

DR MAHMOUD EL-HAJ

SCC, LANCASTER UNIVERSITY, LA1 4WA, UK
elhaj.uk m.el-haj@lancaster.ac.uk



EXPERIENCE

09/2019 – CURRENT

LECTURER, LANCASTER UNIVERSITY

NLP/Computer Science Lecturer at the School of Computing. Teaching/ Research/ Supervision.
Mainly specialized in multilingual natural language processing and machine learning.

12/2012 – 09/2019

SENIOR RESEARCH ASSOCIATE, LANCASTER UNIVERSITY

Worked as an NLP and machine learning researcher, working on interdisciplinary projects between computing, linguistics, school of medicine, and accounting & finance to automatically analyze big data, extract information and build NLP resources.

07/2010 – 11/2012

DATA MINING DEVELOPER, UK DATA ARCHIVE, UNIVERSITY OF ESSEX

Worked on a JESC funded project to automatically index the UK Data archive's big data using NLP.

EDUCATION

AUGUST 2012

PHD, UNIVERSITY OF ESSEX

PhD in Computer Science, mainly working on automatic text summarization and creating NLP resources for under-resourced languages.

AUGUST 2008

MSC, UNIVERSITY OF JORDAN

MSc in Information Systems working on automatic summarization of under-resourced languages.

PROFESSIONAL SERVICES

- General Chair of the 1st, 2nd and 3rd Financial Narrative Processing Workshop at LREC'18 (Japan), NoDaLiDa'19 (Finland) and DSI'21 (UK). <http://wp.lancs.ac.uk/cfie/>
- Area Chair for EMNLP 2021 Conference, Dominican Republic.
- General Chair of the 1st Joint FNS-FNP Workshop at COLING 12 December 2020, Spain.
- Programme Committee for the 16th NAACL: HLT Workshop, 2018, USA.
- Programme Committee for the WEB 2018, Nice, France
- Programme Committee for the 3rd OSACT Workshop, 2018, Miyazaki, Japan
- Programme Committee for the 3rd ACLing 2017 Conference, Dubai, UAE.
- Programme Committee for the 6th ICALP Conference, 2017, Fez, Morocco
- Organiser & advisor of the Arabic NLP Workshop series 2017-2021
- Organiser and coordinator of MultiLing workshop series 2011-2019
- Programme Committee for the Big Data and NLP workshop hosted at IEEE Big Data 2016
- Summer School Tutor UCREL NLP Summer School 2016-2018 Lancaster University
- Programme Committee for the Corpus Linguistics 2015-2019.
- Organiser of the UCREL Corpus Research Seminars (CRS) at Lancaster University.
- 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

ACADEMIC GRANTS

Project: 'Welsh Automatic Text Summarisation'

Funder: The Welsh Government, UK. The University of Cardiff and Lancaster University.

Amount: £90,000 (USD-\$112,000).

Duration: 1 year [starting May 2021].

Summary: A collaboration led by Cardiff University, the work will contribute to the automated tools available in the Welsh language and facilitate the work of those involved in document preparation, proof-reading, and (in certain circumstances) translation. The tool will also allow professionals to quickly summarise long documents for efficient presentation. For instance, the tool will allow educators to adapt long documents for use in the classroom. It is also envisaged that the tool will benefit the wider public, who may prefer to read a summary of complex information presented on the internet or who may have difficulties reading translated versions of information on websites.

Project: 'An Assessment of Corporate Disclosures of IFRS 15: Revenue from Contracts with Customers'.

Funder: [IAAER/KPMG Research](#)

Amount: £17754 (USD-\$25,000).

Duration: 2 years [Starting April 2020].

Summary: The grant with HEC Business school in Montreal, Canada uses Information Extraction and Natural Language Processing (NLP) techniques and study disclosures of firms adopting the new accounting standard: IFRS 15. This kind of financial text mining has significant implications for holding companies to account in their reporting.

Project: 'Arabic USAS Semantic Tagger'

Funder: Research Incentive Fund grant R19068 from Zayed University Office of Research.

Amount: £5000 (USD-\$7000)

Duration: 1 year [starting June 2020].

Summary: A collaboration Zayed University and New York Abu-Dhabi University, the work provides the first Arabic Semantic Tagger, which we called AraSAS <https://arasas.herokuapp.com/>. The Project was funded by Zayed University and supervised by myself and Professor Paul Rayson at Lancaster University.

NON-ACADEMIC GRANTS

Project title: 'Detect Risk Warning through analysing UK Conference Calls'

Funder: Hedge Fund Company in London (NDA doesn't allow disclosing the name)

Amount: £5000 (USD-\$7000)

Duration: 2 months (January – March 2015)

Summary: The funded from one of Europe's leading investment companies was 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.

Project title: 'Support for the 'Analysing Narrative Aspects of UK Preliminary Earnings Announcements and Annual Reports: Tools and Insights for Researchers and Regulators'.

Funder: UK Financial Reporting Council (FRC)

Amount: £63,000 (USD-\$88707)

Duration: September 2019 to July 2020

Summary: The amount was to help in automatically analyse UK annual reports using NLP and machine learning. This to automatically detect the structure of those reports and help reporting standardisation in the UK using NLP and Machine Learning and Computational Linguistics.