CURRICULUM VITAE ET STUDIORUM

Istituto Di Linguistica Computazionale 'A. Zampolli' Email: fahad.khan@ilc.cnr.it
Consiglio Nazionale delle Ricerche
Via Moruzzi 1
Pisa, Italy

Education

Ph.D. Computer Science, University of Nottingham, 2007-2012. Supervisor: Dr Natasha Alechina.

Dissertation: Regular Path Temporal Logic and its Variants.

M.Sc. Mathematical Logic, University of Manchester, 2004 -2005.

Dissertation: The Guarded Fragments of First Order Logic. Supervisor: Dr George Wilmers.

Courses Taken Included: Predicate Logic, Non-monotonic Logics, Introduction to Grammatical Theory, Topics in Pragmatics, Types for Programs and Proofs

B.Sc. (Hons) Computer Science, *Upper Second*, University of Glasgow, 1998 - 2002.

Dissertation : An Equational Reasoning Assistant Using Haskell. Supervisor: Dr John O'Donnell.

Courses Taken Included: Artificial Intelligence, Advanced Programming, Neural Networks, Databases, Graphics and Multimedia, Operating Systems. Languages learned: Java, C, C++, HTML.

Research Biography

My research career was initially focused on theoretical computer science. For my doctoral thesis I studied dynamic temporal logics for modelling software resources (see V, VI).

Since starting at ILC I have mainly concentrated on the design, development and use of formal ontologies in the context of computational linguistics and also within the general area of the digital humanities. This has led to my working on standards for the representation of diachronic lexical and lexicographic information in LMF. At the beginning of my career at ILC I was involved with the IMAGACT project and worked on the derivation of an ontology of action types from an English language corpus annotated for action verbs (see 20, 18, 17, 13). At the same time I also became interested in linguistic linked data and the numerous possibilities it offered for textual and lexical scholarship. I was involved in the publication of part of a legacy lexical dataset, the Parole-Simple-Clips lexicon (see IV), and am currently working on converting the Perseus TEI-DICT version of the Liddell Scott Jones dictionary into linked open data (see 9).

I am also engaged in the creation of a model for representing diachronic lexico-semantic data in RDF, with a special application to classical languages (see 8, 7, 15, 1) as well as the conversion of important historical philological/lexicographic resources into XML using the lexical markup framework LMF (see 3).

As a result of these activities, I was able to participate as an expert in the W3C OntoLex working group and have been recently nominated as the project leader of the ISO 24613 LMF thematic domain group, devoted to the extension of the Lexical Markup Framework with diachronic information. I have also co-organised a workshop on such topics as part of one the major conferences of the sector (see c).

I have been interested in the use of ontologies to model natural language meaning (see 10, 19). Recently I have been working on the development of narrative ontologies within the context of a semantic web based querying system to facilitate research on literary texts (see 5).

As part of my work on the Memorata Poetis project, I have worked on the conversion of a taxonomy of themes and motifs used to index a corpus of annotated texts into an OWL ontology in order to facilitate better querying and analysis of the texts in question (see 4, II, 2).

I have been working with my colleagues at ILC to investigate the documentation and citation of language resources (see 6) in linguistic research with a specific focus on Italian language resources (see 11), resources for Ancient Greek (see 9) and Classical Arabic (see 3). These efforts are set against the background of a larger project carried out at ILC-CNR regarding the use of infrastructures and web services for digital humanities and computational philology (see III).

Academic Positions

Assegno di Ricerca Istituto Di Linguistica Computazionale 'A. Zampolli', December 2016 - Current

Title: PARTHENOS Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies.

Role: Specialist on ontologies for the humanities.

Details:Ontologies for the Humanities, in particular for the area of language resources: collection of practices and use cases for documentation tools and digital linguistic dataset catalogs for Digital Humanities; study of formal patterns used; analysis of models for the industry work flows; modeling domain ontologies, their expression in the Semantic Web languages and their integration into a common ontology for the contents of the Cultural Heritage.

Assegno di Ricerca Universitá Ca'Foscari Venezia \Istituto Di Linguistica Computazionale 'A. Zampolli', December 2015 - December 2016

Title: Memorata Poetis (BANDO Rep. n. 538/2015 Prot. n. 52694-III/13)

Role: Models and semantic resources to enhance Memorata Poetis search engine.

Details: The post-doctoral research project foresees the ongoing development and refinement of the ontology created during my previous post doctorate, as well as its application in the advanced querying and research on the annotated corpus of Memorata Poetis with a special focus on the Latin

corpus. I am also involved in the use of formal and quantitative models for the semantic study of the texts annotated during the Memorata Poetis Project.

Assegno di Ricerca, Istituto Di Linguistica Computazionale 'A. Zampolli', Consiglio Nazionale delle Ricerche, October 2013 - December 2015

Title: Memoria Poetica e Poesia della Memoria. Ricorrenze Lessicali e Tematiche nella Versificazione Epigrafica e nel Sistema Letterario (BANDO N. 126.048\ASS.008.2013 PROT. 0000831 del 29.08.2013)

Role: Development of formal models and computational instruments for the morphosyntactic and semantic analysis of Greek, Latin and Arabic literary texts.

Details:Memorata Poetis is a project focusing upon the study of multilingual intertextuality between epigraphic and literary texts, which have been manually annotated in order toidentify relevant themes and motifs. The languages treated in the project are ancient Greek, Latin, Arabic and Italian.

I helped revise and restructure the semantic relations between the themes and motifs which have been so far identified, in order to develop a more efficient and well structured taxonomy in order to assist in the analysis and study of the annotated text. To this end I created an OWL ontology containing the themes and motifs arranged according to ontology modelling principles.

Further activities: I helped to develop an RDF model based on the lemon model for representing diachronic lexico-semantic information called lemonDIA. I was an active participant in the Ontolex W3C community working towards a new version of the *lemon* model.

Assegno di Ricerca, Istituto Di Linguistica Computazionale 'A. Zampolli', Consiglio Nazionale delle Ricerche, September 2012 - September 2013.

Title: IMAGAct: Ontologia per Immagini dell'azione per l'acquisizione de L2 in Ambiente Multilingue. (BANDO N. 126.048\ASS.003.2012 PROT. 0000662 del 26.07.2012)

Role: Validating an interlinguistic ontology of action based on images; verification of action types and of corpus instances assigned to them; evaluation of their productivity in English by means of competence-based methodologies.

Details: The ImagACT project aimed to create a repository of action verbs, ontologically organised around prototypical action scenes in the form of both video recordings and 3D animations. The ImagACT project was funded in Italy by the PAR/FAS program of the Tuscany Region and was a joint undertaking between the University of Florence, ILC-CNR, Pisa, and the University of Siena. I was involved in designing an annotation scheme for the corpus and in studying how action types could be extracted from the annotated data.

Further activities: Conversion of the Parole Simple Clips Lexicon into RDF for publication as linked data.

Research

Journals

I Mugelli, G., Boschetti, F., Del Gratta, R., Del Grosso, A., Khan, F., Taddei, A. (2016, Forthcoming).A user-centered design to annotate ritual facts in Ancient Greek tragedies. Bulletin of the Institute of Classical Studies. Volume 59, Issue 2.

II Khan, F., Arrigoni,S., Boschetti, F., Frontini,F. (2016) Restructuring a Taxonomy of Literary Themes and Motifs for More Efficient Querying. In Materialities of Literature, Vol 4. No. 2, 2016.

III Del Gratta, R., Boschetti, F., Del Grosso, A., Khan, F., Monachini, M. (2015). Cooperative Philology on the Way to Web Services: The Case of the CoPhiWordNet Platform. In International Workshop on Worldwide Language Service Infrastructure (pp. 173-187). Springer International Publishing

IV Del Gratta, R., Frontini,F., Khan, F., Monachini, M. (2015). Converting the PAROLE SIMPLE CLIPS Lexicon into RDF with lemon. In Semantic Web Journal, 2015.

V Alechina, N., Dastani, M. , Khan, F. , Logan,B., and Meyer. J- J. Ch. (2010). Using theorem proving to verify properties of agent programs. In M. Dastani, K. Hindriks, and J-J. Ch. Meyer, editors, Specification and Verification of Multi-agent Systems, chapter 2, pages 134. Springer, 2010.

VI Khan, F. (2010).The Decidability of RPTL. STAIRS 2010: 151-161

Conference Papers

1. Khan, F., Bellandi, A,. and Monachini, M. (2016) Tools and Instruments for Building and Querying Diachronic Computational Lexica. In LT for DH: Language Technology Resources and Tools for Digital Humanities (COLING 2016).

2. Arrigoni, S., Khan, F., Monachini, M., and Boschetti, F. (2016). Misurare Memorata Poetis: prime statistische. AIUCD 2016 Book of Abstracts, 7-9 September 2016, Venezia, Italy

3. Nahli O., Frontini F., Monachini M., Khan F., Zarghili A., Khalfi M. (2016). Al Qamus al Muhit, a Medieval Arabic Lexicon in LMF. In Tenth International Conference on Language Resources and Evaluation (LREC 2016).

4. Khan, F., Frontini, F., Boschetti, F., Monachini, M. (2016). Converting the Liddell Scott Greek-English Lexicon into Linked Open Data using lemon. In Digital Humanities 2016: Conference Abstracts. Jagiellonian University and Pedagogical University, Krakw, pp. 593-596.

5. Khan, F., Bellandi,A., Benotto, G., Frontini,F., Giovannetti, E., Reboul, M. (2016). Leveraging A Narrative Ontology to Query a Literary Text. In 7th International Workshop on Computational Models of Narrative (CMN'16).

6. Del Gratta R., Frontini F., Monachini M., Pardelli G., Russo I., Bartolini R., Khan F., Soria C., Calzolari N. (2016). LREC as a Graph: People and Resources in a Network. In Tenth International Conference on Language Resources and Evaluation (LREC 2016).

7. Khan F., Daz-Vera J., Monachini M. (2016). Representing Polysemy and Diachronic Lexico-Semantic Data on the Semantic Web. In Second International Workshop on Semantic Web for Scientific Heritage (SW4SH 2016)

8. Khan F., Daz-Vera J., Monachini M. (2016). The Representation of an Old English Emotion Lexicon as Linked Open Data. In LDL 2016 5th Workshop on Linked Data in Linguistics: Managing, Building and Using Linked Language Resources (LREC 2016)

9. Boschetti, F., Del Gratta, R., Frontini,F., Khan, F., Monachini M. (2015). (Re)thinking the BLARK for Ancient Greek. In 4th Biennial Workshop on Less-Resourced Languages.

10. Khan, F., Frontini,F. (2015). Using Ontologies to Model Polysemy in Lexical Resources. Proceedings of IWCS 2015, London, UK.

11. Del Gratta, R., Frontini,F., Monachini M., Pardelli, G., Russo, I., Bartolini,R., Goggi, S., Khan, F., Quochi, V., Soria, C., Calzolari, N (2015). Visualising Italian Language Resources: a Snapshot

12. Khan, F., and Frontini,F. (2014). "Publishing PAROLE SIMPLE CLIPS as Linguistic Linked Open Data". Proceedings of CLIC-IT 2014, Pisa, Italy.

13. Moneglia, M., Brown,S., Frontini,F., Gagliardi,G., Khan, F., Monachini M., Panunzi, A. (2014). "The IMAGACT Visual Ontology. An Extendable Multilingual Infrastructure for the Representation of Lexical Encoding of Action". Proceedings of LREC 2014, Reykjavk, Iceland.

14. Del Gratta, R., Frontini,F., Khan, F., Mariani,J., Soria,C. (2014). "The LREMap for Under-Resourced Languages". Proceedings of CCURL 2014, Reykjavk, Iceland.

15. Khan, F., Boschetti, F. and Frontini,F. (2014). "Using lemon to Model Lexical Semantic Shift in Diachronic Lexical Resources". Proceedings of the 3rd Workshop on Linked Data in Linguistics (LDL-2014), Reykjavk, Iceland.

16. Del Gratta, R., Frontini,F., Khan, F., Mariani,J., Soria, C.(2014). The LRE Map for Under-Resourced Languages. CCURL 2014.

17. Russo, I., De Felice, I., Frontini,F., Khan, F. and Monachini M. (2013). (Fore)seeing actions in objects. Acquiring distinctive affordances from language. NLPCS 2013

18. Russo, I., Frontini,F., De Felice,I., Khan, F. and Monachini M. (2013). Disambiguation of Basic Action Types through Nouns' Telic Qualia. GL 2013.

19. Khan, F., Frontini,F., Del Gratta, R,., Monachini M. and Quochi, V. (2013). Generative Lexicon Theory and Linguistic Linked Open Data. GL 2013.

20. Frontini F., De Felice I., Russo I., Khan F., Monachini M., Gagliardi G. (2012). Panunzi A. Verb interpretation for basic action types: annotation, ontology induction and creation of prototypical scenes. In: CogALex-III - COLING 2012 - 3rd Workshop on Cognitive Aspects of the Lexicon (Mumbai, India, 15 Dicembre 2012). Proceedings, pp. 69 - 80. The COLING 2012 Organizing Committee, 2012.

Professional Activities

a Project Leader in ISO Working Group: Multi-part development of LMF - Part 3: Diachrony-Etymology

b Expert in the Ontolex W3C Community

c Member of Organising Committee for Joint Second Workshop on Language and Ontology and Terminology and Knowledge Structures (LangOnto2 + TermiKS). LREC 2016.

d Member of Program Committee for 4th Workshop on Linked Data in Linguistics: Multilingual Knowledge Resources and Natural Language Processing. LREC 2015.

e Member of Program Committee for 3rd Workshop on Linked Data in Linguistics: Multilingual Knowledge Resources and Natural Language Processing. LREC 2014.

f Involved in the drafting of several national and Horizon 2020 research proposals in my domain of expertise.

Training

1st Summer Datathon on Linguistic Linked Open Data (SD-LLD 2015); Cercedilla. Courses attended include: Performing Multilingual Word Sense Disambiguation and Entity Linking for the Web of Data; Use of NLP Interchange Format

The 11th Reasoning Web Summer School 2015; Berlin. Courses attended include: RuleML, Foundations of Answer Set Programming, Fuzzy Description Logics and Applications, Stream Reasoning

Foundations and Challenges of Change in Ontologies and Databases Research School, 29-31 January 2014. Courses attended include:Belief Revision in Description Logics, Temporal Dynamic Description Logic

The European Summer School in Logic, Language and Information (ESSLLI), 2008 Hamburg. Courses attended include: A Model theory for extensions of modal logic, A Modal logics for games and multi-agent systems

Miscellaneous

Languages Spoken: English (native), Italian (speaking, writing and reading C1 level), Urdu (spoken and written), French (basic).

IT and Programming Skills: Python, RDF, OWL, XML, TEI-XML, SQL, Prolog, SPARQL.

Pisa
Anas Fahad Khan