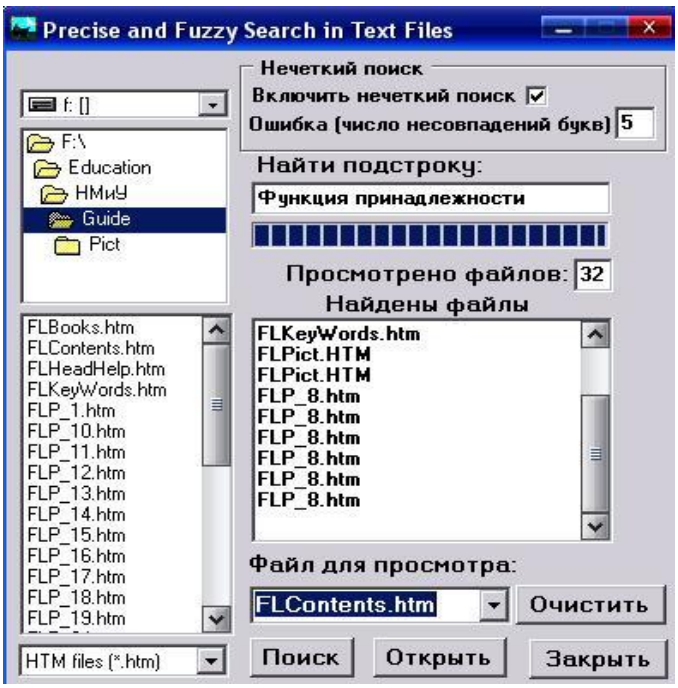

Fig. 2. **A clear and fuzzy information search program**

The block diagram of the information retrieval program interface events is shown in fig. 3.

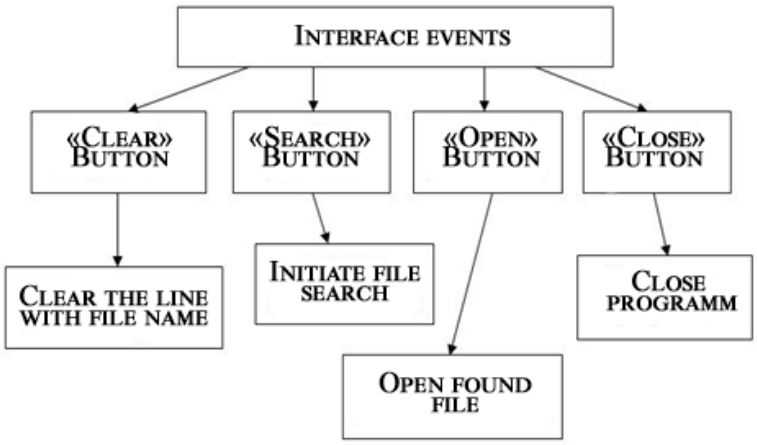

Fig. 3. **Block diagram of information retrieval program interface events**

Output

The features of search engines used to search for information by keywords on the Internet are considered. Fuzzy search engines have advantages over traditional clear systems, as they allow for a fuzzy (approximate) search that takes into account the distance to keywords.

List of sources

1. Ashmanov I.S., Ivanov A.A. Website promotion in search engines. M.: Williams, 2007.

2. Kirillov A.V. Search engines: components, logic, ranking methods // Business Informatics. 2009. № 4.

3. Kolisnichenko D.N. Search engines and website promotion on the Internet. M.: Dialectics, 2007.

4. Lande D.V. Search for knowledge in the Internet. M.: Dialectics, 2005.

5. Lande D.V., Snarsky A.A., Bezsudnov I.V. Internetics: navigation in complex networks: models and algorithms. M.: Librocom, 2009.

6. Levenshtein V.I. Binary codes with correction of character dropouts and insertions // Problems of information transmission. 1965. № 1.

7. Chu H., Rosenthal M. Search engines for the World Wide Web // American Society for Information. 2006. Vol. 33.

8. Risvik K.M., Michelsen R. Search engines and web dynamics // Computer Networks. 2002. Vol. 39.

9. Tarakeswar M.K., Kavitha M.D. Search Engines // Journal of Computer Applications. 2011. Vol. 4.

10. Sonkin M.A., Leshik Yu.V. Application of fuzzy search algorithms in forest fire monitoring systems // Trudy TSTU. 2012. № 2.

11. Labinsky A.Yu. Fuzzy search for textual information // Natural and man-made risks (physico-mathematical and applied aspects). 2019. № 2. P. 42–45.

Information about the article: submitted for editing: 26.10.2024; accepted for publication: 19.11.2024

Information about authors:
Labinsky Alexander Yu., associate professor of the applied mathematics and information technology chair of Saint-Petersburg university of State fire service of EMERCOM of Russia (196105, Saint-Petersburg, Moskovskiy ave., 149), Phd in technical sciences, associate professor e-mail: labynsciy@yandex.ru, https://orcid.org/0000-0001-2735-4189, SPIN: 8338-4230