Mohamed Eldesouki

Curriculum Vitae (last update: April 2022)

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

SUMMARY

- I am a Research engineer with a senior engineering skills and a wide technical and academic background in the fields of Natural Language Processing (NLP) and Machine Learning (ML) with a broad experience covering a diverse set of software development tools, languages and methodologies, gathered throughout almost 8 years of experience.
- I have participated in designing, implementing and maintaining ML-related projects such as dialect identification for Arabic Speech, Search Engines, text processing toolkits, Content-based Recommendation Systems, Auto-completion and correction for mobile devices, and Intelligent Tutoring Systems (ITS). Most of these projects have been conducted within a funded research projects and their results have been documented and published in top-tier conferences or Scientific Journals.
- I got a strong academic foundation in NLP, Information Retrieval (IR) and Information Extraction (IE) gained through conducting research activities through my education and working within notable NLP groups such as QCRI.

EXPERIENCE

Teaching Assistant/Full-time PhD student — Sep. 2019 - Present
School of Engineering and Computer Science, Concordia University, Montreal, Canada

Software Engineer/Research Associate — Dec. 2015 - Sep. 2019
Qatar Foundation (QCRI), Doha, Qatar

Senior R&D Engineer — Jan. 2015 - Dec. 2015
OMS Company, Cairo, Egypt

Senior Research 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.
- Published 20+ research articles in top-tier conferences and Journals in the domain of Artificial Intelligence and Natural Language Processing (NLP). See [full list].
- 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.

PROJECTS

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

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.
Technology used: Python, socket programming, Tornado web framework, Kaldi toolkit, Tensorflow.

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]
Technology used: Python, Java, Django, Tensorflow, Tomcat, and FastAPI.

Natasy Deep Learning Lib (https://github.com/disooqi/Natasy) Jun 2017 - Present
A deep learning library designed and developed to be both easy 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 maintainer of Natasy open-source project. I implemented the majority of the 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, and more.
Technology used: Python, and Numpy

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 designed and implemented many parts of the aforementioned technologies 2) I have provided the dataset for building Sentiment Analyzer, NER, and dialect identification, and 3) I have conducted research building both Sentiment Analyzer and NER.
Technology used: Python, Java, Scikit-learn, Django, Spark Streaming

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.
Technology used: Python, Java, CRF++, AMIRA

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.
Technology used: C# .NET

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 used: Java and Java EE

TECHNOLOGY

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

Programming Languages:
Proficient in Python, 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:* Django, Flask, and Java EE.
Web Frontend Technologies: HTML, XHTML, CSS, JavaScript, HTML DOM, Ajax, XML, XML DOM, JSON and Bootstrap framework, ReactJS.
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,

- Teaching experience at both Concordia University, and Cairo University. At Cairo University, I had the responsibility to prepare and teach two courses, namely introduction to programming using Python and fundamentals of NLP.