

Junte Zhang

✉ juntezhang@gmail.com
🌐 www.zhang.nl

Information technologist specialized in search engines and analytics who builds bridges across technologies, disciplines and between people

Personal Data

Gender	Male	Date of birth	July 20, 1982
Nationality	Dutch	Marital status	Single

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.
- 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 of Groningen to design and implement a Semantic Web system for their bibliographic catalogs. Work published and presented at international conferences.

Work Experience

- April 2017 – **Sr. Search Engineer**, *Swisscom Directories AG*, Zürich, Switzerland.
now End-to-end responsibility for designing, planning, implementation and testing search algorithms on local search (phone book, Yellow Pages, restaurants and other SMEs). Daily collaboration in creating search functionalities with a team of back-end engineers, product managers, user experience experts, designers and web developers.
- Nov 2014 – **Sr. Search Engineer**, *LexisNexis*, Leiden, the Netherlands.
Mar 2017 Technical lead on information retrieval, in particular design (infrastructure & application) and back-end development of search engines (e.g. ingestion of terabytes of data, query parsing, retrieval algorithms) for patent data in bibliographic form and in cross-language full-text.
Designed infrastructure topology for superior search efficiency, designed and implemented expert patent search functionality such as stemmed vs. non-stemmed search, proximity search, case-sensitive search, cross-language retrieval, document filtering based on document IDs, fields and user-generated content, advanced classification search, and patent family de-duplication algorithms.
○ *Technologies used*: Java, Node.js, AngularJS, Python, Bash scripting, XSLT, Elasticsearch, OpenStack, MongoDB, RabbitMQ, Jenkins/Bamboo, Git, JIRA.
- Jan 2014 – **Software Developer, Information Analyst, Technical Lead**, *Vancis Advanced ICT Services B.V.*,
Nov 2014 Amsterdam, the Netherlands.
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 for software solutions 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*: Hadoop, OpenStack, OpenShift, Bash scripting, Perl, PHP, Python, jQuery, (My)SQL, SOAP, CLI-based APIs, UML, SVN, Git.

- Nov 2012 – **Information Specialist**, *NIOD Institute for War, Holocaust and Genocide Studies, Royal Netherlands Academy of Arts and Sciences, Amsterdam, the Netherlands, 0.2 FTE.*
 Dec 2013 Worked on the European Holocaust Research Infrastructure (EHRI) project, working with archivists and (international) IT specialists on a digital research environment for historians.
 ◦ *Technologies used:* Perl, JSON, XML, XSLT, Apache Solr, Git.
- Oct 2011 – Dec 2013 **Software Developer**, *Meertens Institute, Royal Netherlands Academy of Arts and Sciences, Amsterdam, the Netherlands.*
 Design and development of advanced search engines and user interfaces for a plethora of complex and large amounts of structured data (i.e. Dublin Core, TEI, WSDL, DIDL, custom markup) for scientists. I delivered software to provide digital access to historical editorial cartoons in newspapers.
 ◦ *Technologies used:* Apache Solr, PHP, Perl, jQuery, XML, XSLT, Bash scripting, SQL, R, SVN.
- Aug. 2011 – **Consultant**, *VLC Projects B.V., Amsterdam, the Netherlands.*
 Sept. 2011 Developed ideas on enterprise search and worked on technical solutions. Worked on content management in Alfresco and Drupal. Gave trainings and did presentations.
- 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, 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), which resulted in numerous publications and presentations at large international conferences.
 ◦ *Technologies used:* XML databases, Lucene, Indri, Perl, SPSS, Bash scripting, L^AT_EX en BibT_EX.
- Feb. 2006 – **Intern**, *University Library Groningen, Groningen, the Netherlands.*
 July 2006 Worked on the development of a Semantic Web based on bibliographic information stored in MARC21 records at the Digital Library department.
 ◦ *Technologies used:* XML, XSLT, Perl, Protégé, Sesame, SeRQL.

Professional and Technical Skills

Research Experienced with quantitative and qualitative research following the scientific method, including experimental user studies, reporting on and presenting the results in scientific articles and at major international conferences.

Coding and Programming Java, Python, Perl, PHP, Bash scripting, R, Javascript, XSL, XQuery, SQL, L^AT_EX.

Application Stack Experienced with working in Linux; use of Microsoft Office products; crafting production-ready search engines with user-friendly graphical interfaces in Lucene/Solr/Elasticsearch; comfortable with working with SOAP, RESTful services, Web and CLI-based APIs and libraries.

Other European driving licence B.

Natural Languages

Dutch Fluent, both oral and in writing.

English Fluent, both oral and in writing.

German Limited working proficiency.

Chinese Fair, only oral.

Grants and Awards

- *Association for Computers in the Humanities (ACH)* Young Scholar bursary in 2008; \$1000.
- *European Conference on Digital Libraries (ECDL)* Student Scholarship in 2009; €360.
- *e-Humanities* grant in 2012 by the Netherlands Academy of Arts and Sciences to work on access to historical editorial cartoons in newspapers; €30K
- *Reed Tech Innovation Award 2016* for innovative contributions to project Summit and the subsequent commercial release of TotalPatent One; handed out one time per year to an employee.

Miscellaneous

- Hobbies: movies, making sushi, traveling.
- Interested in cultural heritage (storytelling, identity), history and science fiction.
- Additional information with descriptions of my completed projects and selected publications can be found on my LinkedIn profile: nl.linkedin.com/in/juntezhang.

Publications

- [1] Junte Zhang, Kees Ribbens, and Rob Zeeman. The digital archiving of historical political cartoons: An introduction. In Carsten Eickhoff and Arjen P. de Vries, editors, *DIR*, volume 986 of *CEUR Workshop Proceedings*, pages 52–53. CEUR-WS.org, 2013.
- [2] Junte Zhang, Marc Kemps-Snijders, and Hans Bennis. The CMDI MI Search Engine: Access to language resources and tools using heterogeneous metadata schemas. In Panayiotis Zaphiris, George Buchanan, Edie Rasmussen, and Fernando Loizides, editors, *TPDL*, volume 7489 of *Lecture Notes in Computer Science*, pages 492–495. Springer, 2012.
- [3] Junte Zhang. Supporting serendipitous and focused search. In Max L. Wilson, Tony Russell-Rose, Birger Larsen, and James Kalbach, editors, *EuroHCIR*, volume 909 of *CEUR Workshop Proceedings*, pages 79–82. CEUR-WS.org, 2012.
- [4] Junte Zhang. System evaluation of archival description and access. *SIGIR Forum*, 45(2):109–110, 2011.
- [5] Junte Zhang. *System Evaluation of Archival Description and Access*. SIKS dissertation series 2011-22, University of Amsterdam, 2011.
- [6] Junte Zhang and Jaap Kamps. A search log-based approach to evaluation. In Mounia Lalmas, Joemon M. Jose, Andreas Rauber, Fabrizio Sebastiani, and Ingo Frommholz, editors, *Research and Advanced Technology for Digital Libraries (ECDL 2010)*, volume 6273 of *Lecture Notes in Computer Science*, pages 248–260. Springer, 2010.
- [7] Junte Zhang and Jaap Kamps. Search log analysis of user stereotypes, information seeking behavior, and contextual evaluation. In *IiX '10: Proceeding of the third symposium on Information interaction in context*, pages 245–254, New York, NY, USA, 2010. ACM.
- [8] Junte Zhang. XML retrieval for digital archives. *Bulletin of IEEE Technical Committee on Digital Libraries*, 6(1), 2010. <http://www.ieee-tcdl.org/Bulletin/current/Zhang/zhang.html>.
- [9] Khairun Nisa Fachry, Jaap Kamps, and Junte Zhang. The impact of summaries: What makes a user click? In Maarten van der Heijden, Max Hinne, Wessel Kraaij, Maria van Kuppeveld, Suzan Verberne, and Theo van der Weide, editors, *Proceedings of the 10th Dutch-Belgian Information Retrieval Workshop (DIR 2010)*, pages 47–54. Radboud Universiteit Nijmegen, 2010.
- [10] Junte Zhang and Jaap Kamps. Focused search in digital archives. In Gottfried Vossen, Darrell D. E. Long, and Jeffrey Xu Yu, editors, *Web Information Systems Engineering (WISE 2009)*, volume 5802 of *Lecture Notes in Computer Science*, pages 463–471. Springer, 2009.
- [11] Junte Zhang and Jaap Kamps. Searching archival finding aids: Retrieval in original order? In Maristella Agosti, José Luis Borbinha, Sarantos Kapidakis, Christos Papatheodorou, and Giannis Tsakonias, editors, *ECDL*, volume 5714 of *Lecture Notes in Computer Science*, pages 447–450. Springer, 2009.
- [12] Junte Zhang, Alia Amin, Henriette S. M. Cramer, Vanessa Evers, and Lynda Hardman. Improving user confidence in cultural heritage aggregated results. In James Allan, Javed A. Aslam, Mark Sanderson, ChengXiang Zhai, and Justin Zobel, editors, *SIGIR*, pages 702–703. ACM, 2009.
- [13] Alia Amin, Junte Zhang, Henriette S. M. Cramer, Lynda Hardman, and Vanessa Evers. The effects of source credibility ratings in a cultural heritage information aggregator. In Katsumi Tanaka, Xiaofang Zhou, and Adam Jatowt, editors, *WICOW*, pages 35–42. ACM, 2009.
- [14] Junte Zhang and Jaap Kamps. A content-based link detection approach using the vector space model. In Shlomo Geva, Jaap Kamps, and Andrew Trotman, editors, *INEX*, volume 5631 of *Lecture Notes in Computer Science*, pages 395–400. Springer, 2008.
- [15] Junte Zhang, Khairun Nisa Fachry, and Jaap Kamps. Effective access to digital assets: An XML-based EAD search system. In *Proceedings of the 2nd Symposium on Digital Curation: Practice, Promise and Prospects (DigCCurr 2009)*, pages 49–56. University of North Carolina at Chapel Hill, 2009.
- [16] Khairun Nisa Fachry, Jaap Kamps, Rianne Kaptein, Marijn Koolen, and Junte Zhang. The University of Amsterdam at INEX 2008: Ad hoc, book, entity ranking, interactive, link the wiki, and XML mining tracks. In *INEX 2008 Workshop Pre-proceedings*, pages 66–92, 2008.

- [17] Khairun Nisa Fachry, Jaap Kamps, and Junte Zhang. Access to archival material in context. In *IliX '08: Proceedings of the second international symposium on Information interaction in context*, pages 102–109, New York, NY, USA, 2008. ACM.
- [18] Junte Zhang and Jaap Kamps. Link detection in XML documents: What about repeated links? In Andrew Trotman, Jaap Kamps, and Shlomo Geva, editors, *Proceedings of the SIGIR 2008 Workshop on Focused Retrieval*, pages 59–66. University of Otago, Dunedin New Zealand, 2008.
- [19] Junte Zhang, Khairun Nisa Fachry, and Jaap Kamps. Access to archival finding aids: Context matters. In Birte Christensen-Dalsgaard, Donatella Castelli, Bolette Ammitzbøll Jurik, and Joan Lippincott, editors, *ECDL*, volume 5173 of *Lecture Notes in Computer Science*, pages 455–457. Springer, 2008.
- [20] Ismail Fahmi, Henk Ellermann, Peter Scholing, and Junte Zhang. Improving Berrypicking Techniques Using Semantically Similar Information in a Historical Digital Library System. In Stefanos Kollias and Jill Cousins, editors, *Workshop on Semantic Interoperability in the European Digital Library (SIEDL 2008) at ESWC 2008, Tenerife, Spain, June 2, 2008, Proceedings*, pages 2–13, 2008.
- [21] Junte Zhang, Khairun Nisa Fachry, and Jaap Kamps. Automatic link-detection in encoded archival descriptions. In Lisa Lena Opas-Hänninen, Mikko Jokelainen, Ilkka Juuso, and Tapio Seppänen, editors, *Digital Humanities 2008, Conference Abstracts*, pages 226–228, Oulu, Finland, 2008. The University of Oulu.
- [22] Ismail Fahmi, Peter Scholing, and Junte Zhang. Digitally Deducing Information about the Early American History with Semantic Web Techniques. In *Digital Humanities 2008 (DH 2008), Oulu, Finland, June 25-29, 2008, Proceedings*, pages 244–246, 2008.
- [23] Khairun Nisa Fachry, Jaap Kamps, Marijn Koolen, and Junte Zhang. Using and detecting links in wikipedia. In Norbert Fuhr, Jaap Kamps, Mounia Lalmas, and Andrew Trotman, editors, *INEX*, volume 4862 of *Lecture Notes in Computer Science*, pages 388–403. Springer, 2007.
- [24] Khairun Nisa Fachry, Jaap Kamps, Marijn Koolen, and Junte Zhang. The University of Amsterdam at INEX 2007. In Norbert Fuhr, Mounia Lalmas, and Andrew Trotman, editors, *Pre-Proceedings of INEX 2007*, pages 388–402, 2007.
- [25] Junte Zhang, Ismail Fahmi, Henk Ellermann, and Gosse Bouma. Mapping Metadata for SWHi: Aligning Schemas with Library Metadata for a Historical Ontology. In Mathias Weske, Mohand-Said Hacid, and Claude Godart, editors, *Web Information Systems Engineering - WISE 2007 Workshops, WISE 2007 International Workshops, Nancy, France, December 3, 2007, Proceedings*, volume 4832 of *Lecture Notes in Computer Science*, pages 103–114. Springer, 2007.
- [26] Ismail Fahmi, Junte Zhang, Henk Ellermann, and Gosse Bouma. SWHi System Description: A Case Study in Information Retrieval, Inference, and Visualization in the Semantic Web. In Enrico Franconi, Michael Kifer, and Wolfgang May, editors, *The Semantic Web: Research and Applications, 4th European Semantic Web Conference, ESWC 2007, Innsbruck, Austria, June 3-7, 2007, Proceedings*, volume 4519 of *Lecture Notes in Computer Science*, pages 769–778. Springer, 2007.