

Muhammad Faheem

*PhD in Computer Science and Networks
Database, Web*

University of Ottawa
800 King Edward
K1N 6N5, Ottawa, Canada
☎ +1 343 777 4960
✉ mfaheem@uottawa.ca
🌐 ssrg.site.uottawa.ca/~faheem/



Education

- 2011–2014 **PhD in Computer Science and Networks**, *Télécom ParisTech*, Paris, France.
Topic: Intelligent content acquisition in Web archiving
Supervisor: Prof. Pierre Senellart
- 2008–2010 **European Master's program in Computational Logic (double degree program)**, *Vienna University of Technology*, Austria; and, *Free University of Bozen-Bolzano*, Italy.
Topic: Implementation of Semantic Desktop Tool
Supervisor: Prof. Enrico Franconi
- 2003–2007 **Bachelor of Science in Computer Science**, *COMSATS Institute of Information Technology*, Lahore, Pakistan.
- 2000–2002 **Intermediate Computer Science (Maths, Statistics, Computing)**, *Forman Christian College*, Lahore, Pakistan.

Academic career

- Mar 2015–to
date **Post-doctoral researcher**, *University of Ottawa*, Ottawa, Canada.
Topic: Rich Internet Application (RIA) modeling and testing
Working with: Prof. Guy-Vincent Jourdan and Prof. Gregor v. Bochmann
- 2013/14 **Teaching assistant**, *Télécom ParisTech*, Paris, France, Databases.
- 2013/14 **Teaching assistant**, *Télécom ParisTech*, Paris, France, statistical computing with R Language.
- Feb 2013–
Apr 2013 **Visiting research student**, *Oxford University, Department of computer science*,
Oxford, United Kingdom.
Topic: Automatic construction of the Web application knowledge base written in XML
Worked with: Giovanni Grasso, Tim Furche, and Christian Schallhart
- 2012/13 **Teaching assistant**, *Télécom ParisTech*, Paris, France, Programming in Java.
- Mar 2012 **Visiting research student**, *Free University of Bozen-Bolzano, KRDB Research Centre*, Bolzano, Italy.
Topic: Intelligent access to information
Worked with: Prof. Enrico Franconi and Prof. Sergio Tessaris
- Feb 2011–
Jul 2011 **Research Intern**, *Free University of Bozen-Bolzano, KRDB Research Centre*,
Bolzano, Italy.
Topic: KRDB Technologies applications
Worked with: Prof. Enrico Franconi

Participation in contract research projects

- Mar 2015– to date **Software Security**, Collaborating with IBM for research related to Software Security.
- 2011– 2013 **Arcomem**, European Union FP7 project on the Archiving of the social Web. Worked on Intelligent content acquisition in Web archiving. Lead developer of the *Application-Aware Helper* component; Integration with other components.

Research interests

- Rich Internet Application (RIA) session reconstruction, modeling and testing
- Information Retrieval, Data Mining, Data Extraction and Integration
- Web Archiving
- Description Logics and logics for Knowledge Representation
- Data Structure

Awards

- Erasmus Mundus scholarship recipient for EMCL course
- Merit Scholarship recipient during the Bachelors degree

Information technology

Programming languages	C, Objective C, C++, Java, Prolog	Semantic Web	RDF, RDFS, OWL
XML technologies	DOM, DTD, XML Schema, XPath, XSLT, XQuery	Web	(X)HTML, CSS, JavaScript
DBMS	MYSQL, SQL Server, Oracle, PostgreSQL	MISC	LaTeX, Selenium, PhantomJS, Eclipse, Netbeans

Publications

International publications

Muhammad Faheem, Salman Hooshmand, Akib Mahmud, Gregor Bochmann, Guy-Vincent Jourdan, Russ Couturier and Iosif-Viorel Onut. D-ForenRIA: Distributed Reconstruction of User-Interactions for Rich Internet Applications. In *Proc. WWW*, Montreal, Canada, April 2016.

Muhammad Faheem, Sara Baghbanzadeh, Salman Hooshmand, Gregor Bochmann, Guy-Vincent Jourdan, Seyed Mir Taheri, Iosif-Viorel (Vio) Onut, ForenRIA: The Reconstruction of User-Interactions from HTTP Traces for Rich Internet Applications. In *Proc. Twelfth Annual IFIP WG 11.9 International Conference on Digital Forensics*, New Delhi, India, January 2016.

Muhammad Faheem, Pierre Senellart, Adaptive Web Crawling through Structure-Based Link Classification. In *Proc. ICADL*, Seoul, Korea, December 2015.

Muhammad Faheem, Pierre Senellart, Vassilis Plachouras, Florent Carpentier, Julien Masanès, Thomas Risse, Patrick Siehndel, Yannis Stavarakas, Arcomem Crawling Architecture. *Future Internet*, 2014.

Muhammad Faheem, Pierre Senellart, Demonstrating Intelligent Crawling and Archiving of Web Applications. In *Proc. CIKM*, San Francisco, USA, October 2013.

Muhammad Faheem, Pierre Senellart, Intelligent and Adaptive Crawling of Web Applications for Web Archiving. In *Proc. ICWE*, Aalborg, Denmark, July 2013.

Muhammad Faheem, Pierre Senellart, Intelligent crawling of Web applications for Web archiving. In *Proc. PhD Symposium of WWW*, Lyon, France, April 2012.

National publications

Muhammad Faheem, Pierre Senellart, Crawl intelligent et adaptatif d'applications Web pour l'archivage du Web. *Ingénierie des Systèmes d'Information*, 2014.

Muhammad Faheem, Pierre Senellart, Une démonstration d'un crawler intelligent pour les applications Web. In *Proc. BDA*, Nantes, France, October 2013. Demonstration. Conference without formal proceedings.

Muhammad Faheem, Pierre Senellart, Collecte intelligente et adaptative d'applications Web pour l'archivage du Web. In *Proc. BDA*, Nantes, France, October 2013. Conference without formal proceedings.

PhD thesis

- Title *Intelligent content acquisition in Web archiving*
- Supervisors Prof. Pierre Senellart
- Description The steady growth of the World Wide Web raises challenges regarding the preservation of meaningful Web data. Tools used currently by Web archivists blindly crawl and store Web pages found while crawling, disregarding the kind of Web site currently accessed (which leads to suboptimal crawling strategies) and whatever structured content is contained in Web pages (which results in page-level archives whose content is hard to exploit). We focus in this PhD work in the crawling and archiving of publicly accessible Web applications, especially those of the social Web. A Web application is any application that uses Web standards such as HTML and HTTP to publish information on the Web, accessible by Web browsers. Examples include Web forums, social networks, geolocation services, etc. We claim that the best strategy to crawl these applications is to make the Web crawler aware of the kind of application currently processed, allowing it to refine the list of URLs to process, and to annotate the archive with information about the structure of crawled content. We add adaptive characteristics to an archiving Web crawler: being able to identify when a Web page belongs to a given Web application and applying the appropriate crawling and content extraction methodology.

Master thesis

- Title *Implementation of Semantic Desktop Tool*
- Supervisors Prof. Enrico Franconi
- Description New technologies coming from different fields of research may converge for implementing new and more sophisticated office automation systems. One of these technologies allows modern operating system to store in the file system Meta information regarding content of various types of documents such as instant messages, media files, contacts and office documents. However, the organization of this Meta Information in a knowledge base, that could lead users to perform intelligent query, is a new challenge for ontology engineers. With increasing storage capacities in personal computer, searching the World Wide Web has become more efficient than searching one's own personal computer. In the thesis, we address the problem of retrieving cross data that comes from different documents within the file system using an ontology which was built through a bottom up process. Starting from the data - meta information about documents - retrieved using spotlight APIs, we build a first simple knowledge base constituted by a database in fourth normal form. We develop methodology to devise a conceptual schema from the constraints specified over the raw data. The database serves both to lead the creation of the conceptual space that is formed by the T-Box of the ontology represented by Description Logic Syntax and as data that are used to fill the assertion space which is the A-Box of the ontology.