

# Jie Gao

44(0) 741 27 82628

[j.gao@sheffield.ac.uk](mailto:j.gao@sheffield.ac.uk)

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

<https://uk.linkedin.com/in/jie-gao-a8a68631>

## RELEVANT EXPERIENCE

Research Assistant, OAK Group, University of Sheffield

Mar 2014 – Present

- Working in European FP7 SETA project (<http://setamobility.eu/>), FP7 WeSenseIt Research Project ([www.wesenseit.com](http://www.wesenseit.com)), the SPEEAK-PC industry project (<http://staffwww.dcs.shef.ac.uk/people/F.Ciravegna/Speak-PC>) funded by Innovate UK, and Move More Sheffield initiative. Proficient in:
  - Coding, analysing, troubleshooting** large scale distributed system, **7+ years solid commercial experience**
  - Information Extraction** and Natural Language Processing (**NLP**) R&D experience, **3+ years**
  - Developing and deploying multi-tier application in **JEE** application development environment (Java 6.x or above, SOAP/RESTful web service, Tomcat 6.x, MySQL 5.x)
  - commercial/research experience** (java, python, GATE/JAPE Platform, NLTK, scikit-learn)
  - Machine learning** (feature engineering, supervised training & diagnose, unsupervised learning/clustering, evaluation) for Information Extraction (Named Entity Recognition, Entity typing, Sentiment Analysis), 2+ years research experience
  - Working with **Cloud technology** in Linux/Ubuntu environment (Amazon EC2, Sheffield Iceberg HPC cluster), 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
- Teaching Assistant for “COM3504 The Intelligent Web” module (>70 students)

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 which covers the area of :

- Information Extraction & Retrieval
- RDF Databases (Sesame, OWLIM-SE), Semantic Web and Linked Data
- Semantic technologies for Enterprise Information Integration, and Web compliance

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

## PUBLICATIONS

- 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)*. To appear
- 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 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 2<sup>nd</sup> 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 (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)
  - **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)
  - **Software/OS proficiency:** Eclipse, IntelliJ, Git, SVN, MATLAB, WEKA, GATE Platform, protégé, Linux/Ubuntu, AWS EC2
  - **Languages:** Chinese (Native), English (working proficiency)
  - **Work eligibility:** Tier-2 working visa