



Khiati Zakaria Abdel-ilah

My research interests lie at the intersection of machine learning and natural language processing. I am particularly interested in word embedding, part of network architectures for statistical language modelling, which can transform words or phrases from a certain corpus into vectors of real numbers for language modelling purposes. Currently, I am working on cross-lingual word embedding, which consists of embedding multiple languages into a single semantic space. Nonetheless, I am not restricted to the topic mentioned above, I also enjoy working with multiple languages, or more specifically, connecting languages and transferring resources from one language to another.



zakria.ai@kaist.ac.kr

(082) 10-5521-8905

Daejeon, South Korea

Experience

Mar 2016 - Present

Graduate Assistant, Users & Information Lab (U&I Lab), KAIST

- Lab assistant
- Teaching assistant

June 2017 - August 2017

Team manager/Teacher Assistant at Data Science Expert Training Course

- Team and workshop manager
- Assist members of the team to publish their work on a workshop

May 2017 and October 2017

Workshop Instructor, Issac KAIST, KAIST, Daejeon, Korea

- Java and Web Development Workshop
- Python Programming Workshop

Mar 2015 - Jan 2016

Graduate Assistant at Semantic Web Research Center(SWRC), KAIST

- Web-page: semanticweb.kaist.ac.kr
- Lab assistant
- Implemented an Entity Linking module (for Korean and English) used in OKBQA-3 hackathon 2016
- and a pipeline for Named Entity Recognition(NER) and Named Entity Disambiguation(NED)
- Participated in building DBpedia Korea using the pipeline and module mentioned above

Sep 2012 - Jul 2014

Graduate Assistant at Intelligent Information System Lab, Korea University

- Web-page: iis.korea.ac.kr
- Lab assistant
- Participated in the national project BK21+

Jul 2012 — Aug 2012

Internship at HSBC Bank, at the IT department, Algiers, Algeria

Jan 2004 — Present

Volunteering at FOREM, a NGO that does Humane action and solidarity in Algeirs, Algeira

- Web-site: forem.dz/index.php/en/to-know-forem

EDUCATION

1

March 2015 — Present

Ph.D in computer science at Korea Advanced Institute of Science and Technology (KAIST)

2

September 2012 — July 2014

PUBLICATIONS

- "AWS-SGNS: A Simple extension to Skip-Gram Model using Dice Aligner to find an adequate window size for multi-lingual embedding", Work in progress, intended to be submitted for EMNLP 2018.
- "Agglomerative Hierarchical Clustering for Information Retrieval using Latent Semantic Index" (SocialCom, 2015)
- "Agglomerative Hierarchical Clustering Using Latent Semantic Analysis in Information Retrieval" (KIPS, 2014)
- "Collaborative Movie Recommendation Method Using Sentiment Analysis" (KIPS, 2014)
- "An Improved Method for Measurement of Gross National Happiness Using Social Network Services" (HumanCom, 2013)
- "OSGI for the management and implementation of dynamic applications" (B.S. Thesis, 2010)

Skills

Laboratory/Research

Skills

Data processing, Statistical analysis,
Programming, Database management, etc.

LANGUAGES

English (Fluent), French (Fluent), Arabic (Fluent),
Korean (Intermediate)

Computer languages

Java, Python, Matlab, etc.

Research Statement

My research interest lies in the area of natural language processing (NLP) where I have worked on statistical approaches and computational models to extract semantic information from text. I am particularly interested in closing the gap between commonly spoken languages like English or other European languages, for which there exist an abundance of NLP resources and technologies, and minority languages that often lack even the most basic NLP resources and tools such as Arabic and Korean.

During my masters, I worked at the Intelligent Information System Laboratory (IIS Lab) as graduate assistant under the supervision of Prof. Chung In-Jeong. My research area was in-between NLP and Machine Learning. In the laboratory, I was involved in several projects such as The Brain Korea 21 Plus (BK21+), a human resource development program funded by the Korean Ministry of Education. My master thesis was about retrieving snippets from search engines and embedding them using Latent Semantic Analysis (LSA), clustering the output vectors using Agglomerative Hierarchical Clustering (AHC). I published a paper by the end of my master which explains how to use LSA efficiently for clustering purposes and topic analysis.

On March 2015, I started pursuing a doctorate in Computer Science at Korea Advanced Institute of Science and Technology (KAIST) under the supervision of Prof. Key-Sun Choi. I also joined Semantic Web Research Center (SWRC) where I worked as a graduate assistant. My research topics were focused on NLP and Semantic Web where I started building an interest in tasks that extracts/enrich information from/into Knowledge bases such as DBpedia. More precisely, I worked on Entity Linking which is a task that relied on a knowledge base (KB) to identify entities from a given text and then link them to their corresponding links in the KB. Most of these entities are very ambiguous, and thus they need a strong model to make sure every named entity is correctly linked using its context. I developed a model that was able to successfully find and disambiguate named entities in Korean and was even able to accomplish results that are on par with state-of-the-art models such as AGEDISTS in English. The model was part of project financed by Hancom and was also used in several other applications such as Open Knowledge Base and Question-Answering (OKBQA) and DBpedia Korea.

On the second year of my doctorate, I decided to shift my focus to a less application oriented laboratory and thus I have joined the Users and information Laboratory (U&I Lab) under the supervision of Prof. Alice Oh. My current research topic is multi-lingual word embedding and its application to other NLP tasks such as cross-lingual transfer. Recently, I have been working on an idea that combines traditional embedding methods such as Dice Aligner and current state-of-the-art methods such as Word2Vec to come up with a better representation for multi-lingual embeddings. I am also using my work to collaborate with other researchers to work on different tasks such as topic analysis. One of the current collaborations involves clustering political tweets from three different languages (English, French, and German) and then analyze the results to identify politicians that have similar/opposite point of views.

Finally, my goal is to develop or improve new techniques that could serve as a bridge between different languages, especially the less fortunate languages that are lacking in research development. Such type of transfer can help a lot of NLP tasks ranging from machine translation to all sort of

other NLP tasks that rely on a multi-lingual environment.

REFERENCES

- Alice Oh, Associate Professor, Computer Science, KAIST, alice.oh@kaist.edu
- Chung In-Jeong, Professor, Computer Science, Korea University, chung@korea.ac.kr

