This short report is intended to provide a brief description of my experience at MEDAR's conference. The report starts by an overview of the basic motivation of Human Language Technology followed by interesting facts about Arabic language and some significant observations. Next, an overview of the event and its organizer's is given. Then, my research interests and paper presented are summarized. After that, my personal conference's experience is covered highlighting most of interest to me. Finally, it finishes by acknowledgments to key individuals.

Shabib AlGahtani ✉ *algahtani@cs.man.ac.uk*

A large fraction of the rapidly growing information available to us is unstructured in the form of text, images, video and audio. The web content, as a knowledge repository, is doubling each 15 months with 80% of its content stored in natural language textual format. Human Language Technology (HLT) is concerned with finding practical techniques to help in processing and producing multi-lingual natural language text and speech, thus making it easier for people to interact with machines.

Arabic, one of the web languages, is the mother tongue of more than 250 million in the Arab states and the religious language of more than 1.5 billion Muslims all over the world. It is ranked 6th by United Nations in terms of its importance. According to www.internetworldstats.com, the world's web users' growth between 2000 and 2008 was 342%. Impressively, the Middle East, excluding African states, has scored the highest rate of web users' growth in same period, about 1300% increase. Expecting that most of them are Arabic speakers, the web Arabic content will have to increase dramatically which require a serious effort to address the need for an Arabic HLT tools and resources.

As a response to the high demand of Arabic HLT solutions, European Commission has supported the consortium of Mediterranean Arabic Language and Speech Technology (MEDAR)'s project, one of few events specialized in Arabic language. That initiative aims at establishing a wide cooperation between European Union and Mediterranean Arabic speaking countries bridging the gap between academia and industry to promote Arabic HLT. Toward its aim's fulfillment, MEDAR consortium organized The 2nd International Conference on Arabic Language Resources and Tools in Cairo, held on 22nd-23rd April, 2009. Its objectives mainly includes highlighting

challenges, tools and resources survey and establishing network between major players in the field of Arabic HLT.

About me, I am a second year PhD student at the school of computer science, the University of Manchester, United Kingdom. The university operates the National Center of Text Mining (NACTEM) where I'm conducting my research on Arabic Information Extraction under the co-supervision of Mr. John McNaught and Mr. William Black. My specific topic is on Named Entity Recognition and Co-reference resolution in Modern Standard Arabic adopting corpus-based methods. Those two tasks involve the development of a part of speech tagger as preprocessing step which my publication in the conference was about.

My experience in the conference started a day earlier in its organized tutorials in Cairo University where it had been a great opportunity for me to listen to professional invited speakers who have been contributing decent work to the field of Arabic language processing, Dr. Habash and Dr. Diab from Columbia University. I was glad to learn the main characteristics of Arabic language that need to be considered for any processing attempt. Dr. Habash clearly covered Arabic ambiguity linking each Arabic feature to the complexity it adds and the analysis level it affects. Dr. Mona talked about their CADIM project concerned with Arabic Dialect processing which I have long thought it has not been touched yet. Given the many aspects of Arabic, topics covered during those tutorials were invaluable for me; enriching my background on Arabic language processing and helping me to properly understand and appreciate people's work presented in the conference the next two days.

Later, in the two days conference, I had the chance to present my paper about Arabic Part of Speech Tagging Using Transformation-Based Learning, when I had the chance to explain my approach and discuss the results with interested researchers. Also, I had an initial agreement with Mr. Hamdy Mobarak, from Sakhr Software working also on tagging, to work on maintaining a standard test data that could be used in order to evaluate and compare both my tagger and his. That attempt is to surrogate the lack of standard dataset that would hopefully be considered by MEDAR. Meeting people from the LDC was another great opportunity gained from attending the conference. I had a discussion with Dr. Christopher and Dr. Moammori regarding the Arabic TreeBank used in my experiments, particularly on results of my experiments; larger training data unexpectedly giving a lower accuracy. That was due to tagset modifications and tagging inconsistency. As understood from them, the TreeBank is being revised and in the way for reproducing the revised version soon. For the sake of the next phase of my PhD work on Arabic

NER, I was fortune to meet Dr. Al-Kharashi, from King Abdulaziz City for Science and Technology, who presented his work on person names generation and triggers. The list that he maintained would certainly be a rich resource for my proposed named entity recognizer. Given that I am adopting a corpus-based method in NER study, another indispensable resource that I was lucky to know about is the multi-lingual named entity corpus for Arabic, English and French, being built by ELDA and presented in the poster session.

During presentations attended, I had the chance to know more about the state-of-the-art in the field, learn about standing challenges, maintain a network of interested professionals in my field, present my first ever paper published in the conference proceedings and to know the major players. Most importantly, I had a better view on the bright future of the Arabic HLT by looking at tremendous effort from both organizers and participants.

Finally, I would like to acknowledge MEDAR community for their utmost support and considering me for the travel grant. Special thanks to Dorte Haltrup Hansen and Helene Mazo for their kind assistance before and during the conference which helped me to make the most out of this great experience.