

Introduction to Text Summarization and Other Information Access Technologies

Horacio Saggion

Natural Language Processing Group
University of Sheffield
211 Portobello Street - Sheffield - S1 4DP
England - United Kingdom
saggion@dcs.shef.ac.uk

Abstract

In recent years we have witnessed an explosion of on-line unstructured information in multiple languages, making natural language processing technologies such as automatic text summarization increasingly important for the information society. Text Summarization provides users with condensed descriptions of documents, allowing them to make informed decisions based on text summaries. Text summarization can be combined with Information Retrieval (IR) and Question Answering (QA) to provide users with focus-based or query-based summaries which are targeted towards the users' specific needs. When the information a user looks for is spread across multiple sources, text summarization can be used to condense facts and present a non-redundant account of the most relevant facts found across a set of documents.

The objective of this IJCNLP 2008 tutorial is to give an overview of a number of technologies in natural language processing for information access including: single and multi-document summarization, cross-lingual summarization; and summarization in the context of question answering.

The tutorial will discuss summarization concepts and techniques as well as its relation and relevance to other technologies such as information retrieval and question answering. It will also include description of available resources for development, training and evaluation of summarization components. A summarization (multi-document and multilingual) toolkit will be used for demonstration purposes. A number of ques-

tion answering components relevant for the creation of definitional summaries and profiles will also be demonstrated.

Biography

Dr. Saggion is a research fellow in the Natural Language Processing group, Department of Computer Science, University of Sheffield, England, UK. His area of expertise is Text Summarization. He works in national and international projects on information extraction, ontology-based information extraction, question answering, and text summarization. He obtained his PhD. in 2000 from Université de Montréal, Département d'Informatique et de Recherche Opérationnelle. He has published over 50 works in conferences, workshops and journal papers. Together with his research career, he has been an active teacher, he was assistant professor and researcher at Universidad de Buenos Aires, and Universidad Nacional de Quilmes, and teaching assistant at Université de Montréal. He has participated in a number of summarization and question answering evaluations including DUC 2004, DUC 2005, MSE 2005, TREC/QA 2003, TREC/QA 2004, TREC/QA 2005. He has recently organised the workshops "Multi-source Multi-lingual Information Extraction and Summarization" and "Crossing Barriers in Text Summarization Research" in the RANLP Conferences. He has given courses and tutorials on Text Summarization and other technologies such as question answering in a number of international venues such as ESSLLI and LREC.