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@gmail.com
(082) 10-5521-8905
Daejeon, South Korea

Experience

June 2017 - Present

Team manager/Teacher Assistant at Data Science Expert Training Course

Mar 2016 - Present

Research Assistant at Users & Information Lab, KAIST
- Web-site: uilab.kaist.ac.kr
- Teacher Assistant of courses provided by the lab

Mar 2015 - Jan 2016

Research Assistant at Semantic Web Research Center(SWRC), KAIST
- Web-site: semanticweb.kaist.ac.kr
- Implemented the Entity Linking module (for Korean and English) used in OKBQA-3 hackathon 2016: 3.okbqa.org
- Implemented several pipelines that linked Korean and English language such as Named Entity Recognition(NER) and Named Entity Disambiguation(NED)
- Participated in several national projects such as DBpedia Korea

Sep 2012 - Jul 2014

Research Assistant at Intelligent Information System Lab, Korea University
- Web-site: iis.korea.ac.kr

Jul 2012 — Aug 2012

Internship at HSBC Bank, at the IT department, Algeirs, 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

PUBLICATIONS

- "Agglomerative Hierarchical Clustering for Information Retrieval using Latent Semantic Index" (SocialCom, 2015)
- "정보 검색에서의 잠재 의미 분석 방법을 이용한 응집 계층 군집화 기법 연구" (한국정보처리학회, 2014)
- "감정 분석을 이용한 협업적 영화 추천 방법" (한국정보처리학회, 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" (Bachelor thesis, 2010)

EDUCATION

① March 2015 — Present

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

KAIST

2

September 2012 — July 2014

Master in computer Science at Korea University

3

September 2006 — July 2010

B.S. in mathematics and computer sciences at University of Science and Technology Houari Boumediene (USTHB)

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 research/student assistant. My research area was focused on NLP tasks and Machine Learning while I prepared myself to pursue a doctorate. 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), then clustering the output vectors using Agglomerative Hierarchical Clustering (AHC). I published a paper on that by the end of my master; the general idea behind the publication was to explain how to use LSA efficiently for clustering purposes and topic analysis.

My first year of doctoral studies, I was part of the Semantic Web Research Center (SWRC) at KAIST. During that time, I completed several tasks that extracted knowledge from resourceful languages, such as English, and transfered it to a less resourceful one such as Korean or French. One of my main tasks was Entity Linking. The entity linking module is a statistical model that relied on a knowledge base (KB) to extract entities from text and identify them as a person, organization, or place, 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 entity is correctly linked using its context. I also took part in a Hackathon called Open Knowledge Base and Question-Answering (OKBQA) that has been around for the past 3 years and consists of knowledge base construction and application. My contribution was the integration of the Entity Linking to the platform, which has been used on the Question Answering task. I have also adapted a stat-of-the-art Entity Linking model called AGDESTIS on the korean language for comparison purposes.

In the second year of my doctorate, I shifted my focus to a less application oriented laboratory and thus I have joined the Users & information Laboratory (U&I Lab) . My current research topic is about multi-lingual word embedding, where we embed several languages into a single space and from that space we could retrieve relatedness between two tokens. Using this method I have been able 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, German) and then analyse the results to find politicians that have similar point of views.

Finally, my goal is to work on new languages and develop/improve new techniques that could serve as a bridge between researches on different languages from machine translation to all sort of other NLP tasks that rely on multilingual 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