

Summarisation through Discourse Structure

Dan Cristea^{1,2}, Oana Postolache^{1,3}, and Ionuț Pistol¹

¹ Al. I. Cuza University, Iași, Romania
cristea@infoiasi.ro, ipistol@infoiasi.ro

² Institute for Theoretical Computer Science,
Natural Language Processing Group, Iași, Romania

³ Computational Linguistics, Saarland University, Saarbrücken, Germany
oana@coli.uni-sb.de

Abstract. In this paper we describe a method to obtain summaries focussed on specific characters of a free text. Summaries are extracted from discourse structures which differ from RST structures by the fact that the trees are binary and lack relation names. The discourse tree structures are obtained by combining constraints given by cue-phrases (resembling Marcu’s method) with constraints coming from the exploitation of cohesion and coherence properties of the discourse (as proved by Veins Theory). The architecture of a summarisation system is presented on which evaluations intended to evidence the contribution of each module in the final result are performed and discussed.

1 Introduction

In this paper we describe an approach to discourse parsing and summarisation that exploits cohesion and coherence properties of texts. We built discourse structures that resemble the RST (Rhetorical Structure Theory [1]) trees, although ours are binary and lack relation names. Discourse tree building resembles the cue-phrase centred approach of Marcu [2] but adds to it constraints coming from the exploitation of the relation that is proved to exist by Veins Theory (VT) [3] between discourse structure and reference chains (a manifestation of cohesion), on the one hand, and between the global discourse structure and the smoothness of centering transitions (a manifestation of coherence) [4], on the other. The output of the parsing process is used to obtain excerpt-type summaries focussed on individual characters mentioned in the text. A combined, pipe-line/parallel/incremental, type of processing is employed.

The involved modules are POS-tagging, FDG-parsing, clause segmentation of sentences in clauses, construction of elementary discourse trees, detection of noun phrases (NPs), anaphora resolution (AR), discourse parsing and summarisation. To master the combinatorial explosion yield by different sources of ambiguity, a beam-search processing is employed. We present the architecture of a discourse parsing system and discuss the evaluation methodology. The final evaluation is realised by comparing the summaries output by the system against those contributed by human subjects.