

- Motivation and Objectives:

The ever-increasing availability of on-line textual databases and the development of the Internet have made of Text Summarization a major research topic. Two major efforts show the increasing interest of the community for text summarization and its evaluation:

(1) Since the year 2000, the National Institute of Standards and Technology (NIST) with support from the Defence Advanced Research Projects Agency (DARPA) is conducting a series of evaluations in the area of text summarization, the Document Understanding Conferences (DUC), providing the appropriate framework for system-independent evaluation of text summarization systems.

(2) The Translingual Information Detection, Extraction and Summarization (TIDES) program is developing advanced language processing technology to enable English speakers to find and interpret information in multiple languages (<http://www.darpa.mil/ipto/programs/tides>). One way of accessing information in a multilingual environment is through the use of cross-lingual summaries, the focus of the recent Multilingual Summarization Evaluation (MSE).

The focus of this LREC 2006 tutorial will be on available resources which may help researchers in developing summarization applications and assess them in controlled experiments. The tutorial will give an overview of summarization concepts and techniques, to then concentrate on available resources for development, training and evaluation.

- Detailed content:

- Summarization concepts: summary typology and examples; human factors in the production of summaries; human production and use.
- Summarization techniques and systems: theoretical framework; superficial features: indicative phrases, term distribution, title, position, etc.; machine learning for sentence extraction; summarization by information extraction; text generation; language models; new techniques such as paraphrase identification and generation.
- Summarization resources: the SUMMAC corpus; the SummBank corpus for cross-lingual summarization; the DUC corpus; the MEAD toolkit; available NLP tools to support summarization including a tool compatible with GATE.
- Evaluation exercises: the SUMMAC evaluation; the Document Understanding Conference (DUC) evaluation and the 2005-2007 Roadmap on Text Summarization; the MSE 2005 evaluation; the Text Summarization Challenge; relation to TREC/QA evaluation.
- Evaluation methods: intrinsic and extrinsic methods; precision and recall; content-based metrics, the ROUGE package and its extension to Basic Elements, the Pyramids evaluation.

- Prerequisites:

Basic knowledge of natural language processing.

- Intended public: students/professionals in: Computer Science, Information Science, Computational Linguistics, and Linguistics.

- Lecturer:

Dr. Horacio Saggion

Department of Computer Science
University of Sheffield
Regent Court
211 Portobello Street - Sheffield - S1 4DP
England - United Kingdom
Tel: +44-114-222-1947
Fax: +44-114-222-1810
saggion@dcs.shef.ac.uk

- About the lecturer: 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 on information extraction, question answering, and text summarization. He obtained his PhD. in 2000 from Université de Montréal, Département d'Informatique et de Recherche Operationnelle; his Master degree from Universidade Estadual de Campinas (UNICAMP), Departamento de Computação in 1995; and his undergraduate degree of "Licenciado" in 1988 from Universidad de Buenos Aires, Departamento de Computación. He has published over 40 works in conferences, workshops and journal papers as well as written two book chapters. He has given talks in several conferences and workshops. Together with his research career, he has been an active teacher, he was assistant professor and researcher at Universidad de Buenos Aires (1986-1997), Universidad Nacional de Quilmes (1995-1996), and Université de Montréal (fall 1998 and fall 1999), and invited researcher at Johns Hopkins University and Université de Paris VII. He has been a member of several scientific program committees in natural language processing and artificial intelligence. He has received awards from several institutions including Ministerio de Educación de la Nación (Argentina), Fundación Antorchas (Argentina), and Université de Montréal (Canada). Dr. Saggion has participated in a number of summarization evaluations including DUC 2004, DUC 2005, and MSE 2005. He has recently organised the workshop "Crossing Barriers in Text Summarization Research" in the RANLP 2005 Conference and acted as editor of the Workshop Proceedings.