

Maxim Zakharov

Current Address: 31A-35 Alpijskaya St, Sochi, 354057, Russia

E-mail: maxime@maxime.net.ru, Site Web: www.maxime.net.ru, Phone number: +7 9882 377057.

OBJECTIVE A full-time position in software development with areas of interest that include networking applications, search engines, information retrieval, or similar.

EDUCATION Moscow State University, Moscow, Russia
Faculty of Computational Mathematics and Cybernetics.
M.S. in Computer Science and Applied Mathematics (major in Computation Systems Software), 1993.

SKILLS SUMMARY *Programming Languages:* C, C++, Perl, PHP, SQL, HTML, XML, CSS, JavaScript, Assembler, Fortran, BASIC.

Tools and Technologies: SDLC, CVS/SVN, autoconf/automake, GNU debugging and profiling tools (gdb, gprof), PostgreSQL, MySQL, Perl DBI, FTP, HTTP (Apache + mod_perl), TCP/IP, CGI, L^AT_EX, Apache modules in C and Perl.

Operating Systems: Unix, FreeBSD, Linux, Solaris, MacOS X (Darwin), Windows

Specific Interests: Open Source, multithreaded and multiprocess programming, problem solving, code optimisation and troubleshooting.

Language Skills: spoken and written Russian (native); spoken and written English (advanced).

EXPERIENCE *Director / Lead Developer* **DataPark Ltd.**
August 2002 – present

DataparkSearch Engine:

Developing and supporting *the DataparkSearch Engine*, a full-featured open sources web-based search engine written in C language and released under GPL. It's designed to organize the search within a web-site, group of web-sites, intranet or local system and works on many UNIX-like operation systems with PostgreSQL or MySQL SQL-servers. The list of key features is available at [DataparkSearch's home page](#).

Neo Popularity Rank:

Invented and implemented *the Neo Popularity Rank*, a link-based ranking algorithm to order the Web. The Neo Popularity Rank allows to highlight most cited documents in the collection indexed. It's similar to Google's PageRank, but it based on the model of neural network, i.e. it supposes all pages are neurons and links between pages are links between neurons. This method is used in the DataparkSearch Engine to adjust relevance score in the search results ordering, and in the SEA algorithm to mark out automatically the most important sentences for the summary of the document. The short description of the Neo Popularity Rank can be found in the article at www.maxime.net.ru/doc/Neo-en.pdf.

TREC-2005 evaluation:

The DataparkSearch Engine was evaluated in the [Genomics Adhoc Track](#) of TREC-2005. The DataparkSearch's model of relevance was compared against other information retrieval systems. Two relevance methods available in DataparkSearch, the Full and the Fast, were compared also against each other.

According to the results achieved, see trec.nist.gov/pubs/trec14/papers/datapark.geo.pdf, the Full method of relevance calculation gives better performance, as expected. But the Fast method is only a little worse, so it may be a better choice for a large collection in the combination of speed and relevance performance.

SEA algorithm:

Implemented the Summary Extraction Algorithm (SEA), an algorithm of the automatic document summarizing. This algorithm is based on the ideas of Rada Mihalcea and Paul Tarau described in the paper “[An Algorithm for Language Independent Single and Multiple Document Summarization](#)”, in [Proceedings of the International Joint Conference on Natural Language Processing \(IJCNLP\), Korea, October 2005](#). The Neo Popularity Rank is used as a ranking algorithm in the SEA. DataparkSearch assigns a higher weight for words from these summaries. That improves the quality of the search. Also the search domain can be limited by these summaries.

Phrase segmenter for sentences in East-Asian languages:

It's based on the segmenter of the Linguistic Data Consortium (LDC) and the University of Pennsylvania (UPENN) written in Perl. DataparkSearch's segmenter was translated into C language and was modified for better performance (as for speed as for better precision and recall), so it slightly outperforms the original LDC segmenter and also it outperforms the S-MSRSeg segmenter from Microsoft Research. See evaluation results at projectile.is.cs.cmu.edu/research/public/tools/segmentation/eval/

mod_dpsearch:

DataparkSearch Engine can be incorporated into the [Apache http-server](#) as a mod_dpsearch module. This allows preload all data and templates which are used to answer search queries and that reduces the average response time. The mod_dpsearch module is implemented for both 1.x and 2.0.x branches of the Apache http-server.

Language and charset detection:

An automatic charset and language guesser has been implemented for DataparkSearch. At this moment, it can recognize more than 80 various charsets and 25 language groups. However, a special learning mode has been implemented to construct a map for new languages and character sets. The detection of charset and language is implemented using “[N-Gram-Based Text Categorization](#)” technique. The DataparkSearch uses 3,6-grams to improve the quality of the recognition for both single and multi-byte encodings, and also uses the smart technique of early rejection for irrelevant candidates to reduce the working time.

Memory debugging:

Implemented the system for memory debugging and leaks detection, which is used to test and control the quality of DataparkSearch. It's based on [Electric Fence 2.2](#) by Bruce Perens, is ported to FreeBSD Operating System and is incorporated into the DataparkSearch's source tree to be compiled in the special memory debugging mode. This system was very helpful in the catching of several wired bugs in memory handling. Also it allows to check memory operations before releasing a new version of DataparkSearch to avoid such problems. The similar debugging system was implemented for file descriptors.

sochi.org.ru:

Planned and coordinated the development of the [Sochi city search engine and directory of the sites](#). Maintained the categorized collection of links to Sochi city related sites. Wrote a bundle of scripts in Perl to check automatically the accessibility of the sites in the directory, to update click counts on the resources in the directory, to gather the news about Sochi from numerous feeds and sources

around the World. The DataparkSearch Engine had been deployed as a search engine over resources collected in this directory and that engine indexes Sochi's internet deeper than any conventional World-span search engine. The [LancOS](#) server for gathering data from air temperature sensors was ported to FreeBSD OS. A similar server was implemented to gather the data from air pressure sensors.

Apache modules:

Wrote a bundle of Apache modules in Perl for branding context advertisements onto web-pages on-the-fly. This allows to monetize published content without any modification of the files in the repository.

Developer

**mnoGoSearch development team
September 2001 – November 2003**

Maintained project's code in C language by tracking defects and feature requests. Picked-up and maintained the cache mode storage. Improved the algorithm of automatic language and charset guessing. Implemented phrase segmenters for Chinese, Japanese and Thai languages. The project has been ported to Apple Darwin OS. Implemented various functionality improvements (include the meta-clusterisation, the ispell-based fuzzy search, the support for content-encoding, the storage for indexed document copies), made source code optimisations and troubleshooting. Updated project documentation.

Automation manager

**Bank of Moscow, Sochi branch
September 1999 – September 2001**

Maintained the accounting system with strong periodical deadlines. Planned and coordinated the development and upgrade for local computer network and for computer equipments (including the Y2K problem troubleshooting). Coordinated setup of a WAN network for inter-office communications and plastic card authorisation. Installed and maintained the client-bank system. Developed the firewall protection for the Internet connection. Wrote a bundle of scripts in RSL language for out-of-time reporting.

Automation manager

**Mosbusinessbank, Sochi branch
September 1997 – September 1999**

Maintained the accounting system with strong periodical deadlines. Maintained the client-bank system. Deployed and maintained the firewall protection for the Internet connection. Wrote a bundle of scripts in RSL language for out-of-time reporting.

Sr. Automation Engineer

**Mosbusinessbank, Sochi branch
October 1996 – September 1997**

Maintained the accounting system with strong periodical deadlines. Maintained the client-bank system. Planned and coordinated the development and upgrade for the local computer network. Wrote a bundle of scripts in RSL language for out-of-time reporting.

Automation Engineer

**UnionCard Sochi Ltd.
April 1995 – October 1996**

Developed and deployed the installation of the plastic cards authorisation center.

Programmer

**MENATEP Bank, Sochi branch
June 1994 – April 1995**

Maintained the accounting system with strong periodical deadlines. Deployed and maintained the inter-office communication network via Sprint Network.

Researcher

**Russian Academy of Science,
Sochi research center
August 1993 – June 1994**

Took part in the development under the "Informatisation of Russia" program. Supported and maintained the city public UUCP network. Wrote a set of small programs and scripts to gather data for the billing system.

ACTIVITIES

Developed [43°N 39°E](#), the search engine as a demo site of DataparkSearch's features. Designed a new layouts of search engine result page, see [www.43n39e.ru/andromeda](#) and [www.43n39e.ru/janus](#). Designed a nice type of graph to visualize the distribution of the Neo Popularity Rank for all sites indexed. See graphs for [Russian sites](#), [Australian sites](#) and [New Zealandian sites](#).

Editorial: blogs [SoNotes](#) and [Founds](#); [the Sochi city history timetable](#); [the World history timetable](#).

Invented and implemented the [Yet Another Realisation of Multiplication Free Arithmetic Code](#), C source.

REFERENCES

MR. MICHAEL BABAKOV	MR. DENIS SHAKALO
Company: DataPark Ltd.	Company: Bank of Moscow, Sochi branch
Role: Technology Director	Role: Sr. Systems Engineer
e-mail: mike@datapark.ru	e-mail: den@sochi.mmbank.ru

LinkedIn profile: <http://www.linkedin.com/in/dpmaxime>