



Dr. Mahmoud El-Haj

CONTACT INFORMATION	NLP Lecturer SCC Lancaster University Bailrigg, Lancaster, UK	<i>Mobile:</i> +44 1524-510348 <i>E-mail:</i> m.el-haj@lancaster.ac.uk <i>WWW:</i> http://www.lancaster.ac.uk/staff/elhaj/ <i>GitHub:</i> https://github.com/drelhaj <i>Google Scholar:</i> https://scholar.google.com
NATIONALITY & LANGUAGES	British (Arabic and English - bilingual native speaker)	
RESEARCH INTERESTS	Language Resources, Data Science, Natural Language Processing (NLP), Health and Medicine, Biomedical Data, Text Summarisation, Corpus and Computational Linguistics, Financial Narrative Processing and Disclosures, Big Data, Interdisciplinary Research, Machine Learning on Text Classification, Crowd-sourcing, Information Extraction.	
EDUCATION	The University of Essex , Essex, UK. Ph.D., School of Computer Science and Electronic Engineering, August 2012. <ul style="list-style-type: none">• Thesis Topic: <i>“Multi-document Arabic Text Summarisation”</i>.• Research Field: Natural Language Processing.• Advisors:<ul style="list-style-type: none">• Prof Udo Kruschwitz (Professor in Computer Science) Area of Study: Natural Language Processing and Information Retrieval.• Dr. Chris Fox (Reader.) Area of Study: Natural Language Processing. The University of Jordan , Amman, Jordan. M.Sc., Computer Information Systems, August 2008. <ul style="list-style-type: none">• Thesis Topic: <i>“Automatic Text Summarisation for Arabic Language”</i>.• Adviser: Dr. Bassam Hammo (Professor in Computer Information Systems).• Area of Study: Natural Language Processing. B.S., Computer Information Systems, June 2005. <ul style="list-style-type: none">• Project Topic: <i>Using NLP for Query Expansion</i>.• Adviser: Dr. Bassam Hammo (Professor in Computer Information Systems).• Area of Study: Natural Language Processing.	

PROFESSIONAL
EXPERIENCE

School of Computing and Communications
Lancaster University
Lancashire, UK.

NLP/Computer Science Lecturer.

September 2019 to Current

- Mainly Specialised in Multilingual Natural Language Processing and Machine Learning research.
- Expert in: a) Arabic NLP, b) Financial NLP, c) Biomedical NLP.
- General Chair of the Financial Narrative Processing (FNP) Workshop series
- General Chair of the Arabic Corpus Linguistics (WACL) Workshop Series
- Programme Chair of the Arabic Natural Language Processing Workshop Series

School of Computing and Communications
and Accounting and Finance Department
Lancaster University
Lancashire, UK.

Senior Research Associate.

October 2016 to 2019

- Working (90%) on the Datini Project on analysing Financial Narratives using NLP. A joint project with the Accounting and Finance Department at Lancaster University. Funded by ESRC
- Working (10%) on the Bio Text Mining Project with the Data Science Institute at Lancaster University. Funded by ESRC.

CASS
Lancaster University
Lancashire, UK.

Senior Research Associate.

January 2015 to October 2016

- Working on the ESRC funded Understanding Corporate Communications project.
http://cass.lancs.ac.uk/?page_id=90

School of Computing and Communications
Lancaster University
Lancashire, UK.

Research Associate.

November 2012 to December 2014

- Worked on the ESRC funded Corporate Financial Information Environment (CFIE) project. <http://ucrel.lancs.ac.uk/cfie>

UK Data Archive, Essex, UK.

Developmental Systems and Data Mining Developer.

February 2011 to November 2012.

- Upgrade the Archive's systems and preservation service.
- Worked on JISC funded project to automatically index the UK Data Archive's big data.
<http://www.data-archive.ac.uk/find/our-projects/skos-hasset?index=2>

UK Data Archive, Essex, UK.

System Migration Assistant.

July 2010 to October 2010.

- Worked on gathering, recording, organising and analysing the UK Data Archive's big data.

Computeach International Centre, Amman, Jordan.

Computer Courses Trainer.

July 2006 to January 2009.

- Computer instructor and licensed examiner.

Macro Apps, Amman, Jordan.

Software Developer.

December 2005 to July 2006.

- Java, and C# with ASP.NET developer.

Saudi Business Machines (SBM), Riyadh, Saudi Arabia.

Computer Engineer.

September 2005 to December 2005.

- Network Administrator, installation, administration service and support to hardware and peripherals.

TEACHING PLACEMENTS

Senior Teaching Assistant - 2017/2018

- I was appointed by Lancaster University to teach the SCC.361: Artificial Intelligence "machine learning and classification" module for the weeks 6-10.
- Teaching third year Computer Science students, I have contributed in designing the module and the lecture materials for 5 weeks of the module.
- I was heavily involved in setting up the final and resit exam, marking and moderation.

Programming Instructor. Code First Girls Organisation, Lancaster - 2018

- Teaching programming courses to help increase diversity in tech. The students were a mix of undergraduate and postgraduate level from different departments.

International Teaching

- Delivering online lectures on Corpus and Computational Linguistics for master level linguistics students at Prince Noura University (PNU), Saudi Arabia - 2018.
- I worked with the University to create a Computational Linguistics Module, designing the module by setting course syllabus, reading list, deliverables and final project suggestions.

Summer School Instructor. UCREL, Lancaster University - 2016-2018

- Teaching Machine Learning Automatic Classification.
<https://github.com/drelhaj/MachineLearning> to a master and PhD level students.

Teaching Assistant at Lancaster University - 2014/2015

- Corpus Linguistics, third year undergraduates in linguistics.
- Information Systems, first year undergraduates in computer science.

Teaching Assistant at The University of Essex - 2010/2011

- Digital Systems Architecture, first year undergraduates in computer science.
- Application Programming, second year undergraduates in computer science.
- Web Development, first year undergraduates in computer science.

PROJECT PROPOSALS

Project proposals I am involved in drafting:

- [1] **Project Title:** “Large Sample Evidence on the Properties and Impact of Strategic Commentary in Annual Reports. Proposal got accepted. Current funder: “ESRC”. I am now working as a full time Senior Research Associate on this Project which is called “Datini”- a collaboration between the School of Computing and the Management School at Lancaster University.
- [2] **Project Title:** “Analysing Spanish Financial Narratives”. Potential Funder: “Spanish Ministry of Science and Innovation”. The proposal was accepted and I serve as a consultant on the project and have already travelled to Madrid earlier this year to share knowledge and help the project team there to reflect what I have already developed on Analysing UK financial data on Spanish.

NON-ACADEMIC GRANTS

- During my time at Lancaster University I received a two month grant from one of Europe’s leading investment companies to develop a financial analysing tool to detect risk warning by automatically analysing the language of UK and US financial reports and conference calls. I have successfully accomplished the task and provided training and support for the company users.
- I also received a 40 from the Financial Reporting Council (FRC) to further automatically and using NLP analyse UK annual report to help reporting standardisation.

PROFESSIONAL SERVICE

- General Chair of the 2nd Financial Narrative Processing Workshop at NoDaLiDa 30 September 2019, Turku, Finland.
- General Chair of the 1st Financial Narrative Processing Workshop at LREC 7-12 May 2018, Miyazaki, Japan.
- Programme Committee for the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies will be held in New Orleans, June 1 to June 6, 2018.
- Programme Committee for the Sixth International Conference on Building and Exploring Web Based Environments (WEB 2018). May 20, 2018 to May 24, 2018 - Nice, France
- Programme Committee for the 3rd Workshop on Open-Source Arabic Corpora and Processing Tools at LREC 7-12 May 2018, Miyazaki, Japan
- Programme Committee for the Third International Conference on Arabic Computational Linguistics (ACLing 2017), Dubai, UAE
- Programme Committee for the 6th International Conference on Arabic Language Processing (ICALP 2017), Fez, Morocco
- Organiser of the Third Arabic Natural Language Processing Workshop co-located with EACL 2017, Valencia, Spain
- Programme Committee for MultiLing 2017: Summarization and summary evaluation across source types and genres co-located with EACL 2017, Valencia, Spain
- Programme Committee for the Big Data and NLP workshop hosted at IEEE Big Data 2016
- Summer School Tutor UCREL NLP Summer School 2016 Lancaster University
- Programme Committee for the Corpus Linguistics 2015 . Lancaster, UK.
- Coordinator of the 7th LSE/LUMS/MBS Conference 2013. London, UK.

- Organiser of the UCREL Corpus Research Seminars (CRS) at Lancaster University.
- Coordinator of the MultiLing Workshop at the ACL 2013 Conference in Sofia, Bulgaria.
- Coordinator of the MultiLing Pilot at the Text Analysis Conference (TAC) 2011.
- Organiser of the disciplinary Language And Computation (LAC) group at Essex University.
- Organiser of the FlatLands 2012 Workshop on NLP 2012 at Essex University, UK.

REVIEWING SERVICES

- Annual Conference of the Association for Computational Linguistics (ACL 2020).
- Language Resources and Evaluation Conference (LREC) 2020.
- The 57th Annual Meeting of the Association for Computational Linguistics (ACL) 2019.
- Language Resources and Evaluation Journal 2019.
- Computational Linguistics journal 2019.
- SAGE Open Journal 2018.
- Dyslexia International Journal 2018.
- Jordanian Journal of Computers and Information Technology (JJCIT) 2018
- 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics 2018
- Sixth International Conference on Building and Exploring Web Based Environments (WEB 2018) 2018
- 3rd Workshop on Open-Source Arabic Corpora and Processing Tools at LREC 2018
- 3rd Workshop on Open-Source Arabic Corpora and Processing Tools at LREC 2018
- 3rd International Conference on Arabic Computational Linguistics (ACLing 2017) 2017
- 6th International Conference on Arabic Language Processing (ICALP 2017), Fez, Morocco 2017
- 3rd Arabic Natural Language Processing Workshop co-located with EACL 2017
- MultiLing 2017: Summarization and summary evaluation across source types and genres co-located with EACL 2017
- Arabian Journal for Science and Engineering 2017
- International Journal of Corpus Linguistics 2017
- Big Data and NLP workshop hosted at IEEE Big Data 2016
- Digging into Data Challenge grant program (project proposal) 2016.
- Computational Linguistics journal 2016.
- ESRC Research Project Proposal (RCUK) 2015.
- International Journal of Corpus Linguistics 2015.
- Journal of Natural Language Engineering 2014, 2015.
- MDPI Future Internet Journal 2014.
- 15th International Conference on Intelligent Text Proc and Comp Ling (CICLing) 2014.
- LRE-Rel Workshop at the 8th LREC Conference 2012.
- Fourth Computer Science and Electronic Engineering Conference (CEEC) 2012.
- 2nd IEEE Conference on Computer and Communication Technology (ICCCT) 2011.
- 32nd European Conference on Information Retrieval (ECIR) 2010.

ON-LINE TOOLS AND RESOURCES

Tools, Software and Repositories

- CFIE-FRSE Web and Desktop tools, to automatically analyse and detect structure of UK Annual reports. CFIE-FRSE was presented in many conferences and invited talks World Wide and is already being used by hundred of researchers and students in the Corpus Linguistics, and Accounting and business research fields.
WEB: <https://cfie.lancaster.ac.uk:8443>.

Desktop: <https://drelhaj.github.io/CFIE-FRSE/>.

- OSMAN Arabic Readability Metric. Java open Source tool for Arabic text readability. The tool calculates readability for Arabic text with and without diacritics (Tashkeel). The tool works better with diacritics added in (we provide a method to allow you add diacritics to plain Arabic text).
<http://drelhaj.github.io/OsmanReadability/>
- Statistical Word Cloud Creator (Java)
<http://drelhaj.github.io/Java.WordCloud.LogLikelihood/>
- Machine Learning Java Suite. Java Code that trains classifiers for chairman's statements, governance and remuneration sections from 1,000 annual financial reports (Part of UCREL NLP Summer School 2017 and 2016 Lancaster University). Machine Learning Tutorial <https://github.com/drelhaj/MachineLearning/>.

Corpora and Datasets *

- Habibi: [First Arabic Lyrics Dataset](#)
- Arabic Dialects Dataset - Dataset of Arabic dialects for GULF, EGYPT, LEVANT, TONESIAN Arabic dialects in addition to MSA.
- Arabic in Business and Management Corpora (ABMC) 2016. 1200 Arabic articles as plain text and also tagged using Stanford Arabic Part of Speech Tagger
- MultiLing 2011, and 2013 corpora, I worked as a coordinator for the Arabic language to provide multidocument articles and summaries.
- KALIMAT a Multipurpose Arabic Corpus Arabic natural language resource of more than 100,000 Arabic articles processed using Arabic NLP tools.
- Essex Arabic Summaries Corpus (EASC) Arabic natural language resource for single document summarisation.

** All above corpora available for free on:*

<http://www.lancaster.ac.uk/staff/elhaj/corpora.htm>

AWARDS

- Winning team for the best audience-facing tool - BBC NewsHack event , London, 2016.
- Best Paper Award at the 4th LTC Conference, Poznan, Poland.
- Fully funded internship at the National Institute of Informatics (NII) in Tokyo, Japan.
- Winning team for the best work at the Hadoop Hackathon at Edinburgh University, UK.