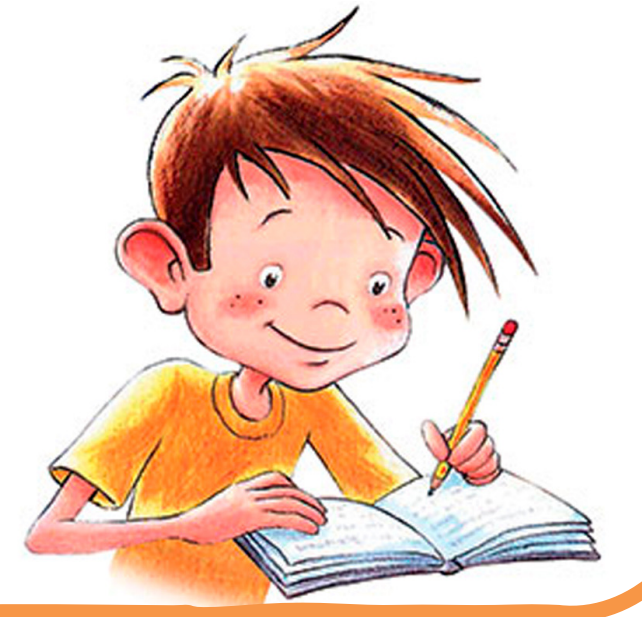




# Análisis de sentimientos basado en minería de texto de la serie de libros Papelucho

Sebastián Villalobos, Manuel Villalobos  
sebastian.villalobos.p@usach.cl , manuel.villalobos@usach.cl



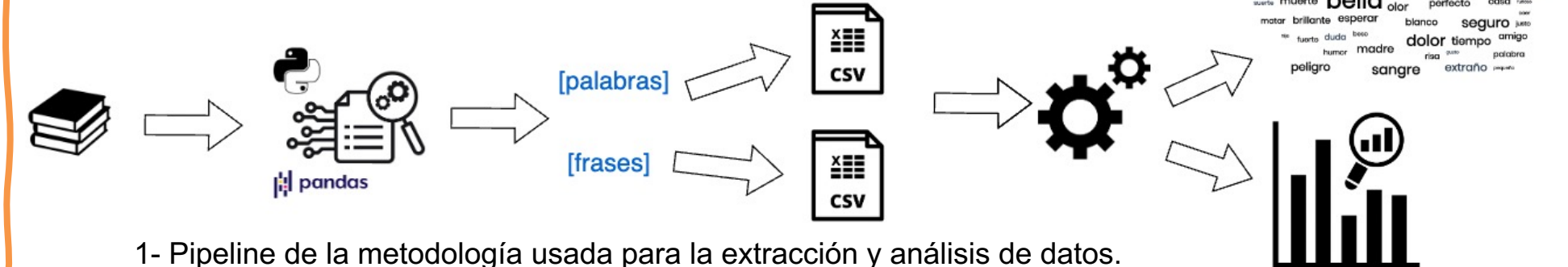
## 1- INTRODUCCIÓN

Según cifras obtenidas en la última publicación de la OCDE, en cuanto a competencias lectoras, Chile obtiene resultados inferiores a los del promedio. Por otro lado, la serie **Papelucho** ha sido estudiada y se ha demostrado que cuenta con características que **ayudan a desarrollar y mejorar estas competencias en los niños lectores**.

Este escenario hace casi indispensable realizar un análisis profundo de estos libros, para así poder caracterizar cuantificando las emociones presentes en el texto, su intensidad durante la narrativa, palabras utilizadas, y su complejidad léxica utilizando técnicas de minería de texto y análisis de sentimientos y "machine learning".

## 2- PROBLEMA

Estudios cualitativos, han indicado que la serie presenta características favorables para el aprendizaje indicando diferencias y similitudes con otros libros. Sin embargo, **no se han realizado estudios que corroboren esto desde el punto de vista cuantitativo**. Diferentes análisis basados en minería de texto y el análisis de sentimientos han permitido efectuar análisis similares en otros contextos usando otros libros.



## 3- SOLUCIÓN PROPUESTA

Dada esta problemática es posible formular las siguientes preguntas de investigación:

- (1) ¿El uso de minería de texto coincidirá con los hallazgos identificados para la serie Papelucho en forma cualitativa por la literatura?
- (2) ¿Qué libros de la serie resultan similares y diferentes entre sí, según sus elementos literarios y sus sentimientos?
- (3) ¿Qué características particulares posee la serie Papelucho a nivel literario y sentimientos que le diferencian o asemejan a otros libros?

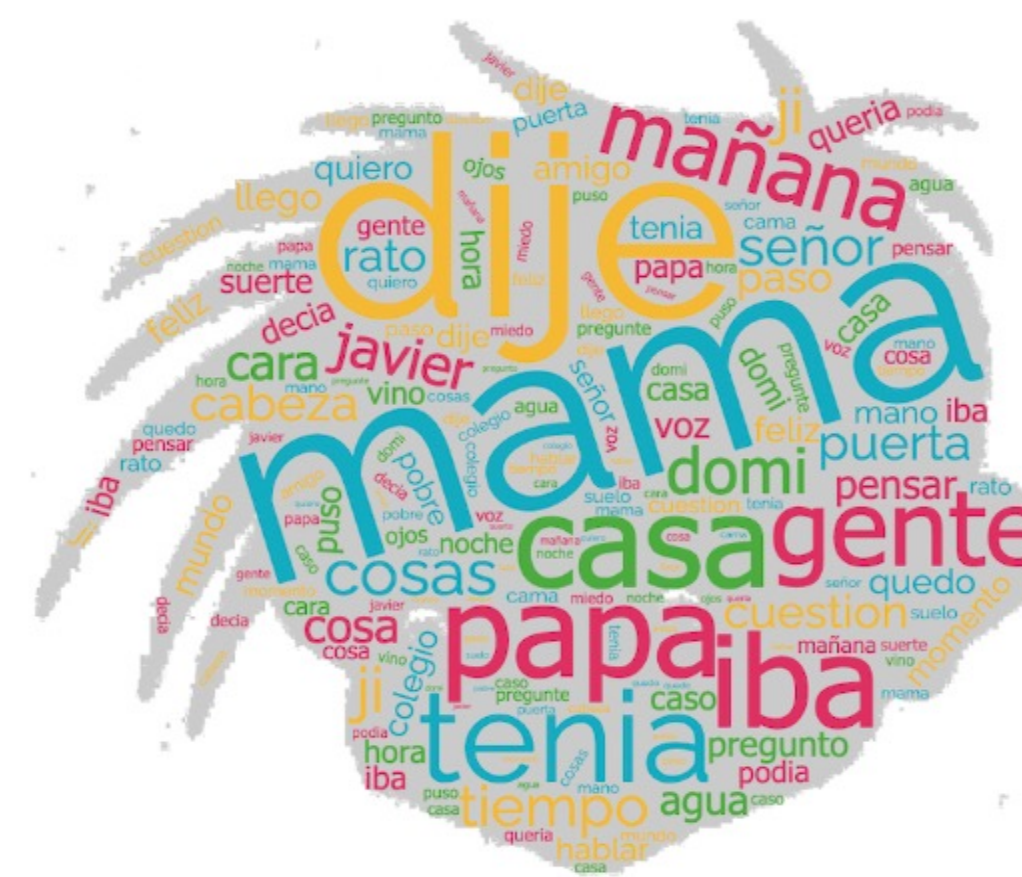
Este trabajo tuvo como objetivo responder estas preguntas utilizando una metodología dividida en cuatro etapas:

- Búsqueda en la literatura de resultados de análisis cualitativos.
- Análisis léxico a nivel cuantitativo, análisis de sentimientos.
- Aplicación de algoritmos de selección de características y agrupamiento.

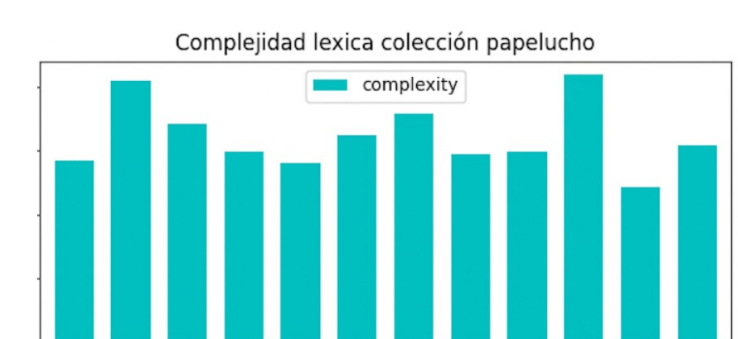
La importancia de realizar esta investigación está dada por el aporte como una herramienta de comparación utilizando los resultados obtenidos con otros libros "similares" dirigidos a desarrollar las competencias lectoras.

## 4- RESULTADOS

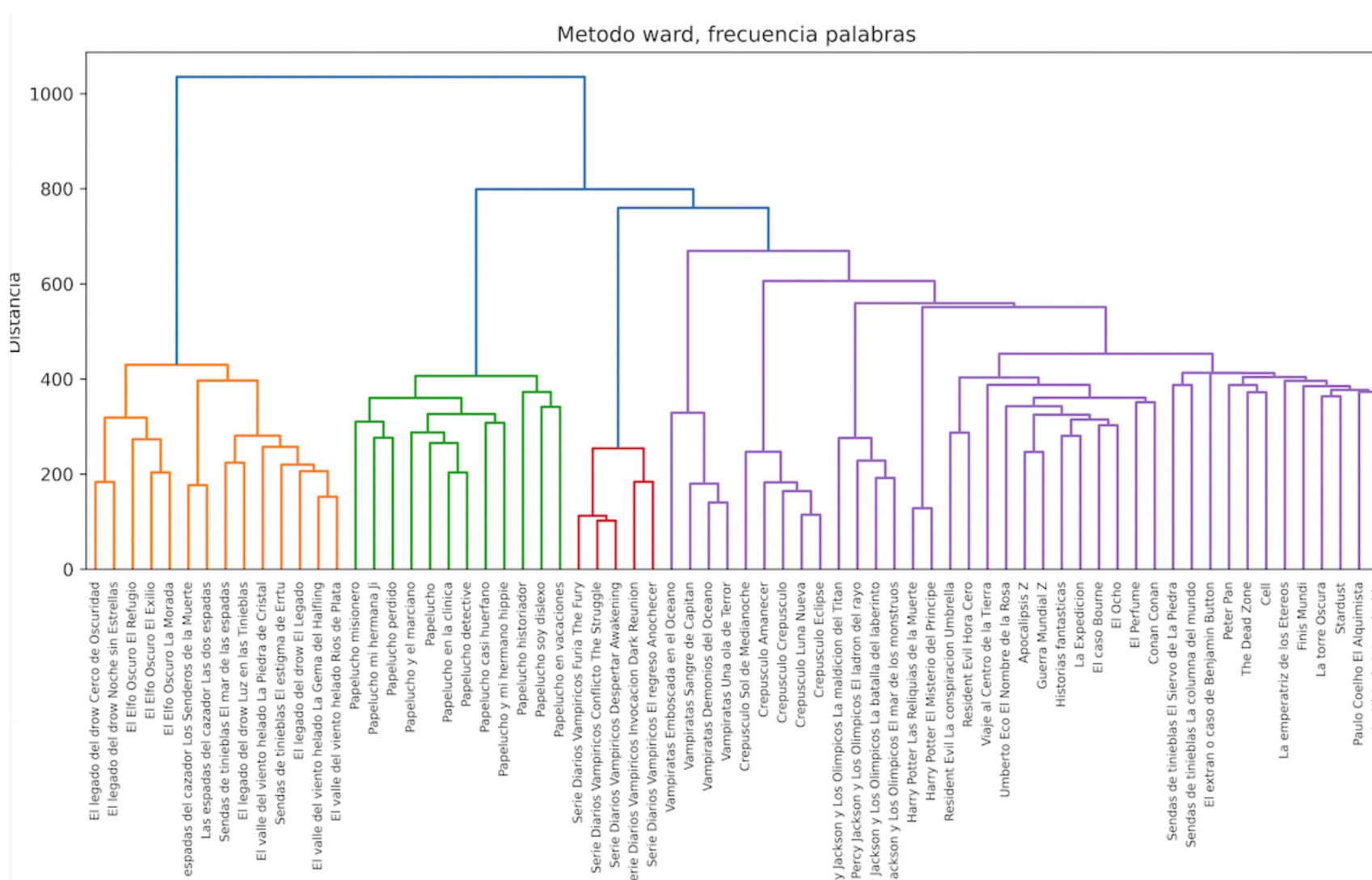
Los resultados obtenidos fueron generados luego de realizar agrupaciones utilizando distintos criterios, como la **intensidad de sentimientos positivos/negativos** en función de la narración, **emociones predominantes** en función de la narración, **complejidad léxica y su frecuencia de palabras**. Estas agrupaciones lograron caracterizar la serie de libros Papelucho y según estos criterios encontrar similitudes y diferencias con otros libros analizados.



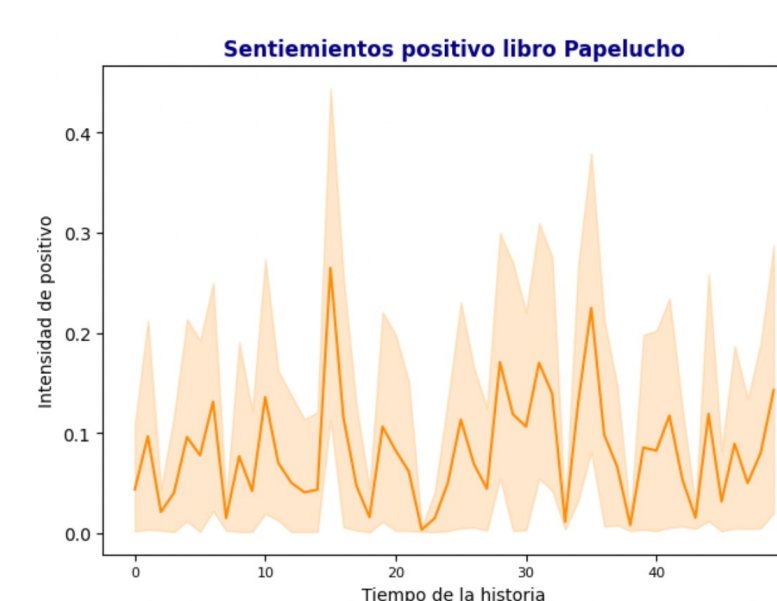
1- Wordcloud de frecuencia de palabras de la colección de libros Papelucho.



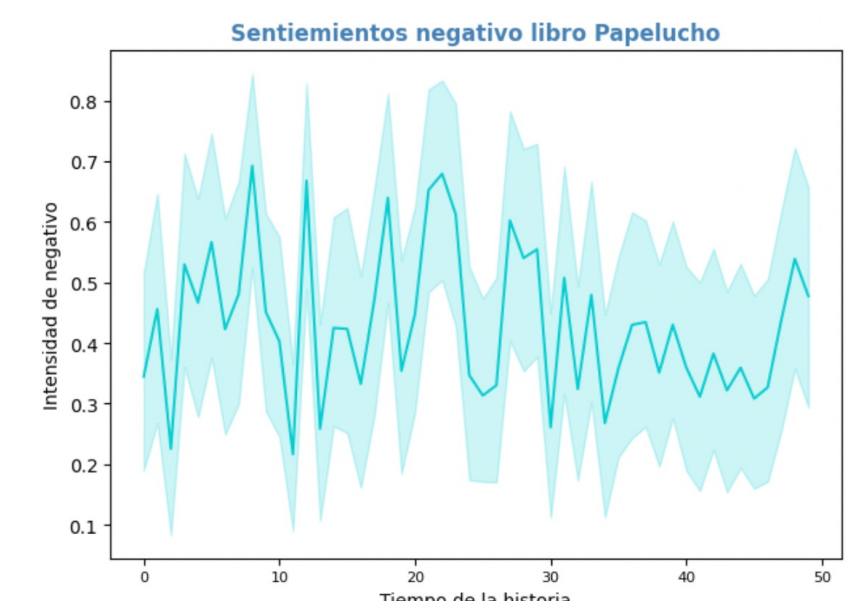
2- Cantidad de palabras distintas por cada texto analizado de la colección Papelucho.



3- Agrupación de textos similares utilizando el método "Ward" en función de frecuencia de palabras en la base de datos de libros completa utilizada en esta investigación.



4- Intensidad de sentimientos positivos y negativos representados en cada gráfico en función del tiempo de narración en la colección de libros Papelucho.



## 5- REFERENCIAS

- Experiencias de lectura literaria en educación básica: **Papelucho** de Marcela Paz, un estudio de caso, [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0123-48702019000200097](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-48702019000200097)
- Patrones cuantitativos en novelas chilenas de los siglos XIX a XX (2015), <https://www.redalyc.org/pdf/1345/134544960014.pdf>

- Sentiment Analysis and Opinion Mining <https://www.morganclaypool.com/doi/abs/10.2200/S00416ED1V01Y201204HLT016>