Mohamed Eldesouki (PhD Student)

Curriculum Vitae (last update: June 2020)

School of Engineering and Computer Science, Concordia University
Montreal, QC, Canada
+1 (438) 526-8875 - m_d2019@encs.concordia.ca
https://mohamed.eldesouki.ca/

RESEARCH INTERESTS

Machine Learning, Deep Learning, Information Retrieval, Natural Language Processing, and Question Answering.

EXPERIENCE

Research Associate   Dec. 2015 - Sep. 2019
Qatar Computing Research Institute (QCRI), Doha, Qatar

Senior Research Software Engineer   Jan. 2015 - Dec. 2015
OMS Company, Cairo, Egypt

Senior Research Software Engineer   Aug. 2013 - Dec. 2014
Taya IT Company, Cairo, Egypt

Research Engineer   Jul. 2012 - Aug. 2013
MGD Company, Cairo, Egypt

Research Assistant   Jul. 2009 - Jun. 2012
Institute of Statistical Studies and Research (ISSR), Cairo University, Cairo, Egypt.

EDUCATION

Master (M.Sc) in Computer Science, Cairo University   Jan. 2012
(Master by Research)
Rank: Top of my class (with grade 82.2%)
Thesis Title: AN INTELLIGENT AGENT FOR ARABIC WEB INFORMATION RETRIEVAL
Supervisors: Dr. Kareem Darwish, QCRI/HBKU, Dr. Mervat Gheith, and Dr. Waleed Arafa, Cairo University, Egypt.
Specialized in Web Information Retrieval and Web Personalization systems for Arabic language.

Postgraduate Diploma in Computer Science, Cairo University   May 2006

Bachelor of Computers and Information, Cairo University   May 2003
Majoring in: Information Systems

ACHIEVEMENTS & RECOGNITION

- I have received a full scholarship from Gina Cody School of Engineering and Computer Science, Concordia University to pursue PhD in Computer Science.

- I have been awarded a Concordia University International Tuition Award of Excellence, valued at approximately $37,915.00.

- Graduated the first of my class in the Masters program with a final grade of 82%, with a recommendation to complete the PhD.

- Developed an *Arabic dialect identification system* using features extracted from only the transcripts of a speech recognition system. Using this system, I achieved the FIRST place in accuracy and THIRD place in F1 among 18 participants in the DSL Shared Task 2016 of the VarDial 2016 workshop for Arabic dialect identification [LINK].

- I obtained a scholarship to join *Wikimania Conference in Poland in 2010* [LINK] due to being an active Wikipedia member who have created and contributed in writing more than 1800 articles in both Arabic and English Wikipedia projects.

PUBLICATIONS

2020

- Kareem Darwish, Mohammed Attia, Hamdy Mubarak, Younes Samih, Ahmed Abdelali, Llus Mrquez, **Mohamed Eldesouki**, Laura Kallmeyer, (2019), *Effective Multi Dialectal Arabic POS Tagging*, Natural Language Engineering, Volume 1, Issue 1, April 2020, Pages 18, doi:10.1017/S1351324920000078. [JOURNAL]

2019

- **Mohamed Eldesouki**, Naassih Gopee, Ahmed Ali, Kareem Darwish, (2019), *FarSpeech: Arabic Natural Language Processing for Live Arabic Speech*, In proc. of INTERSPEECH 2019: Show & Tell Contribution, Sep. 2019, Pages 2372-2373, Graz, Austria. [DEMO]

- Younes Samih, Hamdy Mubarak, Ahmed Abdelali, Mohammed Attia, **Mohamed Eldesouki**, Kareem Darwish, (2019), *QC-GO Submission for MADAR Shared Task: Arabic Fine-Grained Dialect Identification*, In proceedings of the Fourth Arabic Natural Language Processing Workshop, Aug. 2019, Pages 290-294. [WORKSHOP]

- Hamdy Mubarak, Ahmed Abdelali, Kareem Darwish, **Mohamed Eldesouki**, Younes Samih, Hassan Sajjad, (2019), *A System for Diacritizing Four Varieties of Arabic*, In proceedings of the EMNLP 2019 and the 9th IJCNLP: System Demonstrations, Nov. 2019, Pages 217-222. [DEMO]

- Salvatore Romeo, Giovanni Da San Martino, Yonatan Belinkov, Alberto Barrón-Cedeño, **Mohamed Eldesouki**, Kareem Darwish, Hamdy Mubarak, James Glass, and Alessandro Moschitti, (2019), *Language Processing and Learning Models for Community Question Answering in Arabic*, Information Processing & Management, Volume 56, Issue 2, March 2019, Pages 274-290. [JOURNAL]

2018

- **Mohamed Eldesouki**, Suwon Shon, and Ahmed Ali, James Glass (2018), *QCRI-MIT Live Arabic Dialect Identification System*, QCRI - MIT CSAIL Annual Meeting Qatar National Convention Centre March 27-28, 2018. [POSTER]

- **Mohamed Eldesouki**, Suwon Shon, and Ahmed Ali, (2018), *QCRI-MIT Live Arabic Dialect Identification System*, ICASSP, Calgary, Canada. [DEMO]

- Kareem Darwish, Hamdy Mubarak, **Mohamed Eldesouki**, Ahmed Abdelali, Younes Samih, Randah Alharbi, Mohammed Attia, Walid Magdy, and Laura Kallmeyer, (2018), *Multi-Dialect Arabic POS Tagging: A CRF Approach*, In 11th edition of the Language Resources and Evaluation Conference (LREC), 7-12 May 2018, Miyazaki (Japan).

2017

- Younes Samih, **Mohamed Eldesouki**, Mohammed Attia,Kareem Darwish, Ahmed Abdelali, Hamdy Mubarak and Laura Kallmeyer, (2017), *Learning from Relatives: Unified Dialectal Arabic Segmentation*. In Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017), Vancouver, Canada, 432-441.

- Walid Magdy, and **Mohamed Eldesouky**, (2017), *ClassStrength: A Multilingual Tool for Tweets Classification*, Proc. of the 2017 IEEE/ACM International Conference on Advances in social networks analysis and mining (ASONAM), Sydney, Australia, 593-596.

- **Mohamed Eldesouki**, Younes Samih, Ahmed Abdelali, Mohammed Attia, Hamdy Mubarak, Kareem Darwish and Kallmeyer Laura, (2017), *Arabic Multi-Dialect Segmentation: bi-LSTM-CRF vs. SVM*, http://arxiv.org/abs/1708.05891.

- Younes Samih, Mohammed Attia, **Mohamed Eldesouki**, Ahmed Abdelali, Hamdy Mubarak, Laura Kallmeyer and Kareem Darwish, (2017), *A Neural Architecture for Dialectal Arabic Segmentation*. In Proc. of The 3rd Arabic Natural Language Processing Workshop (WANLP-2017) co-located with EACL 2017, Valencia, Spain, pages 46-54.

- Kareem Darwish, Hamdy Mubarak, Ahmed Abdelali and **Mohamed Eldesouki**, (2017), *Arabic POS Tagging: Don't Abandon Feature Engineering Just Yet*, In Proc. of The 3rd Arabic NLP Workshop (WANLP-2017) co-located with EACL 2017, Valencia, Spain, P. 130.

2016

- **Mohamed Eldesouki**, Fahim Dalvi, Hassan Sajjad, and Kareem Darwish, (2016), *QCRI @ DSL 2016: Spoken Arabic Dialect Identification Using Textual Features*, Proc. of the 3rd Workshop on NLP for Similar Languages, Varieties and Dialects, (VarDial 3), Osaka, Japan, P. 221.

2012

- **Mohamed I. Eldesouki**, *An Intelligent Agent for Arabic Web Information Retrieval*, (2012), Master's Thesis, Cairo University.

2011

- **Mohamed Eldesouki**, Waleed Arafa, Kareem Darwish, Mervat H. Gheith, (2011), *Representing Arabic Documents Using Controlled Vocabulary Extracted from Wikipedia*, In Proc. of The 11th Conference on Language Engineering (ESOLEC'11), Cairo, Egypt.

- **Mohamed Eldesouki**, Waleed Arafa, Kareem Darwish, Mervat Gheith, (2011), *Using Wikipedia for Retrieving Arabic Documents.* In Proceedings of Arabic Language Technology International Conference (ALTIC 2011), Alexandria, Egypt.

2009

- **Mohamed I. Eldesouki**, Waleed M. Arafa, Kareem M. Darwish,(2009), *Stemming techniques of Arabic Language: Comparative Study from the Information Retrieval Perspective*, The Egyptian Computer Journal. 36(1):30-49.

RESEARCH PROJECTS

All the details of the following projects are available in my website with some demonstrations:

Farasa Project (http://farasa.qcri.org/)   Jan. 2016 - Aug. 2019
Farasa is a fast and accurate text processing toolkit for Arabic text. Farasa can handle both Modern Standard Arabic (MSA) and the different Arabic dialects. Farasa can perform word segmentation, lemmatization, Part-Of-Speech tagging, text Diacritization, Dependency and constituency Parsing, and spell checking and correction.
My task in Farasa was to improve Arabic Dialect Segmentation, and Part-Of-Speech tagging. Furthermore, I was responsible for providing annotated data. Using a deep neural architecture of Bidirectional LSTM-CRF, I achieved an average of 93.4% and 92.8% for both segmentation and POS tagging, respectively for 4 Arabic dialects + MSA. [more details]

DialectID Project (https://dialectid.qcri.org/)   Oct. 2017 - Sep. 2018
At Qatar Computing Research Institute (QCRI), I have worked in collaboration with MIT CSAIL on the DialectID project that aims to automatic dialect identification in Arabic Broadcast Speech into five Arabic dialects namely; Modern Standard Arabic (MSA), Egyptian dialect, Levantine dialect, Moroccan dialect, and Gulf dialect.
I built DialectID as an online live identification system for the Arabic speech. The key features of the system I built are: full duplex communication based on websockets, very scalable (+100K users), can do speech segmentation, supports Kaldi's GMM and "online DNN" models, and Python, Java, Javascript clients are available.

Natasy Deep Learning Lib (https://github.com/disooqi/Natasy) Jun 2017 - Present
A deep learning library designed and developed to be both eassy to use and source code readable. It is a straightforward implementation of different algorithms and techniques of deep learning in Python. You can use it for small projects and/or educational purposes.

I am the main developer of Natasy. I implemented all algorithms from scratch. Natasy supports fully-connected NN, RNN, CNN architectures. I provided implementation for most of the well know functions and algorithms (+25) for activation, initialization, optimization, dropout, ...

Buzzdiggr project (http://www.buzzdiggr.com/)   Jan. 2015 - Dec. 2015
A real-time monitoring platform that listens to social media platforms and the web for

mentions of commercial brands then provides an array of powerful features for analysis.
I was responsible to transfer and maintain the state-of-the-art science and technologies in the field of Topic detection, NER, Sentiment analysis, and Arabic Dialects Identifications where, 1) I have provided the dataset for building Sentiment Analyzer, NER, and dialect identification, and 2) I have conducted research building both Sentiment Analyzer and NER.

Taya Arabic Processing Suite (TAPS) project   Aug. 2013 - Dec. 2014
A multi-purpose Arabic processing port which facilitates Arabic language analysis through advanced tools and techniques, TAPS is designed to adapt and adjust to all search engines. Some of the key features of TAPS are Morphology Analyzes, Named-Entity Recognition (NER), Keyboard Layout Detection/Correction, Spell Correction, Language Identification, Dynamic Document Clustering, and Text Summarization.
I worked on TAPS project for around one year and half when I was a senior research engineer at Taya IT company with Dr. Ossam Emam. I built the NER and the spellcheck error detection & correction, and auto-completion.

MGD Intelligent Tutoring System   Jul. 2012 - Aug. 2013
An Intelligent tutoring systems (ITS) that can assist human teachers identify the weakness of their students in different Math skills (addition, subtraction, multiplications, etc) and then suggests a tailored strategic path of exercises to strengthen these skills for the students.

ArabAgent Project   Jan 2009 - Apr 2013
A content-based recommendation system that is used to help the user customizing his or her access to the Web based on his interests and preferences that have been gleaned from his activities over the internet. The customization is done through filtering news articles and personalizing web search.
ArabAgent project is my Master Research Project. I worked on it during my master period and I continued working on it 2 years after I earned the degree.

TECHNOLOGY   Proficient and familiar with a vast array of programming languages, concepts and technologies, including:

Programming Languages:
Proficient in Python, C/C++, Java, Prolog and Lisp and familiar with C# .NET and PHP.

Machine learning & Scientific packages; Tensorflow, PyTorch, Keras, Scikit-learn, Octave, NumPy, SciPy, matplotlib, Jupyter notebook, CRF++, YASMET, YamCha.

NLP & IR packages; NLTK, Indri (Lemur project), Solr (Lucene), FARASA toolkit, MADAMIRA, SRILM (SRI Language Model), Kaldi.

Databases: *Relational DBs*; MySQL, MS SQL server 2012 PostgreSQL. *NoSQL DBs*; MongoDB, Neo4j, Redis.

Technologies and Tools: *Web Platforms:* Java EE, Django, and Flask
Web Frontend Technologies: HTML, XHTML, CSS, JavaScript, HTML DOM, Ajax, XML, XML DOM, Web sevices, JSON and Bootstrap framework.
Other tech.: Git, conda, virtual environment, virtual machines, Docker, Python Packaging.
OS Platforms: Linux, Windows, Mac

OTHER ACTIVITIES

- Reviewing for: ACL 2018, BJIT, RANLP 2017 http://lml.bas.bg/ranlp2017/pc.php,
- Continuing Education: Achieved several certificates through MOOC courses including machine learning, deep learning, and General AI[1].

---

[1]For the full list of certificates please visit my linkedIn profile https://www.linkedin.com/in/disooqi/ and look for Accomplishments-Certifications section

REFERENCES

- **Dr. Ahmed Ali**
  Principal Software Engineer,
  ALT group, Qatar Computing Research Institute (QCRI), Qatar Foundation
  amali@hbku.edu.qa
  Tels. office: +974-44541440, mobile: +974-33072372

- **Dr. Walid Magdy**
  Assistant Professor
  School of Informatics, University of Edinburgh
  wmagdy@inf.ed.ac.uk
  Tels. office: +44 (0) 131 651 5612

- **Dr. Alberto Barron-Cedeno**
  Senior assistant professor,
  Universit di Bologna
  a.barron@unibo.it
  Tels. office: +39 0543 374 557

- **Dr. Ossama Emam**
  IBM Master Inventor
  IBM Egypt
  emam@eg.ibm.com
  Tels. mobile: +20-1001001707