

Jie Gao

44(0) 741 27 82628

jiegaouk@gmail.com

<http://staffwww.dcs.shef.ac.uk/people/J.Gao/>

Msc (Hons), Bsc (Hons), MBCS, CSM

General

- **Coding, analysing, troubleshooting** large scale distributed system, **10+ years solid commercial experience**
- **Research Scientist** with >5 years research experience in the field of NLP and have 9 publications in top-tier peer-reviewed NLP/ML conferences and/or journals (e.g., ICLR, KDD, ASONAM, LREC, IEEE/ACM, etc).
- **Information Extraction** and Natural Language Processing (**NLP**) R&D experience, **3+ years**
- Rich experience in developing and deploying multi-tier application in **JEE, Python and Node JS** application development environment (Java >6.x, Flask, RESTful web service, Tomcat 6.x, MySQL 5.x, PM2, etc.)
- **commercial/research experience** (Java, Python, GATE/JAPE Platform, NLTK, scikit-learn, Allennlp, Pytorch)
- **NLP-based Machine learning** (feature engineering, language modelling, supervised training & diagnose, unsupervised learning/clustering, evaluation) for Information Extraction (Named Entity Recognition, Entity typing, Sentiment Analysis) and social media analysis (rumor detection), 2+ years research experience
- Working with **Cloud technology** in Linux/Ubuntu environment (Amazon EC2, Sheffield Iceberg HPC cluster, SolrCloud, Zookeeper, NGINX), 3+ years R&D experience
- Developing large-scaled **semantic web applications** (RDF, RDFS, OWL, SPARQL) with Graph API technologies (Sesame, Jena or python RDFLib package), 3+ years R&D experience

RELEVANT EXPERIENCE

Research Associate, OAK Group, University of Sheffield

Mar 2014 – Present

- Involved in various research project, including European Foundation funded Europeana DSI-3, two European FP7 projects (WeSenseIt and SETA), two Innovative UK (prev. TSB) funded projects (SPEEAK-PC 101947 with TATA Steel UK and KTP project 007326 for ActiveStandards, the market-leading SaaS platform), three industry/government/academic collaboration projects (with K-Now, JustGiving, MoveMore and Active10)
- 2 years as Teaching Assistant for “COM3504 The Intelligent Web” module (>70 students)
- Invited speaker/presentation (4 times) in international conference / workshop
- Leading the development and am the manager of the scalable, computational and economically efficient infrastructure that collects the data in large-scale from the mobile phones in EU project SETA and Active10 projects (over 129K monthly active users and over 600,000 downloads over past 10 months) on behalf of Public Health England.
- co-PI of Europeana DSI-3 addendum project
- leading & coordinating the research for work package (WP) 3 (“Data Representation, Mining & Analysis”) and responsible for the delivery for two work packages (WP3 and WP4 – terminology driven text mining and knowledge discovery application) in SPEEAK-PC collaboration project with TATA Steel funded by Innovative UK
- peer reviewer for academic journals / conferences / workshops, such as Journal of Natural Language Engineering, ESWC, USES, PLOS ONE
- providing cloud computing environment administration and maintenance for EU/National funded research projects and our group VMs for PhD research in the lab
- technical support for PhD students/colleagues (e.g., data collection via low-cost sensors, database systems, software engineering, web servers and service hosting)

Semantic R&D developer, Crownpeak (prev. ActiveStandards), London

Oct 2013 – Jan 2014

- Follow-up research and development of KTP project to build innovative semantic products on ActiveStandards platform, covering
 - **Semantic Annotation & Indexing** (GATE platform, Sesame/OWLIM 5.x (now GraphDB), Ontotext KIM platform)
 - **Semantic technologies** for Enterprise Information/content integration (Named Entity Recognition, entity linking, RDF, Ontology)
 - **Semantic SEO** and Web Compliance (corpus pre-processing & indexing, keywords extraction, RDFa, Web Accessibility)

KTP Associate, University of Westminster, London

Oct 2011 – Oct 2013

To undertake the Knowledge Transfer Partnership (KTP) objectives as defined in the KTP proposal covering the development of new semantic systems and services on ActiveStandards platform, in order to assist businesses and organizations to discover, understand and act on the actionable knowledge within their information repositories.

The R&D project (KTP project 007326) is funded by the UK Technology Strategy Board. I was the KTP associate and the lead R&D engineer in the project. The main technologies in this project covers the area of :

- Knowledge extraction (more specifically cross-domain NER and relation extraction with GATE),
- Knowledge management and retrieval (RDF, RDFS, OWL) with OWLIM-SE (now rebranded as GraphDB)
- Entity linking and knowledge base enrichment (via linked open data) and Enterprise Information Integration (ontology normalisation)
- Web compliance (SEO and fact checking)

The partnership project has been graded as second higher level which is shortlisted as one of KTP case-studies among 250 national funded projects.

Senior Software Engineer/Java developer, Dextrys, Suzhou, China

Aug 2006 – Jun 2010

- Engage in large scaled enterprise application development with complicated business requirements
- Involved projects include Chartis Brazil Insurance System, Trinity Project for Tripwire, Hanover Insurance system for Hanover, EA Back Office System for Electric Arts Inc., Mayban Fortis InsuranceSystem for eBaoTech.com, Non-Domestic management System for Hong Kong Housing Authority and Greece Reporting SMS System(GRSMS) for Proctor & Gamble
- Main duties
 - Design, enhance application solution and work closely with business analysis team
 - Transform clients' idea and business concepts into carefully designed and well-authorised software engineering solutions;
 - Work closely with UI design team to ensure the usability of the software;
 - Coordinate with QA Test Engineers on Module business testing, System Integration Test, Regression Test, and so forth during the whole software development lifecycle;

EDUCATION & RELEVANT COURSEWORK

M.S. in Web Technology, University of Southampton, Southampton, England

Sept 2011

- Thesis: "Linked Data based enterprise content management"
- **Graduated with Merit**
- **Subject studies:** Enterprise Web Development, Advance Topics on Web Services, Hypertext and Web technologies, Assistive Technologies and Universal Design, Intelligent Agent, Semantic Web Technologies, Rich Internet Applications.
- **Skills developed:** Enhanced the knowledge and understanding of important technologies and issues in designing and building enterprise web systems; decent English written and oral communication; research, interpretation and presentation; ability to analysis and evaluate issues regarding universal design and assistive technologies; Enhanced hands-on experience with building web services in Java; ability to design an agent-based system and evaluate the suitability of an agent-based approach to a given problem; Enhanced the ability to evaluate difference RIA technologies;

B.S. in Computer Science and Technology, Yancheng Institute of Technology, Yancheng, China

Sept 2006

- Dissertation: "Struts based online shopping system"
- Grade: 2:1

Certificates and Training

- Convolutional Neural Networks (License DWLCCDHKCE2N), Coursera (deeplearning.ai), April 2018
- Structuring Machine Learning Projects (License 8JUA2EYXDHAN), Coursera (deeplearning.ai), November 2017
- Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization (License K39M5D7B9Q6Z), Coursera (deeplearning.ai), November 2017
- Neural Networks and Deep Learning (License BXJWXH56N6TZ), Coursera (deeplearning.ai), October 2017
- Machine Learning, Stanford University through Coursera Inc., May 2015
- Data Mining with Weka, University of Waikato, August 2014
- Associate Certificate (License KTP007326-A02), UK Technology Strategy Board, Oct 2013

- M101J: MongoDB for Java Developers, “10gen, The MongoDB Company”, July 2013
- Certified ScrumMasters (License 000285406), GoodAgile, Oct 2013
- 5th GATE Training Course Track 3 (Advanced GATE), University of Sheffield – 2012
- CMI Level 5 in Management, Ashorne Hill Management College, 2011 - 2012

PUBLICATIONS

- Gao, J., Han S., Song X., Ciravegna, F. (2020). “RP-DNN: A Tweet level propagation context based deep neural networks for early rumor detection in Social Media”, In: *The LREC 2020 Proceedings. The International Conference on Language Resources and Evaluation, 11-16 May 2020, Marseille*. LREC 2020 (submitted).
- Han S., Gao, J., Ciravegna, F. (2019). "Neural Language Model Based Training Data Augmentation for Weakly Supervised Early Rumor Detection", *The 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2019)*, Vancouver, Canada, 27-30 August, 2019 [Acceptance Rate: 14%]
- Han S., Gao, J., Ciravegna, F. (2019). "Data Augmentation for Rumor Detection Using Context-Sensitive Neural Language Model With Large-Scale Credibility Corpus", *Seventh International Conference on Learning Representations (ICLR) LLD*, New Orleans, Louisiana, US
- Ciravegna, F., Gao, J., Ireson, N., Copeland, R., Walsh, J., & Lanfranchi, V. (2019, May). "Active 10: Brisk Walking to Support Regular Physical Activity." In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 11-20). ACM.
- Ciravegna, F. Gao, J., Ingram, C., Ireson, N., Lanfranchi, V., & Humasak S. (2018). "Mapping Mobility to Support Crisis Management." *15th Proceeding of International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. ISCRAM Association. 2018.
- Zhang, Z., Gao, J., & Ciravegna, F. (2018). SemRe-Rank: Incorporating Semantic Relatedness to Improve Automatic Term Extraction Using Personalized PageRank. *ACM Transactions on Knowledge Discovery from Data (TKDD)*.
- Yuan, Y., Gao, J., & Zhang, Y. (2017, October). Supervised Learning for Robust Term Extraction. In *The proceedings of 2017 International Conference on Asian Language Processing (IALP)*. IEEE.
- Gao, Jie, and Suvodeep Mazumdar. "Exploiting linked open data to uncover entity types." In *Semantic Web Evaluation Challenge*, pp. 51-62. Springer International Publishing, 2015.
- Zhang, Z., Gao, J. and Ciravegna, F. (2016) [JATE2.0: Java Automatic Term Extraction with Apache Solr](#). In: The LREC 2016 Proceedings. The International Conference on Language Resources and Evaluation, 23-28 May 2016, Slovenia. LREC 2016.
- Zhang, Z., Gao, J., and Gentile, A.L. *The LODIE team (University of Sheffield) Participation at the TAC2015 Entity Discovery Task of the Cold Start KBP Track*. In: Proceedings of the 2015 Text Analysis Conference. TAC Knowledge Base Population (KBP) 2015, 16-17 Nov 2015, Gaithersburg, Maryland USA.

SKILLS & ATTRIBUTES

- **Strong written & oral communication, presentation and demonstrating skills** with confident, articulate and professional speaking ability;
 - Work as PI in European DSI-3 addendum project
 - Work as core developer and coordinator in PHE active10 project
 - Work as the sole associate & researcher in two Innovate UK (prev. TSB) funded projects for four years, both of which are rated as second higher score
 - Deliver regular presentations and demonstrations in project meetings and international conferences
- Able to communicate effectively with the diverse range of people and **strong teamwork ability in multicultural environment**
 - Good team player with >10 years' experience in multicultural working environment for both commercial and academic research projects
 - Work as an embedded member of ActiveStandards team and as a KTP associate in Semantic Computing lab of University of Westminster to complete two-years KTP project with a series of challenging project milestones
 - Work in large EU projects (WeSenseIt and SETA) involving partners from more than 5 countries
 - Complete CMI Level 5 training in Management in Ashorne Hill Management College
 - Trained as a Certified Scrum Master (CSM)
- **Self-directed, Dynamic and Adaptable:**
 - Quickly adapt to new projects or tasks

-
- Researched and Developed a machine learning based entity typing algorithm in python and got 2nd place in OKE 2015 challenge in less than 2 months, despite no previous python experience
 - Researched and Developed a distant supervision learning based sentiment classifier for Italian in one month and further enhance the prototype as a commercial product, despite no previous sentiment analysis experience
 - **Experience** in Java 6.x, Java EE, Python 3.x, Matlab, Javascript, NodeJS, XML/XSD, Database (Solr 5.x/6.x/7.x, MySQL 5.x, OWLIM/GraphDB), Web Service (SOAP, WSDL, RESTful), NLTK, scikit-learn, Information Retrieval (Lucene, Apache Solr 5.x), ORM(Hibernate 3.x, SQLAlchemy), Spring 2.x/3.x (DI/IoC), HTTP Server(Flask, Tomcat 7.x), RDF framework (Sesame or Jena), Cloud computing (SolrCloud, Zookeeper, security)
 - Good **skills in agile methodologies**: TDD (Junit, mockito), Continuous Integration (Bamboo, Jenkins)
 - More than 2 years commercial experience in an AGILE Scrum based environment (KTP project with ActiveStandards development team and PHE active10)
 - **Software/OS proficiency**: Eclipse, IntelliJ, Git, SVN, MATLAB, WEKA, GATE Platform, protégé, Linux/Ubuntu, AWS EC2
 - **Languages**: Chinese (Native), English (working proficiency)
 - **Work eligibility**: Settlement – INDEFINITE LEAVE TO REMAIN