

Preface

TWLT is an acronym of Twente Workshop(s) on Language Technology. These workshops on natural language theory and technology are organised by the Parlevink Project, a language theory and technology project of the Department of Computer Science of the University of Twente, Enschede, The Netherlands. For each workshop proceedings are published containing the papers that were presented.

TWLT 14, has been organised together with the German Research Center for Artificial Intelligence, DFKI Saarbrücken, Germany. The idea for this workshop grew out of a longstanding cooperation between the University of Twente, TNO-TPD in Delft and DFKI. This co-operation manifested itself for the first time in the Twenty-One project, which inspired a whole series of other projects, such as Pop-Eye and Olive, but which also led to a close contact and exchange with independently established projects such as Mulinex and MIETTA for which DFKI was responsible. All of these projects had in common that they were funded by the Telematics Application Programme of the European Commission, all, except for Twenty-One, by the Language Engineering Sector.

Beyond this formal feature, the projects mentioned also had in common that part of their agenda was and is to explore and prove the use and usefulness of language technology in the area of information retrieval. The application domains of these projects are quite different. Twenty-One concentrated on the environmental domain of sustainable development, Pop-Eye and Olive are situated in the area of video retrieval on the basis of linguistic material, Mulinex is a rather general purpose search engine, and MIETTA belongs to the domain of travel and tourism. What distinguishes these domains is clearly the type of media involved: Twenty-One's document base consists mostly of printed papers which had been scanned in and processed. Pop-Eye and Olive deal almost exclusively with video material, where Olive brings in the additional feature of speech processing. Mulinex is concerned with retrieval in the world wide web, and MIETTA combines the search in structured databases with searching the web for all kinds of information. However, there is again one common feature which plays quite a crucial role in all of these projects viz. multi-linguality. As the individual projects all take somewhat different tacks on this issue, one could almost claim that between them they cover the whole range of possible approaches to multi-lingual information retrieval.

Our objective when organising this workshop was to bring together researchers from these projects at a more formal conference to present their different approaches. However we also wanted to initiate a more lively exchange with other prominent scientists in the field. The work they are representing here shows some intriguing overlap with our own work, but in addition some contributions address the issues we feel should play a more prominent role in the next generation of projects that given the success of the previous co-operation we are likely to start in the future. In particular we foresee that the need for more personalised information delivery will urge for advanced tools such as information extraction and filtering, concept-based retrieval, domain modeling, image processing, etc. to be integrated in what we established thusfar. With funding from the Telematica Instituut (initiative of the Dutch government and a number of industrial organisations) TNO and the University of Twente recently started the DRUID project for the exploration of some of these themes and for setting up new collaborative links. We expect that TWLT14 will contribute to the aims of this initiative, and more generally to a better understanding of the potential and needs of this particular research area.

A workshop is the concerted action of many people. We are grateful to the authors and the organisations they represent, for their efforts and contributions. In addition we would like to mention here the people whose efforts have been less visible during the workshop proper, but whose contribution was evidently of crucial importance. Alice Hoogvliet took care of the administrative tasks (registration, hotel reservations, etc.), Thijs Westerveld supported us in disseminating the workshop announcements and Michiel Visser helped with the lay-out of the proceedings.

Previous TWLT workshops

Previous TWLT workshops were

- TWLT1, *Tomita's Algorithm: Extensions and Applications*. 22 March, 1991.
- TWLT2, *Linguistic Engineering: Tools and Products*. 20 November, 1991.
- TWLT3, *Connectionism and Natural Language Processing*. 12 and 13 May 1992.
- TWLT4, *Pragmatics in Language Technology*. 23 September, 1992.
- TWLT5, *Natural Language Interfaces*. 3 and 4 June, 1993.
- TWLT6, *Natural Language Parsing*. 16 and 17 December, 1993.
- TWLT7, *Computer-Assisted Language Learning*. 16 and 17 June 1994.
- TWLT8, *Speech and Language Engineering*. 1 and 2 December 1994.
- TWLT9, *Corpus-based Approaches to Dialogue Modelling*. 9 June, 1995
- TWLT10, *Algebraic Methods in Language Processing*. 6-8 December, 1995
- TWLT11, *Dialogue Management in Natural Language Systems*. 19-21 June, 1996
- TWLT12, *Automatic Interpretation and Generation of Verbal Humor*. 11-14 Sept. 1996
- TWLT13, *Formal Semantics and Pragmatics of Dialogue, Twendial'98*. 13-15 May 1998

For the contents of the previous proceedings, please consult the last pages of this volume.