

Abriendo y analizando los Diarios de Sesiones del Parlamento uruguayo con R

Keywords: open data, open government, text mining, sentiment analysis, web scraping.

Existen diversas iniciativas de datos abiertos en Uruguay, pero aún queda mucho camino por recorrer. Para que un dato sea considerado abierto, debe cumplir con tres características: accesible (publicado en la web), disponible en un formato que sea legible por computadoras y con una licencia clara de uso. En el caso de los Diarios de Sesiones de Diputados y Senadores (DSDS) el segundo punto no se cumple lo que dificulta el acceso y uso de los mismos. En esta presentación se muestra cómo acceder y preparar los DSDS en un formato adecuado para analizarlos, y realizar un análisis usando text mining. Se utilizó el paquete rvest para obtener de forma sistemática los archivos PDF de las sesiones parlamentarias desde enero de 2017 hasta la actualidad desde <https://parlamento.gub.uy/documentosyleyes/documentos/diarios-de-sesion> y pdftools para extraer el contenido de dichos archivos. Al código fuente del proceso se puede acceder en https://github.com/d4tagirl/uruguayan_parliamentary_session_diary. Con los datos en un formato adecuado se realizó un análisis de text mining con el paquete tidytext. Se calculó la frecuencia de las sesiones y su duración, estimada a través de la cantidad de palabras en el Diario de cada Sesión. Se analizó el texto de las sesiones, centrándose en identificar el sentimiento (usando el Lexicon de Rosas y otros, 2012) para conocer la evolución del sentimiento a lo largo del año (Fig.1) y los temas tratados calculando el tf-idf para medir la importancia de una palabra para una sesión, dado el conjunto de sesiones analizadas. Se nota que el sentimiento empieza a ser más positivo a partir de setiembre de 2017, ante la Renuncia del Vice Presidente Raúl Sendic. Se identifican los temas en las sesiones con sentimientos más extremos (Fig. 2): las dos sesiones más negativas fueron las interpelaciones al Ministro del interior Eduardo Bonomi (extrema violencia en el fútbol) y la interpelación al Canciller Nin Novoa (posición a adoptar ante la crisis venezolana).

En este trabajo se utilizó R de forma exitosa para facilitar el acceso a datos abiertos, que aún necesitan mejorar su acceso, y para analizarlos.

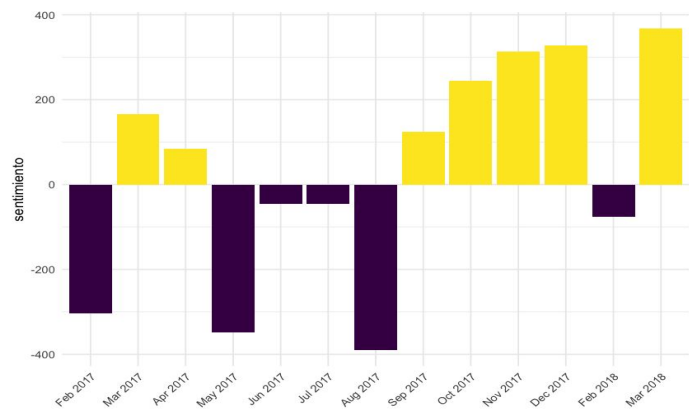


Fig. 1. Evolución del sentimiento de las sesiones de Diputados por mes

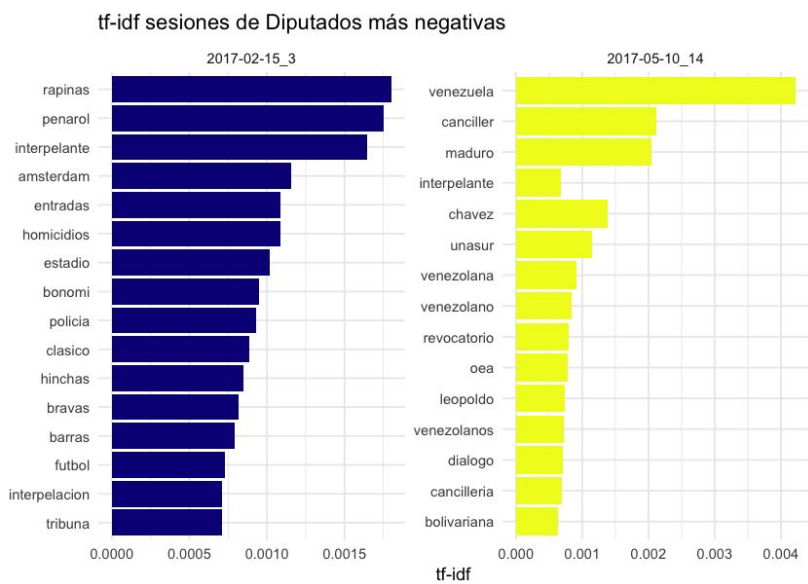


Fig. 2. Tf-idf de los términos con mayor tf-idf, para las dos sesiones de Diputados más negativas