On the 2nd International Conference on Arabic Language Resources and Tools
22-23 April 2009, Cairo, Egypt

Dina A. Said

PhD Candidate and Teaching Assistant

Dept. of Computer Science, University of Calgary, Calgary, AB.

dasaid@ucalgary.ca

1   Introduction

The purpose of this document is to provide information about the 2nd International Conference on Arabic Language Resources and Tools, held in Cairo, Egypt, 22nd-23rd April 2009. The conference was organized by MEDAR project (Mediterranean Arabic Language and Speech Technology).

The aim of the conference is to provide an opportunity for researchers interested in Arabic language resources to meet, explorer the state-of-art in the field, discuss problems and challenges, and finally agree to methods and projects to cooperate and unitify their efforts.

In the following, I will discuss the tutorials held in conjunction with the conference (Section 2). This is followed by providing an overview about the main presentations and posters in section 3. Finally, I'd mention the benefits I personally gained by attending the conference in section 4.

2   Tutorials

Three workshops have been held in coperation with The Faculty of Computers and Information, Cairo University (FCI-CU) The topics of these tutorials are Introduction to Arabic NLP, Arabic Dialect Processing, and Machine Translation: challenges and trends. The choice of these topics was very clever since they range from the necessary background knowledge (Arabic NLP tutorial), new direction in the Arabic research (Arabic Dialect tutorial), and a tutorial that combines between presenting basic knowledge for researchers and addressing possible directions for research (Machine Translation tutorial).

In the tutorial of the Arabic NLP, Dr. Habash has provided an overview about the characteristics of the Arabic Language including letters forms, letter marks (e.g. dots), letter shapes, diacritics, and writing styles. Additionally, he explained briefly the process of morphological analyzing in the Arabic language and addressed some problems that affect the research in the Arabic language such as ambiguity. I found this tutorial to be of a great importance especially for Arabic non-native speakers or researchers who would like to work in the Arabic language and don't have enough background knowledge.

In the second tutorial, Dr. Diab addressed a recent direction in the field of the Arabic language which is the Arabic Dialect (AD). She mentioned the challenges involved in such direction such as the lack of written standards to AD. Additionally, each Arabic county, actually each Arabic city, has its own AD. Sometimes, a word in the AD of a certain country has another meaning in the AD of another country. Dr. Diab emphases the important of studying the AD by mentioning the fact that there is no native speaker to the classical Arabic since Arabs speak only the AD of their country or region. The study of AD would not benefit only the research in Arabic phonetics but also several Internet application since people tend to use their AD in writing in blogs or forums. There has been a debate among researchers attending the tutorial and the conference about the importance of using only the classical Arabic language. Some researchers argue that the research in AD will encourage people to use AD instead of the classical Arabic. However, many others, including myself, think that we are Engineers not politicians . Therefore, we deal with the existing situations and provide solutions to them. If people talk with AD and write with AD, we should provide methods and tools to deal with this fact.

Prof. Way and Dr. Hassan discussed in the third tutorial the problem of Machine translation and the challenges involved in it. They overviewed the state-of-art in Machine Translation. Additionally, they explained the need of parallel corpus to support the translation process.

3   Technical Papers

The technical papers accepted to be presented in the conference were ranged among wide varieties of areas including:

- Large Projects e.g. MEDAR project, project ongoing in the higher institute for applied science & technology (HIAST) in Syria, and projects ongoing in King Abd El-Aziz University in Saudi Arabia. The session was very useful since it provides an overview on these projects and the work achieved there in the area of Arabic research. It seems that a lot of cooperation can be made among these parties to bring their effort together.

- Text Mining including Text Categorization, summarization, Machine Translation and keyword extraction.

- Morphology, Tagging and Diacritization.

In general, the technical presentations and posters were very beneficial . Researchers in the area of Arabic language should have gained knowledge about the state-of-art. Additionally, it was an opportunity for researchers to meet each other and discuss common problems. One of the biggest problem is the problem of lack of standards for Arabic Morphological analyzing. Most people working in the Arabic language have to reinvent the wheel and make their own algorithms for stemming, root extractor, etc... Furthermore, there is a lack of benchmark Arabic datasets for applications such as text categorization, translation and summarization. Most researchers have to collect their own datasets. Unfortunately, even these collected datasets are not available for other researchers to use. There was a suggestion for MEDAR to provide the software programs and

datasets used in the technical papers on the conference website. The conference organizers promised to consider of this issue in future conferences.

4   Personal Benefits

The importance of this conference for myself is shaded from the fact that it is one of the few conferences that target the research in the IT field concerning the Arabic language. It was a great opportunity to meet most researchers that are working in the Arabic language all over the world, exchange ideas, and come with new directions. In addition to the main conference, I have attended the three workshops that had been held in conjunction with the conference. Attending these workshops enriched my background about Arabic Natural language processing and Machine Translation. Being in the early stages of my PhD, this conference helped me to have an overview about the state-of-art in the field. Additionally, this assisted me maintaining the necessary network with researchers in that field. Last but not least, I was able to present my paper and got my work published in the conference proceedings.

I'd like to thank the MEDAR community for giving me the opportunity to attend the conference by accepting my application of the travel grant. Thanks should go specifically to Dorte Haltrup Hensen, Mohammed Attia and Helene Mazo for their incredible help. They took care of every single detail and made every effort so that everyone feels comfortable and satisfied.