

Verbatim, that started in 2008 with Luis Sarmiento and Sérgio Nunes (respectively supervised by Eugénio Oliveira and Cristina Ribeiro) explored collections of electronic news and lead to the identification automatic extraction of personality names and citations. It is now being improved by Jorge Teixeira.

Verbatim was the seed for the final SAPO product called VOXX that makes available personalities and respective citations who are currently present on the media.

Verbetes, to which Luis Sarmiento and Jorge Teixeira, supervised by Eugénio Oliveira, gave a major contribution has been launched in 2010. Verbetes, comes as a follow up of Verbatim and, besides personality names and citations is able to identify personality professions and positions, the temporal window in which those references have been recognized and the frequency with which the personalities were mentioned. Through the use of the learning algorithm “Conditional Random Fields” the results have been significantly improved.

In order to improve the information visualization, a tool called “O Mundo Visto Daqui” (“Interactive World as seen from here”) interactively shows automatically identified relationships among different personalities currently cited in the news. Also, the most frequent a co-occurrence exists the stronger is the respective nodes interconnection. LIACC (Jorge Teixeira, Nuno Baldaia) was present in this project with SAPO Labs (Eduarda Rodrigues) and other groups under the supervision of Mário Silva who leaded the REACTION project funded by Portuguese funding Agency, FCT, under a UTAustin-Portugal research program.

Finally, “Time Machine “, ([“A Máquina do Tempo”](#)) inherits contributions from the previously mentioned projects. It creates a network of personalities including their profiles and identifies more closed connected sub-groups. The information extracted from the repositories becomes visible on the web.

References

Jorge Teixeira, Luis Sarmiento, Eugénio Oliveira (2011) “A Bootstrapping Approach to Training a NER with Conditional Random Fields” in Progress in Artificial Intelligence, Lecture Notes in Artificial Intelligence V.7026, pp.664-678.

Luis Sarmiento, Sérgio Nunes, Jorge Teixeira, Eugénio Oliveira (2009) “Propagating Fine-Grained Topic Labels in News Snippets”. In Proceedings of the WI-IAT 2009 Workshop on Intelligent Analysis and Processing of Web News Content (IAPWNC’09), Milan, Italy, 2009.

Luis Sarmiento, Sérgio Nunes, “Automatic Extraction of Quotes and Topics from News Feeds” In Proceedings of the 4th Doctoral Symposium on Informatics Engineering (DSIE’09), Porto, 2009.

Luís Sarmiento, [Definition and Computation of Similarity Operations between Web-specific Lexical Items](#), PhD Thesis, PRODEI Program, FEUP, 2010 (supervisor: E.Oliveira).