Methods and applications in discourse parsing

Proposer(s) / Proposatzailea(k): names / izenak

Mikel Iruskieta

Contact / Kontaktua: email

mikel.iruskieta@ehu.eus

Description / Deskribapena

Discourse Parsing is nowadays a very useful task. Several discourse parsers have been developed for many languages. Some of them are multilingual (Braud et al. 2017 and Iruskieta and Braud 2019).
Firstly, the student will review the state of the art.

Secondly, the student could develop or improve an automatic discourse parser or partial parser and evaluated intrinsically.

And, finally, the parser could be tested extrinsically in a sound NLP task such as sentiment analysis, summarization, question answering.

Keywords: coherence, machine learning, parsing, question answering, sentiment analysis

Goals / Helburuak

To develop an automatic discourse parser under Rhetorical Structure Theory (RST) formalism and explore possible applications.

Requirements / Betebeharrak

Profile: Computer scientists

Framework / Esparrua

Framework here if needed

Tasks and plan / Atazak eta plana

1. Study the state of the art of discourse parsing
2. Design of an automatic discourse parser, based on Rhetorical Structure Theory.
3. Test the parser in several corpora and evaluate it.
4. Explore possible applications of the parser in different NLP tasks: question answering, sentiment analysis, summarization, machine translation...

References:

Bengoetxea, K. Iruskieta, M. (2018) A Supervised Central Unit Detector for Spanish Procesamiento del Lenguaje Natural 60: 29--36. ISSN 1135-5948. DOI 10.26342/2018-60-3

Bosma, Wauter E. 2008. Discourse oriented summarization. Enschede: University of TwenteThesis.

Cheng Li. 2010. Automatic text summarization based on Rhetorical Structure Theory. Proceedings of International Conference on Computer Application and System Modeling (ICCASM 2010). (pp. V13-595-598). Taiyuan, China.

Braud, C., Coavoux, M., & Søgaard, A. (2017). Cross-lingual RST discourse parsing. *arXiv preprint arXiv:1701.02946*.

Feng, V. W., & Hirst, G. (2012, July). Text-level discourse parsing with rich linguistic features. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1 (pp. 60-68). Association for Computational Linguistics.

Iruskieta, M. Braud, C. (2019). EusDisParser: improving an under-resourced discourse parser with cross-lingual data. Proceedings of Discourse Relation Parsing and Treebanking (DISRPT2019), pages 62–71. Minneapolis, MN, June 6, 2019. ACL.

Iruskieta, M. (2016) Corpus exploration of discourse relations in RST TextLink's 1st Training School: Methods and tools for the analysis of discourse relational devices. Valencia, Spain. January 18-22

Joty, S., Carenini, G., & Ng, R. T. (2015). Codra: A novel discriminative framework for rhetorical analysis. Computational Linguistics, 41(3), 385-435.

Li, Jiwei & Daniel Jurafsky. 2016. Neural net models for open-domain discourse coherence, ArXiv:1606.01545v

Maziero, E., Pardo, T., da Cunha, I., Torres-Moreno, J. M., & SanJuan, E. (2011, October). DiZer 2.0-an adaptable on-line discourse parser. In Proceedings of 3rd RST Brazilian Meeting (pp. 1-17).

Sarda, A.T. & A.R. Kulkarni. 2015. Text summarization using neural networks and Rhetorical Structure Theory. International Journal of Advanced Research in Computer and Communication Engineering 4(6). 49-52.Taboada, Maite & William C. Mann. 2006. Applications of Rhetorical Structure Theory. Discourse Studies 8(4). 567-588.

Taboada, Maite & William C. Mann. 2006. Rhetorical Structure Theory: Looking back and moving ahead. Discourse Studies 8(3). 423-459.

Wu, Changxing, Xiaodong Shi, Jinsong Su, Yidong Chen & Yanzhou Huang. 2017. Cotraining for implicit discourse relation recognition based on manual and distributed features. Neural Processing Letters. 1-18

Xue, Nianwen, Hwee Tou Ng, Sameer Pradhan, Attapol T. Rutherford, Bonnie Webber, Chuan Wang & Hongmin Wang. 2016. CoNLL 2016 Shared Task on multilingual shallow discourse parsing.Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. (pp. 1-19).

Zhao, Zhenyu, Guozheng Rao & Zhiyong Feng. 2017. DFDS: A Domain-Independent Framework for Document-Level Sentiment Analysis Based on RST. In Lei Chen, Christian S. Jensen, Cyrus Shahabi, Xiaochun Yang & Xiang Lian (eds.), Web and Big Data: First International Joint Conference, APWeb-WAIM 2017, Beijing, China, July 7–9, 2017, Proceedings, Part I (pp. 297-310). Cham: Springer International Publishing.