Software engineer and technical lead with strong problem solving skills and experience working at a leading search engine on search quality as well as on crawling and distributed indexing. I have a keen interest in project management, software engineering and machine learning.

Software Engineer & Tech Lead

Experience

09.2013 – Present

Yandex, Moscow & University of Amsterdam
Researcher, PhD Candidate

Yandex is the leading internet company in Russia, opearting the most popular search engine and the most visited website. Yandex sites attracted 52 million unique visitors in December 2012.

Joint PhD program between Yandex, and the University of Amsterdam (UvA) co-supervised by Prof. Maarten de Rijke (UvA) and Pavel Serdyukov (Yandex).

I am working both in the ISLA research group at the UvA, where I spend about 30% of my time, and in Yandex research in strong collaboration with the search quality team. My research focuses on aggregated search, freshness, and online machine learning and I liaise with different teams to find the best ways to improve Yandex's systems on a daily basis. I have written one full paper (accepted at ECIR 2014 [1]), and submitted another one to SIGIR 2014 on freshness and online learning. I have attended the following conferences so far: ECIR 2013; CIKM 2013; WAW 2013, and I am a reviewer for IPM (journal).

[1] D. Lefortier, P. Serdyukov, F. Romanenko, and M. de Rijke. Blending Vertical and Web results. ECIR 2014.

09.2011 - Present

Yandex, Moscow
Software Engineer & Tech Lead, Search Team

Member of the content system team, whose role is to crawl and index the Web for Yandex's search engine.

Initially, I was working on our production Web crawler, in which I implemented many new features (e.g. online algorithm to compact B-tree based databases, algorithm to detect newly created spam websites), and which I also helped to maintain and troubleshoot (C++). I proposed, designed and implemented new quality metrics of this system in Python using Map/Reduce that are now widely used when deploying new features. Moreover, I worked on a research project that led to the implementation of new crawling strategies greatly improving the freshness of Yandex's search engine, and I co-authored two conference papers on this [2,3]. I have received multiple bonuses from Yandex for this work.

Then, I joined a new project, which is a rewrite of the highly critical distributed system responsible for building Yandex search indexes – from the first design stages to implementation (C++, Python). This system runs on hundreds of nodes in different data centers, has strict uptime requirements and operates at a really staggering scale. I became the official Tech Lead after few months. I took a significant role in reviewing code and ensuring that best practices were followed across the team. I have been taking many initiatives that shaped the system. Since September 2013, I am still working on this project part-time.

[2] D. Lefortier, L. Ostroumova, E. Samosvat, and P. Serdyukov. Timely Crawling of High-quality Ephemeral Content. CIKM 2013.
[3] D. Lefortier, L. Ostroumova, and E. Samosvat. Evolution of the Media Web. WAW 2013.

02.2010 - 08.2011

Exalead, Paris
Software Engineer & Tech Lead, Web Team

Exalead provides search platforms and search-based applications for consumer and business users. Exalead also has a public search engine.

Member of the Web team whose role is to scrap the Web both for business users and for Exalead's search engine.

I started as an intern supervised by Hugo Zanghi working on a visual search engine written in Java. I have completely re-written the Web server using Netty to improve the performance of the service, and implemented different online graph clustering and layout algorithms specifically designed for the Web graph -- including the ones used now by default in production (3 months).

Then, I led the design and implementation of a new system to automate data mining tasks inside the team in Java (2-3 people). This system succesfully allowed to automate virtually any task that was done semi-manually before. I also helped to maintain and troubleshoot the distributed system crawling the Web for Exalead's search engine, which was written in Java and C++.

01.2008 - 01.2010

LRDE (EPITA Research & Development Laboratory), Paris
Research Assistant

LRDE allows EPITA's best students to discover the world of academic research by getting involved in its projects.

I worked under the supervision of Alexandre Duret-Lutz on the development of Spot, an open source model checking library written in C++ (~80k SLOC). I have been presenting my work in front of the lab on multiple occasions and wrote technical reports about my contributions.

Education

2013-Present — University of Amsterdam (UvA, The Netherlands). PhD in Computer Science.

Information Retrieval, Machine Learning, Statistics, Optimization.

2005-2010 — Ecole pour l'Informatique et les Techniques Avancées (EPITA, France). MS in Computer Science. Specialization: Artificial Intelligence.

Machine Learning, Math, Text & Data Mining, NLP, Game Theory, Operating Systems, Distributed Systems, Network Programming (not inclusive).

2008 — Indian Institute of Technology, Kanpur (IITK, India). Student Exchange Program, Computer Science and Engineering.

Fuzzy Logic, Information Theory, Quantum Computing.

Skills

- Extensive industrial experience with C/C++, Java, Python, Bash and OCaml.
- Web Search, Information Retrieval, Machine Learning, Data Mining.
- Distributed Systems, Scalability, Concurrent Programming.
- Project Management, Experience competing Priorities.
- Map/Reduce.

Languages

Native French, fluent English (TOEIC 980), fluent Russian.

Personality

Hard-working, creative, self-motivated, high learning capability, strong team player, highly capable of leading the development of software projects.

Interests

Photography, literature, languages, traveling.