

Junte Zhang

✉ juntezhang@gmail.com

🌐 www.zhang.nl



Making new things possible with search engines

Personal Data

Gender Male

Nationality Dutch

Education

2007–2011 **Doctorate (PhD)**, *University of Amsterdam*, Amsterdam, the Netherlands.

Dissertation: System Evaluation of Archival Description and Access.

Worked on the evaluation of information retrieval for digital archives of the *National Archives of the Netherlands*. Designed and implemented retrieval algorithms and multiple front-ends of a search engine. Processed and analyzed log files to understand user search behavior and to evaluate the search engine with information retrieval experimentation.

2006–2007 **Master of Science**, *University of Amsterdam*, Amsterdam, the Netherlands.

Human Centered Multimedia (Information Studies).

Major: multimedia systems, including video search. Master's degree thesis based on joined work with researchers of the *National Research Institute for Mathematics and Computer Science* on the credibility of information sources in search engines; work published and presented at major international venues.

2002–2006 **Bachelor, Master of Arts**, *University of Groningen*, Groningen, the Netherlands.

Information Science, note: Master's degree with honors, cum laude.

Major: language technology. Worked for Master thesis with the University Library to design and implement a Semantic Web system for their bibliographic catalogs; published and presented at international conferences.

Work Experience

January 2021 – **Principal Search Engineer (Sr. Staff Engineer)**, *Optimizely*, Amsterdam, the Netherlands.

Now

Team lead and technical lead on design and implementation of “headless” search solutions in a GraphQL (GQL) API labeled as OPTIMIZELY GRAPH used to build static-site generators and dynamic apps with (faceted) search.

Implemented reliable and high performance search-as-a-service with search infrastructure automation with auto-scaling using a Kubernetes operator, index snapshots, monitoring and alerts. Implemented semantic (hybrid) search with vector embeddings with pre-trained language models, including automated production pipeline. Created dynamic indexing schemas to support search features in GQL schemas, e.g. filtering, full-text search with language analysis, query expansion with synonyms, faceting, auto-complete, efficient exports, boosting, special ranking functions. Implemented application-side joins in the search engine to support automated nested linked results.

○ *Technologies used*: Elasticsearch, OpenSearch, C#, Typescript, GraphQL, Go, Python, Java, Bash scripting, CI/CD with Docker/Kubernetes/Azure DevOps/Github Actions, Datadog, Git, JIRA.

May 2019 – **Sr. Search Engineer (Sr. Software Engineer)**, *Elsevier*, Amsterdam, the Netherlands.

December

2020

Ported, proposed, improved and created new search solutions in Life Sciences. Design and implementation of the migration from MarkLogic database to Elasticsearch for EMBASE, which is the leading commercial biomedical and pharmacological bibliographic search engine in the market. Migration successfully started and completed during my tenure, directly resulting in NPS improvement.

Revised and extended indexing schemas to support search functionalities. Re-writing XQuery queries to equivalent but more efficient and accurate Elasticsearch DSL queries, i.e. exact and phrase search, fulltext search, efficient query expansion with synonyms, proximity queries with search fields and wildcards. Implemented advanced (visual and nested) faceting, efficient exports of large amounts of records and highlighting. Improved relevance ranking.

○ *Technologies used*: Elasticsearch, MarkLogic, Java, Python, Bash scripting, XSLT, Antlr, Kafka, AWS S3, CI/CD with Docker/Kubernetes/Jenkins/SonarQube, Git, JIRA.

April 2017 – **Sr. Search Engineer (Sr. Software Engineer)**, *Swisscom Directories*, Zurich, Switzerland.

April 2019

End-to-end responsibility for designing, planning, implementation and testing search algorithms on LOCAL.CH, which is the leading online directory platform (phone book and Yellow Pages) in Switzerland. Collaboration in creating search features to improve the performance of listed businesses and relevance of directory listings to users.

Proposed, designed and implemented a pipeline from Big Query to Solr by calculating near real-time trending, popular and top-rated businesses near a user for delivering recommendations. Implemented more semantic query recommendations. Contributed to operational excellence by improving and optimizing the intricate matching and ranking of a highly customized search engine.

○ *Technologies used*: Solr, Elasticsearch, Java, Python, Bash scripting, Javascript, MongoDB, Google BigQuery, Jupyter Notebook, AWS S3, Kafka, CI/CD with Docker/Kubernetes/Bamboo, Git, JIRA.

- Nov 2014 – **Sr. Search Engineer (Sr. Software Engineer)**, *LexisNexis*, Leiden, the Netherlands.
- Mar 2017 Technical lead on the design, implementation and evaluation of search engines (e.g. ingestion of terabytes of data, query parsing, retrieval algorithms) for bibliographic patent records and in cross-language full-text for TOTALPATENT ONE, which is a leading commercial intellectual property search engine in the market. Coordinating leading role in the delivery and technical evaluation of implementations by external consultants and contractors.
Designed infrastructure topology for superior search efficiency, designed and implemented expert patent (legal) search functionality such as stemmed vs. non-stemmed search, proximity search, case-sensitive search, cross-language retrieval (from English to CJK), document filtering based on document IDs, fields and user-generated content, advanced classification search, and patent family de-duplication algorithms.
- *Technologies used*: Elasticsearch, Java, Node.js, AngularJS, Python, Bash scripting, XSLT, Antlr, OpenStack, MongoDB, RabbitMQ, AWS S3, Jenkins/Bamboo, Git, JIRA.
- Jan 2014 – **Software Engineer, Information Analyst**, *Vancis Advanced ICT Services*, Amsterdam, the Netherlands.
- Nov 2014 Technical lead on software solutions for Cloud services. Business development for these services and introduction of new Cloud services, in particular Platform-as-a-Service (PaaS). Set up MoSCoW analyses as part of RFIs and RFPs. Identified business processes and KPIs by interviewing colleagues and extracting information from systems, storing and integrating different information streams for metering and billing of cloud services. Design, recommendation and implementation of changes with automation.
- *Technologies used*: Bash scripting, PHP, Python, (My)SQL, SOAP, CLI-based APIs, OpenStack, OpenShift, SVN, Git.
- Nov 2012 – **Information Specialist**, *NIOD Institute (KNAW)*, Amsterdam, the Netherlands, 0.2 FTE.
- Dec 2013 Worked on the European Holocaust Research Infrastructure (EHRI) project. EHRI provides online access to information about dispersed sources relating to the Holocaust through its Online Portal, and tools and methods that enable researchers and archivists to collaboratively work with such sources.
- *Technologies used*: Perl, JSON, XML, XSLT, Solr, Git.
- Oct 2011 – Dec 2013 **Software Engineer**, *Meertens Institute (KNAW)*, Amsterdam, the Netherlands.
- Design and development of advanced search engines and user interfaces for a plethora of complex and large amounts of research datasets with different metadata (i.e. Dublin Core, TEI, WSDL, DIDL, custom markup) for scientists as part of a European project. I developed and delivered software to provide digital access to historical editorial cartoons in newspapers with a grant by the Royal Netherlands Academy of Arts and Sciences.
- *Technologies used*: Solr, PHP, Perl, jQuery, XML, XSLT, Bash scripting, SQL, R, SVN.
- Oct. 2007 – **Lecturer**, *University of Amsterdam*, Amsterdam, the Netherlands, 0.25 FTE.
- June 2011 Independently teaching the course “Representation and Retrieval” – mainly on using XML for describing, presenting, and gaining access to personal document collections, including preparing slides, giving lectures, providing feedback to students, supervising and grading course work.
- July 2007 – **Junior Researcher**, *University of Amsterdam*, Amsterdam, the Netherlands, 0.75 FTE.
- June 2011 Member of the Retrieving Encoded Archival Descriptions More Effectively (README) project funded by the Netherlands Organisation for Scientific Research (NWO). The project was about investigating the use and impact of “electronic” encoding of archival description in the standard Encoded Archival Description (EAD) for access to archives from both a system and user-centered perspective. Besides my doctoral dissertation “System Evaluation of Archival Description and Access”, it resulted in numerous publications and presentations at large top-tier scientific conferences on information retrieval in Europe, the United States and Asia.
- *Technologies used*: XML databases, Lucene, Indri, Perl, SPSS, Bash scripting, L^AT_EX en BibT_EX.

Professional and Technical Skills

Research	Experienced with (quantitative) research following the scientific method, including experimental user studies, reporting & presenting the results.	Application Stack	Experienced with Linux; comfortable with creating and working with SOAP, RESTful and CLI-based APIs and frameworks like Java Spring.
Programming and Coding	Java, Python, C#, Typescript, Go, Perl, PHP, Bash scripting, XSLT, XQuery, SQL, JQ, L ^A T _E X. I write clean code with code coverage.	Other	European driving license B.

Natural Languages

Dutch	Fluent, both oral and in writing.	German	Limited working proficiency.
English	Fluent, both oral and in writing.	Chinese	Limited working proficiency.

Miscellaneous

- Hobbies: watching movies & TV-shows, making sushi, traveling.
- Additional information with descriptions of my completed projects, selected publications, and recommendations can be found on my LinkedIn profile: www.linkedin.com/in/juntezhang.