# Newspaper Analysis

Aleksander Dietrichson, Pablo Pagnone

**Abstract**

*Keywords:* text analysis, newspapers, R, shiny, lemmatization

## 1. Objective

Our team is currently working on a suite of products to provide media analytics to e.g. political campaigns. As part of this project we perform some analyses of traditional media including online newspapers to obtain differents metrics (theme of day, relation candidates-daily themes, repercussion of announcements, relations between candidates, importance of candidates in the daily news, etc) that will be used by the political analysts of a campaign.

## 2. Implementation

The first objective was to obtain of news articles, for this we implemented an R package easily configurable according to the structure of each newspaper site. The rvest (Wickham 2016) and xml2 (Wickham, Hester, and Ooms 2018) packages were used for this purpose.

After going through these procedures the data is persisted on Amazon Web Services, we currently use an Aurora Relational Data-store. As a back-up the raw files are stored on AWS S3, which allows us to rerun cleansing procedures and analyses "from source" should this be needed.

For the news analysis our objective was the implementation of linguistic processing to improve the quality of the Spanish text analysis, for this we used the Udpipe package (Straka and Straková 2017). Also we implemented a sentiment analysis useing syuzhet package (Jockers 2015) and the NRC lexicon.

Finally a Shiny App (Chang et al. 2015) was created to display the results of our analysis.

## 3. Access

Url Video: https://www.youtube.com/watch?v=owfI4LU7KeU

## References

Chang, W., J. Cheng, JJ. Allaire, Y. Xie, and J. McPherson. 2015. "Shiny: Web Application Framework for R. R Package Version 0.12.1." Computer Program. http://CRAN.R-project.org/package=shiny.

Jockers, Matthew L. 2015. *Syuzhet: Extract Sentiment and Plot Arcs from Text*. https://github.com/mjockers/syuzhet.

Straka, Milan, and Jana Straková. 2017. "Tokenizing, Pos Tagging, Lemmatizing and Parsing Ud 2.0 with Udpipe." In *Proceedings of the Conll 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 88–99. Vancouver, Canada: Association for Computational Linguistics. http://www.aclweb.org/anthology/K/K17/K17-3009.pdf.

Wickham, Hadley. 2016. *Rvest: Easily Harvest (Scrape) Web Pages*. https://CRAN.R-project.org/package=rvest.

Wickham, Hadley, James Hester, and Jeroen Ooms. 2018. "Xml2: Parse Xml." In. https://CRAN.R-project.org/package=xml2.