

Entrega 1:

Análisis de términos usados en las diferentes leyes de educación aprobadas en España desde 1857

Equipo 2:

- ***Tania González Rodríguez***
- ***José Lucas Ruiz González***
- ***Daniel Alberto Melián León***
- ***Javier Castro González***
- ***Noé Leopoldo Campos Delgado***
- ***Eric Rafael Ramos Suárez***



Índice:

1. Introducción.....	3
2. Problemática a desarrollar.....	3
3. Fuentes de datos.....	3
4. Procesamiento de los datos y transformaciones.....	3
5. Resultados esperados.....	4
6. Diagrama de Gantt del desarrollo del proyecto.....	4
7. Propuesta del modelo de negocio para el desarrollo.....	5
8. Creación de un repositorio de contenidos con los aportes del proyecto.....	5

1. Introducción

El proyecto a desarrollar consiste en analizar la terminología usada en las leyes de educación españolas, con el objetivo de detectar diferencias significativas en los términos o palabras más usadas de manera que podamos detectar a qué se le daba mayor importancia en la educación hace 2 siglos y qué tiene mayor importancia en la actualidad. También se pueden comparar las leyes propuestas por las diferentes formaciones políticas que han aprobado leyes de educación al llegar al poder, y quizás detectar los temas a los que da más importancia un partido político u otro o, por el contrario, quizás veamos que, al menos en el tema de la educación, no hay tantas diferencias entre ellos.

2. Problemática a desarrollar

La problemática consistirá en desarrollar un script que lleve a cabo automáticamente una comparativa utilizando la herramienta R con la que obtener una mayor o menor similitud entre la terminología de los documentos y los temas en los que se centran para poder realizar un estudio de la evolución de las leyes educativas en España.

3. Fuentes de datos

Se han escogido las leyes de educación publicadas en el Boletín Oficial del Estado hasta Noviembre de 2016, descargando los pdfs de las mismas y pasándolos a txt de manera que puedan ser analizados por un script en R. Además se ha buscado la ley Moyano de 1857, que no aparecía en el boe debido a su antigüedad. La encontramos en la página de la junta de andalucía y procedimos, igual que con las anteriores, a pasarla a txt para poder trabajar con ella.

4. Procesamiento de los datos y transformaciones

La herramienta R solo permite analizar archivos txt. Sin embargo, los textos a analizar se encuentran en formato pdf. Así pues, para transformar los datos de pdf a txt hemos realizado los siguientes pasos:

1. Utilizado la herramienta online [PDF To Word Convert](#) transformamos los datos a formato .rtf (Word) con vistas a manejar correctamente la información.
2. Abrimos el archivo Word generado y guardamos como .txt.
3. Enviamos los datos a la herramienta R para el análisis correspondiente.

Cabe destacar que no transformamos los datos de pdf a txt directamente, pues la codificación resultante es errónea y el txt producido, a simple vista, no reconoce tildes ni la letra ñ.

5. Resultados esperados

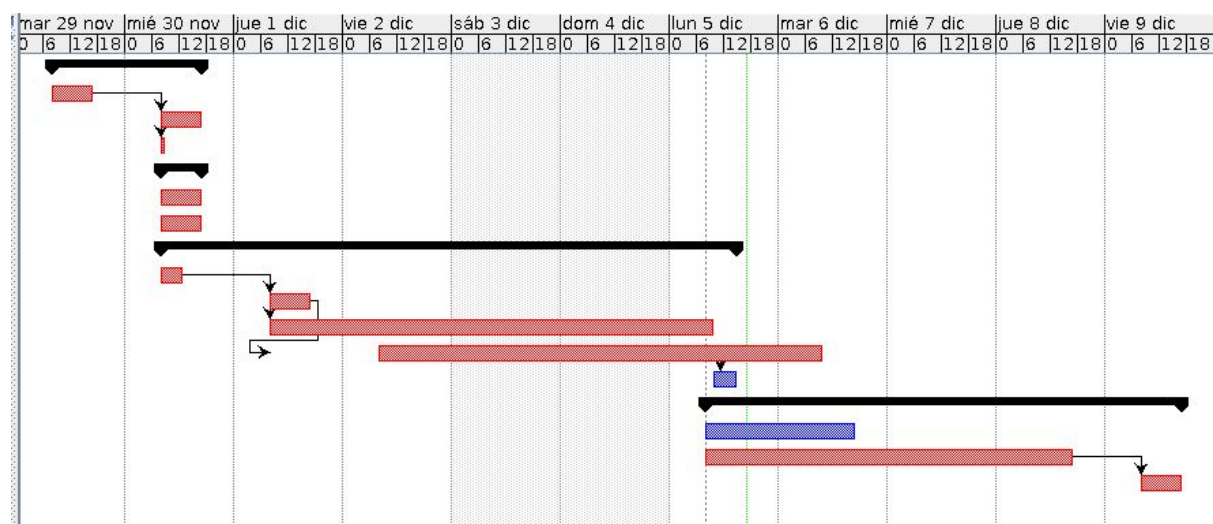
Cuando analicemos los textos se espera encontrar que en la leyes más antiguas se haga más hincapié en términos como: religión, iglesia, santo/a, católica, etc, y que, por el contrario, a medida que avanzamos en el tiempo en las leyes y nos fijamos en la frecuencia de aparición de esos términos, se espera que ésta disminuya o incluso desaparezca.

Por otro lado, parece lógico pensar que en las primeras leyes no se hace mención al género femenino y que en las leyes más modernas los términos como alumna, niñas, etc, aparezcan el mismo número de veces que alumno, niño, etc.

En definitiva, esperamos que el corpus de las primeras leyes y de las últimas sea completamente diferente, ya que históricamente sabemos que la forma de pensar de hace casi 200 años no es en absoluto la misma que la que tenemos actualmente.

6. Diagrama de Gantt del desarrollo del proyecto

	📌	Nombre	Duración	Inicio	Terminado	Predecesores
1	📌	📌Plantear el problema	2 days	29/11/16 8:00	30/11/16 17:00	
2	📌	Plantear el tema del proyecto	1 day	29/11/16 8:00	29/11/16 17:00	
3	📌	Elelgrir proyecto	1 day	30/11/16 8:00	30/11/16 17:00	2
4	📌	Pensar el nombre del grupo	0,125 days	30/11/16 8:00	30/11/16 9:00	2
5	📌	📌Buscar Información	1 day	30/11/16 8:00	30/11/16 17:00	
6	📌	Busqueda de información por libros	1 day	30/11/16 8:00	30/11/16 17:00	
7	📌	Busqueda por internet	1 day	30/11/16 8:00	30/11/16 17:00	
8	📌	📌Realizar trabajo	3,75 da...	30/11/16 8:00	5/12/16 15:00	
9	📌	Buscar información detallada sobre el tema a tratar	0,5 days?	30/11/16 8:00	30/11/16 13:00	
10	📌	Analisis relevante de datos	1 day?	1/12/16 8:00	1/12/16 17:00	9
11	📌	Comparar datos con R	2,25 days?	1/12/16 8:00	5/12/16 10:00	9
12	📌	Extraer información relevante de los txt	2,25 days?	1/12/16 8:00	5/12/16 10:00	10
13	📌	Comprobar correcto funcionamiento del proyecto en R	0,5 days?	5/12/16 10:00	5/12/16 15:00	12
14	📌	📌Informe final	5 days?	5/12/16 8:00	9/12/16 17:00	
15		Primer prototipo de entrega	2 days?	5/12/16 8:00	6/12/16 17:00	
16		Realizar informe	4 days?	5/12/16 8:00	8/12/16 17:00	
17	📌	Comprobar informe	1 day?	9/12/16 8:00	9/12/16 17:00	16
18	📌	📌Presentación	0,125 d...	13/12/16 10:30	13/12/16 11:30	
19	📌	Exposición	0,125 da...	13/12/16 10:30	13/12/16 11:30	



7. Propuesta de modelo de negocio para el desarrollo

Nuestra empresa está formada por un grupo de historiadores especializados en analizar la evolución de los países a través de su legislación.

Las leyes y las modificaciones de las mismas reflejan la mentalidad de las personas de la época, el autoritarismo del gobierno del momento e incluso lo desarrollada que se encuentra dicha sociedad.

A partir de estos datos nos encargamos de realizar análisis para universidades, institutos de historia...

8. Creación de un repositorio de contenidos con los aportes del proyecto.

<https://github.com/alu0100785265/ProyectoTIO>