

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.

JUNE 2005, AUGUST 2008

BSC, MSc, UNIVERSITY OF JORDAN

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

PROFESSIONAL SERVICES

- General Chair of the 1st, 2nd, 3rd & 4th Financial Narrative Processing Workshop at LREC'18 (Japan), NoDaLiDa'19 (Finland), DSI'21 (UK) and LREC'22 (France). <http://wp.lancs.ac.uk/cfie/>
- Programme Committee for LREC 2022, Marsielle, France
- 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

GRANTS

Project: 'FreeTxt: supporting bilingual free-text survey and questionnaire data analysis'

Funder: The Welsh Government, UK.

Amount: £80,647 (USD-\$106,000).

Duration: 1 year [starting 1 March 2022].

Summary: This project aims to bridge this gap by building the novel 'FreeTxt' toolkit which is designed to support the analysis and visualisation of multiple forms of open-ended, free-text data in both English and Welsh. FreeTxt will draw on existing open-source bilingual corpus-based utilities and methodologies, repackaging these and taking them in a new direction so that they are relevant to new audiences/user-groups. We will work closely with project partners Cadw and National Trust Wales to co-design, co-construct and test FreeTxt to ensure that the resource is fit-for-purpose and fairly and consistently meets the needs of Welsh and English-language responses.

Project: 'Welsh Automatic Text Summarisation'

Funder: The Welsh Government, UK.

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: £17,754 (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: £5,000 (USD-\$7,000)

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.

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

Funder: Hedge Fund Company in London (Company name concealed by an NDA)

Amount: £5,000 (USD-\$7,000)

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.

SELECTED PUBLICATIONS

- **El-Haj, M.**, Paul Rayson, Ismail El Maarouf, Najah-Imane Bentabet, Dominique Mariko, Estelle Labidurie, Marina Litvak, George Giannakopoulos, Ahmed AbuRa'ed, Nadhem Zmandar. "Review of the State of the Art in Financial Narrative Processing". Financial narrative processing in Spanish / coord. by Antonio Moreno Sandoval, 2021, ISBN 978-84-18802-42-3, pp. 51-98
- **El-Haj, M.**, Antonio Moreno Sandoval, Jose Antonio Jimenez Millan. "Machine Learning Models for Classifying Spanish Beaters and Non-Beaters Financial Reports". Financial narrative processing in Spanish / coord. by Antonio Moreno Sandoval, 2021, ISBN 978-84-18802-42-3, pp. 179-198
- Nadhem Zmandar, **El-Haj, M.** and Paul Rayson. "Multilingual Financial Word Embeddings for Arabic, English and French" at 2021 IEEE International Conference on Big Data (Big Data) on 15-18 December 2021.
- **El-Haj, M.**, Paul Rayson, Nadhem Zmandar. "The 3rd Financial Narrative Processing Workshop (FNP 2021)". Lancaster, UK [hybrid] on 15-16 September 2021.
- Nadhem Zmandar, Abhishek Singh, **El-Haj, M.**, Paul Rayson. "Joint abstractive and extractive method for long financial document summarization" at The 3rd Financial Narrative Processing Workshop (FNP 2021). Lancaster, UK [hybrid] on 15-16 September 2021.
- Dominique Mariko, Hanna Abi Akl, Estelle Labidurie, Stephane Durfort, Hugues de Mazancourt, **El-Haj, M.** . "Financial Document Causality Detection Shared Task (FinCausal 2021)" at The 3rd Financial Narrative Processing Workshop (FNP 2021). Lancaster, UK [hybrid] on 15-16 September 2021.
- Nadhem Zmandar, **El-Haj, M.**, Paul Rayson, Ahmed Abura'Ed, Marina Litvak, George Giannakopoulos, Nikiforos Pittaras. "The Financial Narrative Summarisation Shared Task (FNS 2021)" at The 3rd Financial Narrative Processing Workshop (FNP 2021). Lancaster, UK [hybrid] on 15-16 September 2021.
- Ismail El Maarouf, Juyeon Kang, Abderrahim Ait Azzi, Sandra Bellato, Mei Gan, **El-Haj, M.**. "The Financial Document Structure Extraction Shared Task (FinTOC2021)" at The 3rd Financial Narrative Processing Workshop (FNP 2021). Lancaster, UK [hybrid] on 15-16 September 2021.
- **El-Haj, M.**, Vasiliki Athanasakou, Sira Ferradans, Catherine Salzedo, Ans Elhag, Houda Bouamor, Marina Litvak, Paul Rayson, George Giannakopoulos, Nikiforos Pittaras. "Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation" at The 28th International Conference on Computational Linguistics (COLING'2020), Barcelona, Spain [online] on 12 December 2020.
- **El-Haj, M.**, . "Financial Narrative Summarisation". Proceedings of the Workshop MultiLing 2019/2020: Summarization Across Languages, Genres and Sources. RANLP Conference 2019/2020, Varna, Bulgaria.
- **El-Haj, M.**, Paulo Alves, Paul Rayson, Martin Walker, and Steven Young. "Retrieving, classifying and analysing narrative commentary in unstructured (glossy) annual reports published as PDF files". Published (Open Access) in Accounting and Business Research (ABR) 2020.
- **El-Haj, M.**, and Rayson, Paul and Walker, Martin and Young, Steven and Simaki, Vasiliki. "In Search of Meaning: Lessons, Resources and Next Steps for Computational Analysis of Financial Discourse". Published in the Journal of Business Finance and Accounting (JBFA) 2019.
- **El-Haj, M.**, Rayson, P., Alves, P., Herrero-Zorita, C., and Young, S.. "Multilingual Financial Narrative Processing: Analysing Annual Reports in English, Spanish and Portuguese". Book chapter at the to the World Scientific Publishers (WSPC) 2019.
- **El-Haj, M.**, Rayson, P., Alves, P., and Young, S.. "Towards a Multilingual Financial Narrative Processing System". 1st Financial Narrative Processing Workshop (FNP2018) at the 11th edition of the Language Resources and Evaluation Conference (LREC'18). May 2018. Miyazaki, Japan