

How to apply:

To apply you can either apply online or download the University's application form from: http://www.ox.ac.uk/admissions/postgraduate_courses/index.html You will need to submit references, transcript and a statement of research interests (in the slot marked "research proposal")

You should quote the following studentship code:

10COMP-GG-WEB – if you have seen the advert on the Computing Laboratory web pages

10COMP-GG-JOBS – if you are applying from jobs.ac.uk

10COMP-GG-FIND – if you are applying from findaphd.com

All applications are subject to a £25 application fee.

Further Information: These studentships are associated with an ERC Advanced Investigators Grant recently awarded to Professor Georg Gottlob from the Oxford University Computing Laboratory (OUCL) for the project "DIADEM – Domain-centric Intelligent Automated Data Extraction Methodology". The aim of the DIADEM project is to provide the logical, algorithmic and methodological foundations for the knowledge-based extraction of structured data from websites belonging to *specific application domains within various industries and areas of general interest*. One core contribution will be a comprehensive multi-dimensional *logical data model* that will be used to simultaneously represent the content of a large website, its structure, user-interaction patterns and all meta-information and knowledge (factual and rule-based) that is necessary to automatically perform the desired extraction tasks. Moreover, new methods and algorithms will be designed for integrating existing machine-learning based approaches and methods for document analysis and annotation into a knowledge-based framework. It is envisioned that such foundations will enable us to build powerful systems that autonomously explore websites of a given domain, understand their structure, and extract and output richly structured data in formats such as XML and RDF, thus seeding a future Semantic Web. The breakthrough in automatic data extraction this project strives for would be most beneficial to two interrelated technologies which are the hottest emerging topics in web search: *vertical search*, web search in specialised domains; and *object search*, the search for web data objects rather than web page. The project will be directed by Professor Georg Gottlob and the project staff will part of the OUCL Information Systems Group whose faculty members are currently Prof. Michael Benedikt, Dr. Phil Blunsom, Prof. Georg Gottlob, Prof. Ian Horrocks, Dr. Dan Olteanu, and Prof. Stephen Pulman.

Selection Criteria: Applicants for these posts must satisfy the usual requirements, see <http://web.comlab.ox.ac.uk/oucl/prospective/dphil/dphil-criteria.pdf>, and should have a good background in theoretical computer science, web data processing, web programming, and should be skilled in the Java programming language and possibly also in C++. Finally, we require good writing, communication, presentation, and organization skills. Skills in one or more of the following areas will count as additional qualification: logical and algorithmic fundamentals of data extraction, design and study of new (logic-) programming languages, knowledge representation, machine learning, domain ontology design, semi-structured data, description logics, ontological reasoning, web site analysis, natural language processing, and software engineering.

Closing date: The closing date for applications has been extended to 30th April 2010

If you have any questions about the applications procedure please email Julie@comlab.ox.ac.uk