

Introduction

This volume contains a description of the CoNLL-2012 Shared Task and the participating systems. The CoNLL-2012 shared task was on modeling multilingual unrestricted coreference in the OntoNotes data. This was an extension of the CoNLL-2011 shared task and involved automatic anaphoric mention detection and coreference resolution across three languages – English, Chinese and Arabic – using the OntoNotes 5.0 corpus, given predicted information on the syntax, proposition, word sense and named entity layers as input. The goal was to identify anaphoric mentions – both entities and events – and perform coreference resolution to create clusters of mentions representing the same entity or event in the text. The English and Chinese language portion of the OntoNotes data comprises roughly one million words per language from newswire, magazine articles, broadcast news, broadcast conversations, web data and conversational speech. The English corpus also contains a further 200k of the English translation of the New Testament. The Arabic portion is smaller, comprising 300k of newswire articles. One of the challenges for the shared task participants (though they were limited by the time constraints of the task) and also for continuing research going forward was to find effective ways to bring these multiple layers of information to bear on the coreference task to improve upon the current state of the art. An additional challenge for participants of this year’s shared task was to develop systems that perform well across languages. We were happy to see many competitive systems in both English and Chinese. The results for Arabic are encouraging as well, in spite of the smaller data set.

As is traditional with CoNLL, we had two tracks – an open and a closed track. Since world knowledge is an important factor in coreference resolution, even in the closed task participants were allowed to use some limited, outside sources, including WordNet and a pre-computed table predicting number and gender information for noun phrases for the English task. This information is not available for Chinese and for Arabic due to lack of similar resources. For the open task, as usual, participants were allowed to use any other source of information, such as Wikipedia, gazetteers, etc., that did not violate the evaluation criteria designed to protect the test set. A total of 17 participants submitted system outputs and one participant withdrew because they found a bug in their system. Among the remaining 16 participants, 15 submitted system description papers. All 16 systems participated in the English task, 15 systems participated in the Chinese task and 8 systems participated in the Arabic task. There were 15 entries in the closed track and 3 in the open track. We hope that the data set of this year’s shared task will provide a useful benchmark and spur further research in this important sub-field of language processing.

Sameer Pradhan, Alessandro Moschitti and Nianwen Xue
Organizers of the CoNLL-2012 Shared Task