

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

E.T.S. DE INGENIERÍA INFORMÁTICA - DEPARTAMENTO DE INTELIGENCIA ARTIFICIAL

**Análisis de entornos colaborativos con tecnologías de análisis de redes
y Learning Analytics**

Memoria del trabajo

Autor:

Félix Adame Toledano

Director:

Dr. Antonio Rodríguez Anaya

Máster Universitario en Ingeniería Informática

Septiembre de 2023

Trabajo de Fin de Máster presentado en la Escuela Técnica Superior de Ingeniería
Informática de la Universidad Nacional de Educación a Distancia para la obtención del
Máster Universitario en Ingeniería Informática

Resumen

El presente Trabajo Fin de Máster consiste en la realización y prueba de una aplicación que permita el análisis de sentimientos y emociones, así como el análisis de la red social subyacente de los foros de los cursos virtuales de los diferentes estudios que se pueden cursar en la UNED y por extensión se puede utilizar para analizar otros foros o cualesquiera colecciones de mensajes o textos.

Sobre la base teórica y práctica de trabajos previos se ha construido una aplicación que pretende ser más amigable y fácil de usar y con un potencial uso práctico y de investigación.

Este trabajo se apoya en dos trabajos previos dentro del ámbito de la UNED, como son otro trabajo de fin de máster de la alumna Andrea Rey Presas y la aplicación web *appForum*¹, presentada en las XI Jornadas de Investigación en Innovación Docente de la UNED, si bien es importante reseñar que esta es una aplicación totalmente distinta a cualesquiera otras, tanto en lo que se refiere a la filosofía como a la funcionalidad, presentación, etc. Lo que aquí nos encontramos es una aplicación de escritorio con una interfaz gráfica basada en *Tkinter* (Python). Se organiza en diferentes módulos funcionales, destacando el módulo de modelos e informes, el cual permite obtener un resumen de los análisis realizados y generar un informe automático en formato PDF.

De los mencionados trabajos se han aprovechado algunas estructuras algorítmicas de los mismos, se ha extendido la funcionalidad, se ha creado una interfaz lo más amigable posible y se ha dotado de flexibilidad para poder trabajar con diferentes tipos de archivos, así como configuraciones de los mismos.

¹ <https://ceur-ws.org/Vol-3224/paper21.pdf>

Se propone un método alternativo para el descubrimiento de sentimientos de los mensajes. Dicho método consiste en la combinación lineal de varios métodos, cuya fórmula es configurable.

Respecto del análisis de emociones se realizan dos análisis diferentes que pueden compararse. Un primer análisis basado en regiones de Russell a partir de un corpus de términos etiquetados con valores de valencia y excitación. El segundo método consiste también en la utilización de un corpus de términos, pero en este caso se les han asignado puntuaciones para cada una de varias emociones consideradas primarias. Este segundo método se ha probado con el corpus de 2.226 palabras puntuadas para cinco emociones primarias creado por (Ferré et al.) en el trabajo *Moved by words: Affective ratings for a set of 2,266 Spanish words in five discrete emotion categories*.

La aplicación se ha probado con datos reales de foros de la UNED anonimizados y se ha utilizado como herramienta para analizar sentimientos en otro trabajo fin de máster: *Estudio sobre la búsqueda de mentor utilizando aprendizaje automático* realizado por D. Juan García Ruiz y dirigido también por el Dr. Antonio Rodríguez Anaya.

Palabras clave

Social Network Analysis; Sentiment Analysis; Discrete emotion categories; Affective norms; Emotional effects on word processing; Learning Analytics.

Abstract

This final master's project consists of the development and testing of an application to carry out both the analysis of feelings and emotions of the messages sent in the forums of the UNED university, as well as the analysis of the social network created in said forums. By extension it can be used to analyze other forums or any collections of messages or texts.

On the theoretical and practical basis of previous work, an application has been built that aims to be friendlier and easier to use and with potential practical and research use.

This work is based on two previous works, such as another master's thesis by the student Andrea Rey Presas and a web application (*appForum*²), presented at the XI Conference on Research in Teaching Innovation of the UNED, although it is important to note that this is an application totally different from any other, both in terms of approach and functionality, presentation, etc. What we find here is a desktop application with a graphical interface based on *Tkinter* (Python). It is organized into different functional modules, highlighting the models and reports module, which allows obtaining a summary of the analyzes carried out and generating an automatic report in PDF format.

Some of the code structures of the works mentioned above have been used and the functionality has been extended, an interface that is as friendly as possible has been created and flexibility has been provided to be able to work with different types of files, as well as their configurations.

An alternative method is proposed for the discovery of message sentiments based on the linear combination of several methods. The formula that defines this method can be configured.

Regarding the analysis of emotions, two different analyzes are carried out that can be compared. A first analysis based on Russell regions from a corpus of terms labeled

² <https://ceur-ws.org/Vol-3224/paper21.pdf>

with valence and excitation values. The second method also consists of using a corpus of terms, but in this case scores have been assigned for each of several emotions considered primary. This second method has been tested with the corpus of 2,226 words scored for five primary emotions created by (Ferré et al.) in the work entitled *Moved by words: Affective ratings for a set of 2,266 Spanish words in five discrete emotion categories*.

The application has been tested with real data from anonymized UNED forums and it has been used as a tool to sentiment analysis in another master's thesis: *Study on the mentor searching using machine learning* carried out by Juan García Ruiz and also directed by Dr. Antonio Rodríguez Anaya.

Key words

Social Network Analysis; Sentiment Analysis; Discrete emotion categories; Affective norms; Emotional effects on word processing; Learning Analytics.

Índice General

Resumen	2
Palabras clave	3
Abstract	4
Key words.....	5
Índice General.....	6
Índice de tablas	12
Índice de figuras	13
Acrónimos	21
Capítulo 1 Introducción	22
1.1. Qué es Learning Analytics.....	22
1.2. Motivaciones para Learning Analytics	23
1.3. Las actividades de Learning Analytics	24
1.4. Estado del Arte.....	24
1.5. Objetivos.....	25
Capítulo 2 Fundamentos de Learning Analytics	28
2.1. Métodos de Minería de Datos	28
2.2. Análisis de Sentimientos y Emociones	30

2.2.1.	Fundamentos del análisis de sentimientos y emociones.....	31
2.2.2.	Análisis de Sentimientos	36
2.2.3.	Análisis de Emociones	43
2.3.	Análisis de Redes Sociales	49
2.4.	La explosión de la IA.....	50
Capítulo 3 Trabajo previo		51
3.1.	Interfaz.....	51
3.2.	Ayuda.....	57
3.3.	Dinámica de Funcionamiento	57
3.4.	Aspectos no cubiertos (oportunidades de mejora)	58
3.4.1.	Persistencia de resultados.....	58
3.4.2.	Generación de informes	58
3.4.3.	Anonimización de datos	58
Capítulo 4 Trabajo realizado		59
4.1.	Aprovechamiento del trabajo anterior	60
4.2.	Aportaciones significativas sobre el trabajo anterior.....	61
4.3.	Estructura de la aplicación creada.....	63
4.3.1.	Estructura de ficheros	63

4.4.	Interfaz de la aplicación y funcionamiento	67
4.4.1.	Preferencias de Tema.....	68
4.4.2.	Configuración	69
4.4.3.	Creación de un proyecto	70
4.4.4.	Análisis de la Red	79
4.4.5.	Análisis de Sentimientos	85
4.4.6.	Análisis de Emociones	92
4.4.7.	Nube de Palabras	99
4.4.8.	Análisis Temporal	100
4.4.9.	Modelos e Informes	103
4.5.	Distribución e Instalación de la aplicación	111
4.6.	Diseño de la aplicación desarrollada	116
4.6.1.	Diagrama de Casos de Uso de la aplicación.....	116
4.6.2.	Diagrama de Clases	117
4.6.3.	Modelado UML	119
4.6.4.	Cronograma del desarrollo y realización del TFM.....	120
Capítulo 5	Pruebas	124
	Comparativa de análisis de Red	124

Comparativa con <i>Gephi</i>	124
Comparativa de Sentimientos	129
Encuesta a usuarios de prueba	132
Capítulo 6 Conclusiones.....	138
Capítulo 7 Posibles líneas de mejora y evolución	142
Apéndice A Ayuda de la aplicación	143
Como usar la aplicación.....	143
Cargar Datos de Foros	143
A.1 Ayuda del módulo de Análisis de Red.....	145
Valores de los Nodos	145
Visualización de la red formada por los participantes atendiendo a diferentes estadísticos	147
Etiquetas de gráfico	147
Red de Sentimientos	148
Herramientas de Gráfico	149
Visualizar Archivo	149
Filtros	150
Exportar Red	153
A.2 Ayuda del módulo de Análisis de Sentimientos	154

Sentimiento medio	154
Tabla de sentimientos	158
Gráficos.....	160
Gráficos de tarta	160
Gráficos de caja.....	160
Análisis de Sentimientos de los mensajes mediante iconos.....	161
Leyenda.....	162
A.3 Ayuda del módulo de Análisis de Emociones	163
Emociones Promedio de cada participante según regiones de Russell.....	163
Tabla de Emociones.....	165
Gráficos de Emociones	166
Regiones de Russell.....	167
Análisis Puramente Aritmético.....	169
Medias Aritméticas por Usuario	170
Leyenda de los iconos.....	171
A.4 Ayuda del módulo de Análisis Temporal	173
Gráfica de Mensajes por Fecha.....	173
Visualización de Datos	174

Gráfica de Mensajes por Día de la Semana	174
A.5 Ayuda del módulo de Modelos e Informes	175
Visualización de Datos	175
Columnas de la 3ª a la 8ª	176
Columnas de la 9ª a la 18ª	177
Columnas de la 19ª a la 23ª	177
Columnas de la 24ª a la 32ª	177
Agrupamiento (<i>Clustering</i>).....	178
Configuración del agrupamiento	179
Función codo.....	179
Gráficos de Usuarios.....	180
Informes en PDF	183
Exportación a Excel	186
Apéndice B Encuesta a usuarios probadores.....	188
Apéndice C Informe PDF Generado Automáticamente por la Aplicación.....	207
Referencias bibliográficas.....	230
Índice alfabético.....	235

Índice de tablas

Tabla 1. Acrónimos	21
Tabla 2. Objetivos del presente trabajo.	26
Tabla 3. Resultados de la encuesta sobre experiencia de usuario inicial.	133
Tabla 4. Resultado de la encuesta. Para qué sirve la aplicación.....	134
Tabla 5. Resultado de la encuesta. Qué ofrece la aplicación.....	135
Tabla 6. Resultado de la encuesta. Propuestas de mejora.	136
Tabla 7. Resultados de la encuesta. Usabilidad y utilidad general.....	137
Tabla 8. Consecución de objetivos del trabajo.	140

Índice de figuras

Figura 1. Interconexión entre neuronas. Fuente: https://www.shutterstock.com/	34
Figura 2. Perceptrón con cinco señales de entrada. Fuente: https://es.wikipedia.org/wiki/Perceptrón	34
Figura 3. Ilustración original del trabajo de Scholberg, 1952. Representación en dos dimensiones (P-U y A-R) de 72 fotografías de expresiones faciales. El eje P-U representa <i>pleasantness-unpleasantness</i> (agrado-desagrado) y el eje A-R <i>attention-rejection</i> (atención-rechazo).	45
Figura 4. Ilustración original del trabajo de Scholberg, 1954. Sólido que representa las tres dimensiones de las emociones correspondientes a expresiones faciales.	46
Figura 5. Adaptación de los resultados de Russell, 1980 en castellano para este trabajo.	46
Figura 6. Óvalo de Posner de las emociones	48
Figura 7. Página principal de la aplicación del trabajo previo.	52
Figura 8. Pantalla de análisis de sentimientos de la aplicación andel trabajo anterior... ..	52
Figura 9. Página de resultados de sentimientos de la aplicación del trabajo anterior.	53
Figura 10. Interfaz del análisis de redes de la aplicación del trabajo anterior.....	54
Figura 11. Análisis de red (Estadístico grado) de la aplicación del trabajo anterior.....	54
Figura 12. Valores estadísticos de los nodos.....	55
Figura 13. Análisis de sentimientos y emociones de la aplicación anterior.....	55

Figura 14. Algoritmo de <i>clustering</i> de la aplicación anterior.....	56
Figura 15. Análisis de red de sentimientos de la aplicación anterior.	56
Figura 16. Ayuda contextual sobre el formato de los archivos.	57
Figura 17. Aspecto de la interfaz de la aplicación inmediatamente después de abrirse.	68
Figura 18. Preferencias de Tema.	68
Figura 19. Aspecto de la interfaz con el tema Azul RELAX.	69
Figura 20. Menú de <i>Configuración</i>	69
Figura 21. Archivo <i>config.ini</i>	70
Figura 22. Módulo de Proyectos al acceder al mismo tras abrir la aplicación.	70
Figura 23. Ayuda contextual para la creación de proyectos.....	71
Figura 24. Resto de pasos de la ayuda de creación de proyecto.....	72
Figura 25. Estructura de directorios de un proyecto.....	72
Figura 26. Habilitación del botón de carga de datos de foros.	73
Figura 27. Cuadro de diálogo de apertura de archivo de datos de foros.	74
Figura 28. Pantalla de Proyectos con la sección de normalización de archivos.....	75
Figura 29. Mensaje indicando que aún no procede el guardado de proyecto.....	76
Figura 30. Pantalla de procesamiento de datos.....	77
Figura 31. Interfaz de la aplicación una vez realizado el preprocesamiento de los datos de los foros.	78

Figura 32. Cuadro de aviso informando del guardado del proyecto.	79
Figura 33. Vista inicial del módulo de Análisis de Red.	79
Figura 34. Grafo de la red según la Centralidad de Grado.	81
Figura 35. Red según Centralidad de Grado con etiquetas de usuarios y zoom aplicado.	82
Figura 36. Mismo grafo de la figura 35 pero sin <i>zoom</i>	82
Figura 37. Red de Sentimientos según <i>Naive Bayes</i>	83
Figura 38. Red de Sentimientos según el algoritmo <i>PySentimiento</i>	83
Figura 39. Modularidad de la red.	84
Figura 40. Filtrado por fecha en la pantalla de <i>Análisis de Red</i>	85
Figura 41. Grafo de red en un intervalo de tiempo.	85
Figura 42. Tabla de <i>Sentimiento medio combinado</i> en el módulo de Análisis de Sentimientos.	87
Figura 43. Correspondencia entre el Sentimiento Combinado y su traducción literal (Particiones).	88
Figura 44. Tabla de Sentimientos.	89
Figura 45. Gráficos de Sentimientos.	90
Figura 46. Tabla de análisis de sentimientos de cada mensaje con iconos.	91
Figura 47. Emociones promedio.	93

Figura 48. Contenido parcial del archivo utilizado como corpus por defecto para el cálculo de las emociones según Regiones de Russell.	94
Figura 49. Tabla de emociones.....	95
Figura 50. Contenido de un mensaje al pie de la tabla de emociones.	95
Figura 51. Gráficos de Emociones.	96
Figura 52. Análisis de regiones de Russell de cada mensaje.....	97
Figura 53. Análisis llamado <i>aritmético</i>	98
Figura 54. Emociones medias para los participantes según el segundo método aritmético.	99
Figura 55. Nube de Palabras.....	100
Figura 56. Mensajes por fecha en el módulo de análisis temporal.....	101
Figura 57. Mensajes por día de la semana.....	102
Figura 58. Tabla de datos temporales.....	103
Figura 59. Tabla con información de los participantes de los foros.....	104
Figura 60. Gráfico de agrupamiento de usuarios (clustering).	106
Figura 61. Función codo que facilita la elección del número de grupos (clusters).	107
Figura 62. Selección de usuario para dibujar su gráfico.....	108
Figura 63. Gráficos de radar de dos usuarios.	110
Figura 64. Creación de un nuevo proyecto.....	112

Figura 65. Creación de nuevo proyecto utilizando el código fuente existente.....	112
Figura 66. Apertura del proyecto que acabamos de crear.	113
Figura 67. Reconocimiento del archivo <i>requirements.txt</i> por el entorno de desarrollo <i>PyCharm</i>	113
Figura 68. Paquetes que se instalarán por estar recogidos en <i>requirements.txt</i>	114
Figura 69. Indicación de los paquetes que se están instalando en el entorno virtual. ..	114
Figura 70. Descarga del corpus para análisis de sentimientos en español.	115
Figura 71. Instalación de <i>es-core-news-sm</i> e indicación de que existe una nueva versión de <i>pip</i>	115
Figura 72. Descarga de contenido adicional.....	116
Figura 73. Diagrama de casos de uso.	117
Figura 74. Diagrama de clases.....	118
Figura 75. Modelado UML de las clases.....	119
Figura 76. Primera fase del desarrollo del TFM.....	121
Figura 77. Segunda fase del desarrollo del TFM.	121
Figura 78. Previsualización de la red en Gephi.	125
Figura 79. Representación de la red en nuestra aplicación.	126
Figura 80. Datos de los mensajes (aristas) en Gephi.....	126
Figura 81. Datos de los mensajes (aristas) en nuestra aplicación.....	127

Figura 82. Estadísticas de la red en Gephi.....	127
Figura 83. Estadísticas de la red en nuestra aplicación.	128
Figura 84. Correspondencia entre sentimiento medio combinado y sentimiento literal.	132
Figura 85. Valores de los Nodos.	146
Figura 86. Representación de la centralidad de Grado.....	147
Figura 87. Selector de etiquetas para los nodos.....	148
Figura 88. Red de Sentimientos N. Bayes.	148
Figura 89. Red de Sentimientos PySentimiento.	149
Figura 90. Caja de Herramientas de Gráfico (grafo).....	149
Figura 91. Tabla de visualización de archivo.	150
Figura 92. Filtro por fecha.....	151
Figura 93. Visualización de la red mostrando Page Rank con un filtro de fecha aplicado.	151
Figura 94. Grafo de Cercanía con auto-respuestas.....	152
Figura 95. Grafo de Cercanía sin auto-respuestas.	152
Figura 96. Botón de Exportar Red.....	153
Figura 97. Vista de sentimiento medio combinado.	156
Figura 98. Correspondencia entre el valor numérico del sentimiento combinado y el literal e icono correspondiente.....	156

Figura 99. Tabla de Sentimientos.	158
Figura 100. Gráficos de Sentimientos.	160
Figura 101. Tabla de análisis de sentimientos de cada mensaje con iconos.	161
Figura 102. Leyenda visible en la pantalla de Sentimientos con Iconos.	162
Figura 103. Pantalla de Emociones Promedio de cada Participante según Regiones de Russell.	163
Figura 104. Archivo utilizado como corpus por defecto para el cálculo de las emociones según Regiones de Russell.	164
Figura 105. Tabla de mensajes y emociones.	165
Figura 106. Detalle del contenido de una celda de la tabla.	166
Figura 107. Gráficos de Emociones.	167
Figura 108. Pantalla de Regiones de Russell de cada mensaje.	169
Figura 109. Tabla de análisis puramente aritmético (análisis 2 con corpus 2).....	170
Figura 110. Medias aritméticas de emociones por usuario y emoción dominante por usuario.	171
Figura 111. Leyenda de los iconos visible en la tabla de emociones dominantes en cada usuario según el método aritmético (método 2).	172
Figura 112. Gráfico de Mensajes por Fecha.....	173
Figura 113. Datos temporales.....	174
Figura 114. Gráfica de mensajes por día de la semana.	175

Figura 115. Tabla de datos de los participantes de los foros.....	176
Figura 116. Gráfico de agrupamiento de usuarios (clustering).	178
Figura 117. Función codo que facilita la elección del número de grupos (clusters). ...	180
Figura 118. Selección de usuario para dibujar su gráfico.....	181
Figura 119. Gráficos de radar de dos usuarios.	182
Figura 120. Visualización del archivo informe_1.html.....	185
Figura 121. Visualización del archivo informe_modelo.html.....	186
Figura 122. Archivo de Excel generado mediante la exportación.....	187

Acrónimos

En la tabla 1 se muestra una lista de los acrónimos utilizados en este trabajo.

Tabla 1. Acrónimos

Acrónimo	Significado
BI	Business Intelligence
CIS	Community Information System
CLA	Community Learning Analytics
CNN	Convolutional Neural Network
CoP	Community of Practice
CSBA	Computer-Supported Behavioral Analytics
CSLA	Computer-Supported Learning Analytics
CSPA	Computer-Supported Predictive Analytics
CSVA	Computer-Supported Visualization Analytics
DM	Data Mining
EDM	Educational Data Mining
LA	Learning Analytics
LSTM	Long Short Term Memory
MOOC	Massive Online Open Courses
SNA	Social Network Analysis
TFM	Trabajo Fin de Máster
CSBA	Computer-Supported Behavioral Analytics
CSLA	Computer-Supported Learning Analytics
CSPA	Computer-Supported Predictive Analytics
CSVA	Computer-Supported Visualization Analytics
K-NN	K- Nearest Neighbors
LDA	Latent Dirichlet Allocation
ML	Machine Learning
NLP	Natural Language Processing
pLSA	Probabilistic Latent Semantic
SA	Sentiment Analysis
SVM	Support Vector Machine

Capítulo 1

Introducción

En esta sección se realiza una breve introducción a los conceptos teóricos fundamentales detrás de la generalización “*Learning Analytics*”. Se hace un breve repaso histórico y se contextualiza con respecto a las ramas a partir de las cuales surge y evoluciona o con las que comparte campos teóricos y de investigación. Además, se describen los objetivos del presente trabajo, contextualizándolo con respecto a lo anterior.

1.1. Qué es Learning Analytics

El término “*Learning Analytics*”, en adelante **LA**, no se usó mucho antes de 2011. No fue hasta que apareció en la 1ª Conferencia Internacional sobre Aprendizaje y Analítica que se convirtió en un término más común. De esta época data el artículo “*Learning Designs and Learning Analytics*” (Lockyer et al., 2011) [1], que explora la relación entre el diseño y las analíticas de aprendizaje.

Según la Wikipedia³, *Learning Analytics* (Análisis del aprendizaje) es la medición, recopilación, análisis e informe de datos sobre los alumnos y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que se produce.

Otra definición la encontramos en el artículo “*Learning Analytics: state of the art*” (Hernández et al., 2022) [2], según el cual “*Learning Analytics*” es un campo que mide, analiza e informa datos sobre los estudiantes y sus contextos para comprender y mejorar el aprendizaje y el lugar en el que ocurre.

³ URL: https://en.wikipedia.org/wiki/Learning_analytics

1.2. Motivaciones para Learning Analytics

La enseñanza a distancia basada en medios informáticos ha ido ganando popularidad de manera exponencial, en consonancia con el crecimiento y evolución de la World Wide Web⁴, tanto a nivel cuantitativo (el acceso a Internet se multiplicó por 100 en la década de 1995 a 2015)⁵, como a nivel cualitativo (Se ha evolucionado de una Internet de páginas web estáticas a una extensión del mundo físico en el llamado ciberespacio, donde se desarrollan todo tipo de actividades comerciales, de entretenimiento, sociales, y por supuesto formativas). La evolución de las tecnologías software y hardware ha derivado en la proliferación de diferentes alternativas para la enseñanza *online*. Desde Plataformas genéricas de código abierto del estilo de Moodle⁶, (Hernández et al., 2022) [2] hasta plataformas privadas del estilo de Valamis⁷, no pudiendo dejar de mencionar las universidades, tanto públicas como privadas, que han potenciado la formación online de manera activa, sobre todo durante la última década. Un concepto que ha impactado de manera importante en la educación es el de MOOC, del inglés (Massive Online Open Courses), en castellano, Cursos Online Masivos y Abiertos. Cursos que desde algunas universidades y otras instituciones, se ofrecen de manera masiva a prácticamente a cualquiera que disponga de una conexión a Internet y esté interesado en mejorar su formación.

En el contexto descrito surge la necesidad de establecer métricas para evaluar los resultados de las acciones formativas y poder implementar ciclos de mejora continua. Es aquí donde entra en juego LA. Las instituciones educativas tienen diferentes motivaciones para utilizar LA. Entre ellas se encuentran por ejemplo, mejorar los resultados de los estudiantes, optimizar su tecnología educativa, reducir la tasa de deserción y otros.

⁴ Red Informática mundial. Conocida normalmente como Internet.

⁵ URL: <https://es.wikipedia.org/wiki/Internet>

⁶ URL: <https://moodle.org/>

⁷ URL: <https://www.valamis.com/>

1.3. Las actividades de Learning Analytics

LA se ha convertido desde hace una década en un área importante de investigación. En particular, las instituciones de aprendizaje buscan formas de recopilar, gestionar, analizar y explotar los datos de los alumnos e instructores para facilitar los procesos de aprendizaje (Klamma, R., 2013) [3]. Las actividades relacionadas con LA son similares a las de cualquier otro campo que requiera analizar datos para mejorar ciertos resultados. Esto nos lleva principalmente a tres áreas que en los últimos años han experimentado un fuerte crecimiento, como son: por un lado el análisis de redes sociales, *Social Network Analysis (SNA)*, por otro lado el análisis de grandes cantidades de datos o **minería de datos** y finalmente el **aprendizaje automático**, aplicando tecnologías de inteligencia artificial. El desarrollo y la mejora de la capacidad de cómputo y las estrategias de procesamiento en la nube, han propiciado, junto con la mejora de los algoritmos empleados, que se disponga actualmente de una capacidad de procesamiento excepcional.

Estas tres disciplinas proporcionan tres pilares básicos a LA.

Más concretamente, dentro del ámbito del presente trabajo, nos centraremos en el análisis de foros de debate de las asignaturas. Los foros son una herramienta de gran valor dentro de la educación a distancia, en tanto en cuanto constituyen a veces el principal medio de comunicación entre los alumnos y el equipo docente, sirviendo también para que los alumnos interactúen entre ellos. Sobre los datos de los foros (mensajes) se pueden realizar las tareas descritas de análisis de redes y aprendizaje automático encaminadas a detectar, por ejemplo, alumnos con dificultades, niveles de participación, patrones de interacciones, etc. Esto se puede apoyar en diferentes técnicas de análisis de sentimientos y emociones por ejemplo.

1.4. Estado del Arte

LA (Song et al., 2011) [4] y *Educational Data Mining (EDM)* (Baker et al., 2009) [5] se han convertido desde hace más de una década en importantes áreas de investigación en el aprendizaje mejorado por la tecnología. LA se centra en la evaluación de los procesos de aprendizaje formal en entornos *on-line* o semipresenciales. (Shum et al., 2012) [6] Tiene su origen en técnicas de Inteligencia de Negocio y Minería de Datos

(Klamma, R., 2013) [3], o más conocidos por sus expresiones en inglés: **Business Intelligence (BI)** y **Data Mining (DM)**. El Informe Horizon 2013 escribía: “*Análisis de aprendizaje, en muchos sentidos, es big data aplicado a la educación*” (Johnson et al., 2013) [7].

1.5. Objetivos

Al abordar el presente Trabajo Fin de Máster se han definido varios objetivos que responden a diferentes vertientes.

Por un lado, y en una vertiente puramente del aprendizaje, se pretende adoptar unos conocimientos suficientes sobre el lenguaje de programación Python como para poder desarrollar una aplicación de escritorio, partiendo de unos conocimientos poco más que superficiales (OBJ1).

En el ámbito técnico, se deben explorar, dentro del ecosistema del lenguaje adoptado (Python), las librerías o módulos puestos a disposición por diferentes desarrolladores, que puedan ser de utilidad para los propósitos de la aplicación y utilizarlos de la manera adecuada (OBJ2).

Por otro lado, y dentro del ámbito de la ingeniería, se persigue, siguiendo las indicaciones de la propuesta, partiendo de código fuente desarrollado por otras personas y con una mínima documentación, ser capaz de comprender su funcionamiento y poder reutilizar parte del mismo, realizando las modificaciones necesarias para el propósito definido (OBJ3).

En el ámbito puramente finalista, se pretende llegar a desarrollar una aplicación de escritorio relativamente fácil de distribuir (instalar) por terceras personas (OBJ4) que se pueda utilizar para analizar los mensajes de un grupo de foros, o por extensión, cualquier colección de mensajes o textos, obteniendo:

- Una representación en forma de grafo de la red social subyacente, donde los nodos sean los participantes y las aristas representen las respuestas de unos a otros. Se debe mejorar la solución del trabajo anterior en términos de usabilidad y funcionalidad (OBJ5).
- Los estadísticos o medidas de centralidad más relevantes para el análisis de redes sociales, así como una tabla con toda la información posible de cada participante (OBJ6).
- Análisis de sentimientos y emociones de los mensajes y por extensión de los participantes, para **poder inferir qué participantes pueden requerir una especial atención**. Debe mejorarse y/o ampliarse la funcionalidad del trabajo anterior (OBJ7).
- Diversas representaciones gráficas que ayuden a conocer la comunidad de los foros en términos agregados (de modo general), así como a nivel individual (OBJ8).
- Una aplicación con una interfaz más amigable, integrada y funcional que la del trabajo anterior (OBJ9).
- Una aplicación con mucha más flexibilidad en lo que se refiere al formato de los datos de entrada (OBJ10).

Además, la creación de la aplicación deberá realizarse siguiendo una metodología ágil⁸ (OBJ11).

En la tabla 2 se representan estos objetivos de manera resumida.

Tabla 2. Objetivos del presente trabajo.

Código	Objetivo	Carácter
OBJ1	Adquirir conocimientos sobre el lenguaje de programación Python	Aprendizaje
OBJ2	Utilización de librerías	Técnico
OBJ3	Comprensión, reutilización y aprovechamiento de código de terceros	Ingenieril
OBJ4	Desarrollo de aplicación fácil de distribuir	Desarrollo
OBJ5	Mejorar la representación de la red social	Desarrollo
OBJ6	Estadísticas, medidas de centralidad y datos de los participantes mejorados y ampliados	Desarrollo
OBJ7	Mejora y ampliación de análisis de sentimientos y emociones	Desarrollo

⁸ https://es.wikipedia.org/wiki/Desarrollo_%C3%A1gil_de_software

Código	Objetivo	Carácter
OBJ8	Representación gráfica de la información de forma mejorada y ampliada a nivel individual y agregado	Desarrollo
OBJ9	Conseguir una aplicación con una interfaz más amigable, integrada y funcional.	Desarrollo
OBJ10	Flexibilizar el formato de los datos de entrada	Desarrollo
OBJ11	Adopción de metodología ágil	Ingenieril

Capítulo 2

Fundamentos de Learning Analytics

En este apartado se realiza una exposición de los fundamentos teóricos en que se basa este trabajo. La aplicación desarrollada contiene lo que podríamos denominar diferentes módulos funcionales: El análisis de sentimientos de los mensajes emitidos por los diferentes participantes de los foros, el análisis de las emociones de dichos mensajes, el análisis estadístico cuantitativo de los mensajes y el análisis de la red social producida en los foros.

2.1. Métodos de Minería de Datos

Los métodos de minería de datos se pueden clasificar en cuatro grupos (Hernández et al., 2022) [2]:

- **Computer-Supported Learning Analytics (CSLA).** Analíticas de Aprendizaje Asistidas por Computador. Utilizan algoritmos de minería de datos para desarrollar la interacción de los estudiantes en el sistema de gestión de la enseñanza (LMS). Identifica oportunidades de aprendizaje evaluando el intercambio entre estudiantes y sus resultados.
- **Computer-Supported Predictive Analytics (CSPA).** Análisis Predictivo Asistido por Computador. Utilizado para predecir el rendimiento y la retención de los estudiantes evaluando varias dimensiones, como la participación, participación y calificaciones.
- **Computer-Supported Behavioral Analytics (CSBA).** Análisis de Comportamiento Asistido por Computador. Muestra el comportamiento y las preferencias o motivaciones de los estudiantes en un ambiente de aprendizaje mientras participan en varias actividades académicas diferentes.

- **Computer-Supported Visualization Analytics (CSVA)**. Ofrece resultados visuales y gráficos relacionados con el comportamiento del individuo en una actividad de aprendizaje.

Como consecuencia, La investigación se concentra en predecir y dirigir el progreso del aprendizaje de los estudiantes individuales bajo la premisa de que el sistema institucional de gestión del aprendizaje (LMS) está desplegado. El apoyo es facilitado mediante recomendaciones de contenidos o de actividades para el aprendizaje. También se ha investigado en la mejora del aprendizaje en entornos colaborativos y de educación a distancia utilizando técnicas de minería de datos (Joshi et al., 2014) [15]. Por el contrario, *Community Learning Analytics (CLA)* comprende la identificación, análisis, visualización y apoyo de procesos de aprendizaje informales regulados por la comunidad. Las comunidades profesionales de práctica, en inglés *Communities of Practice (CoP)* están aprendiendo de manera informal. CoP son grupos de personas que comparten una preocupación o pasión por algo que hacen y que interactúan regularmente para aprender cómo hacerlo mejor (Wenger E., 1998) [8]. En CoP el uso de sistemas de gestión de aprendizaje o capacitación formal es insignificante. Sin embargo, se utiliza un ecosistema de herramientas y recursos de aprendizaje que se denomina sistema de información comunitaria, *Community Information System (CIS)*. La formación o el aprendizaje suelen tener lugar en el trabajo o en entornos informales como puedan ser desplazamientos por motivos de trabajo. LA tradicional a menudo no incluye aspectos de aprendizaje móvil como la consideración del contexto. Además, los números en una comunidad profesional son a menudo demasiado pequeños para sacar conclusiones estadísticas o para aplicar algoritmos de minería de datos educativos de forma estandarizada. Para LA más profundo, se necesita **conocimiento del dominio** para comprender la práctica, generalmente esto va acompañado de una pasión por la comunidad en estudio y un sentido de pertenencia. Por lo general, los miembros expertos de las comunidades asumen la responsabilidad informal de aprendices ya que tienen suficiente conocimiento del dominio sobre la práctica.

2.2. Análisis de Sentimientos y Emociones

Análisis de sentimientos y de emociones están fuertemente relacionados, existiendo cierto número de publicaciones e investigaciones que abarcan ambos aspectos. Se trata de extraer los sentimientos y emociones del usuario. Sin embargo, son dos conceptos diferenciados, por lo cual más adelante abordaremos cada uno de ellos en un apartado diferente. Son de los campos de más interés del Procesamiento del Lenguaje Natural, *Natural Language Processing* (NLP)⁹.

El estilo de vida ha sido remodelado en las dos últimas décadas en multitud de ámbitos a causa de la explosión tecnológica que se ha venido produciendo. La forma en que nos comunicamos, nos relacionamos, consumimos entretenimiento e incluso el modo de formarnos, han sufrido cambios impresionantes. Disponemos de una capacidad enorme de acceso a la información gracias a los motores de búsqueda y de una capacidad también notable de generar contenido en las redes sociales o en los comentarios de los medios digitales

Esta ingente cantidad de información ha constituido importantes oportunidades por un lado, a la vez que también ha planteado enormes desafíos. Los comentarios de los clientes que han comprado o consumido un producto, las críticas de películas, las publicaciones en aplicaciones de microblogging (básicamente Twitter) y el contenido de las noticias (Samuels et al., 2020) [9], han sido objeto de diferentes estudios y artículos en el ámbito del análisis de sentimientos y emociones. Sin embargo, el contexto cambia drásticamente cuando nos referimos al análisis de sentimientos dentro del ámbito de una comunidad educativa. El lenguaje suele ser más neutro y correcto y no es lo más habitual que se utilicen los foros educativos para mostrar sentimientos fuertemente polarizados. Aun así, con las limitaciones que imponen las condiciones de contorno del problema, no es imposible poder realizar análisis de sentimientos en el ámbito educativo. Estos análisis pueden ser útiles para detectar situaciones de frustración o estrés e intentar mitigar el fracaso y el abandono académico.

⁹ https://en.wikipedia.org/wiki/Natural_language_processing

2.2.1. Fundamentos del análisis de sentimientos y emociones

En este apartado se abordan los fundamentos teóricos en los que se apoyan los análisis de sentimientos y de emociones y que pueden ser aplicados en ambos.

2.2.1.1. Aprendizaje Automático

Los algoritmos de aprendizaje automático se clasifican en dos grandes grupos: **aprendizaje supervisado**, cuando se dispone de un conjunto de datos (*dataset*) etiquetados y **no supervisado**, cuando no se dispone del mismo (Borrajo et al., 2006) [32].

En el aprendizaje supervisado se puede entrenar un algoritmo con un subconjunto de los datos etiquetados, conjunto de entrenamiento (*train-set*) y realizar la validación con el resto de datos, datos de prueba (*test-set*). Puesto que se conoce la etiqueta con anterioridad, se puede comparar esta con el resultado del algoritmo, comprobando la bondad del algoritmo y el entrenamiento del mismo. Por ejemplo, un texto etiquetado como positivo debería ser clasificado como positivo.

En el aprendizaje no supervisado se busca resolver una tarea sin información previa. Por ejemplo, es útil para tareas de agrupación (*clustering*). Si disponemos de un conjunto de datos no etiquetados como pertenecientes a una categoría o clase concreta, pero disponemos de los valores de sus atributos, se pueden aplicar algoritmos de *clustering* para intentar encontrar las clases que hay de ejemplares. El aprendizaje no supervisado a menudo requiere de ensayo y error y de múltiples iteraciones. Por ejemplo, si no sabemos cuántas clases de ejemplares hay, puede ser necesario realizar varias pruebas presuponiendo que hay diferentes números de clases hasta encontrar el número de grupos óptimo. Dos de los tipos de algoritmos más populares para el agrupamiento son **k-medias** (*k-means*) (Borrajo et al., 2006) [32] y **agrupamiento jerárquico**.

El agrupamiento jerárquico construye una jerarquía de varios niveles de clústeres mediante una estructura de árbol.

A continuación se introducen algunos de los algoritmos de aprendizaje automático más utilizados en el análisis de sentimientos y emociones.

Agrupación K-medias

Se trata de una técnica de agrupación basada en optimización iterativa. Al algoritmo se le proporciona como entrada el número de clústeres, el conjunto de ejemplos (datos) y una medida de similitud. El algoritmo k-medias asigna los ejemplos a k clústeres distintos según la distancia a los centroides de cada clúster. Los primeros centroides se determinan de manera aleatoria y en sucesivas iteraciones del algoritmo se van recalculando las coordenadas de los nuevos centroides. (Borrajó et al., 2006) [32].

Máquinas de Vectores Soporte, Support Vector Machines (SVM)

Introducidas en los años 90 por Vapnik y sus colaboradores Boser et al. (1992) (Carmona E.J., 2013) [14]. Consisten en resolver un problema de clasificación binaria o de regresión, trasladando el problema original (primal) a un espacio de dimensión mayor (por lo general bastante mayor), denominado espacio de características, aplicando una función de transformación denominada función Kernel. En este nuevo espacio se puede calcular un hiperplano que soluciona el problema denominado dual. La solución en el problema dual puede trasladarse al primal. Las máquinas de vectores soporte se han utilizado con éxito en problemas de análisis de sentimiento (Kawathekar S., 2012) [22].

Naïve Bayes

Se trata de un clasificador probabilístico muy utilizado cuando se dispone de conjuntos de entrenamiento menores. Se basa en el teorema de Bayes, que consiste en que la probabilidad condicionada de que un evento X ocurra, dada la evidencia Y está determinada por la regla de Bayes (Borrajó et al., 2006) [32] , (Gamallo et al., 2014) [17]:

$$P(X/Y) = P(X) P(Y/X) / P(Y)$$

K-NN

El método K-vecinos más cercanos, *K-Nearest Neighbors (K-NN)*, es un algoritmo no paramétrico y pertenece a la denominada familia de técnicas de aprendizaje vago (*lazy learner*). Son técnicas que parten de un esquema sencillo y crean estructuras abstractas de generalización de los ejemplos. K-NN se basa en la clasificación por cercanía en un espacio vectorial. Una variante es K-NN ponderado, donde se proporcionan pesos a los elementos del conjunto de entrenamiento. En cualquier caso se debe seleccionar el valor de k, calcular la distancia (generalmente la distancia euclídea) a los k ejemplos más cercanos y asignar el ejemplar a la clase cuyo número de vecinos dentro de los k es máximo.

Redes Neuronales

Las redes neuronales están inspiradas en el funcionamiento del sistema nervioso biológico (Borrajó et al., 2006) [32].

Aun cuando aún se desconoce mucho acerca del funcionamiento real del proceso cognitivo y de aprendizaje de los seres vivos, se sabe que las neuronas (células del sistema nervioso) están especializadas en la recepción de estímulos y conducción del impulso nervioso entre ellas mediante conexiones llamadas sinapsis, o con otros tipos de células, como las de las fibras musculares¹⁰. Constan además de un núcleo (como todas las células), de un axón y numerosas dendritas. Las neuronas recogen información electroquímica procedente de las neuronas vecinas a través de las dendritas y en función de un determinado código de activación propio de cada neurona, transmiten información a través del axón a las dendritas de otras neuronas o a células denominadas efectoras (figura 1), que forman parte de otros sistemas que no son el nervioso (músculos y glándulas).

¹⁰ URL: <https://es.wikipedia.org/wiki/Neurona>

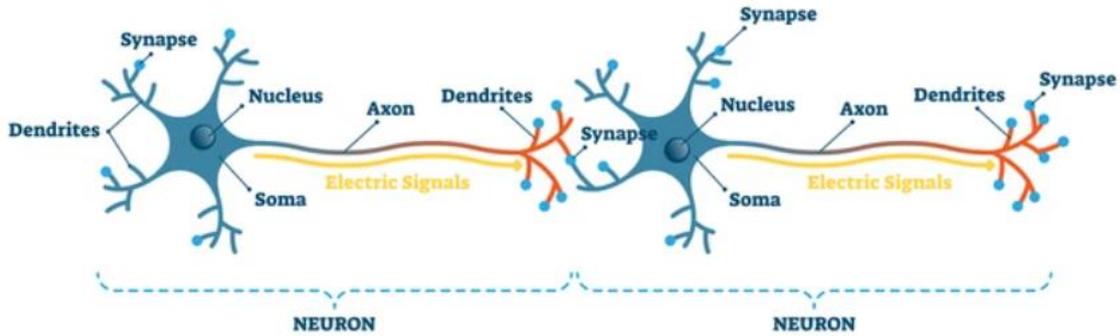


Figura 1. Interconexión entre neuronas. Fuente: <https://www.shutterstock.com/>

Las redes neuronales artificiales son consideradas máquinas conexionistas (Borrajó et al., 2006) [32]. A diferencia de los programas de ordenador convencionales, procesan la información de entrada obteniendo una salida como respuesta (estímulo-acción), no aplicándose un algoritmo de manera ciega. La respuesta puede depender de diferentes factores estructurales o funcionales, pudiendo la red “aprender” en función de los resultados obtenidos mediante un mecanismo de retroalimentación.

Uno de los primeros modelos es el conocido como perceptrón, (Borrajó et al., 2006) [32], ideado por Rosenblatt F., 1962 [33]. Se le considera el modelo matemático más simple de una neurona. En la figura 2 puede verse el esquema de un perceptrón de cinco señales de entrada.

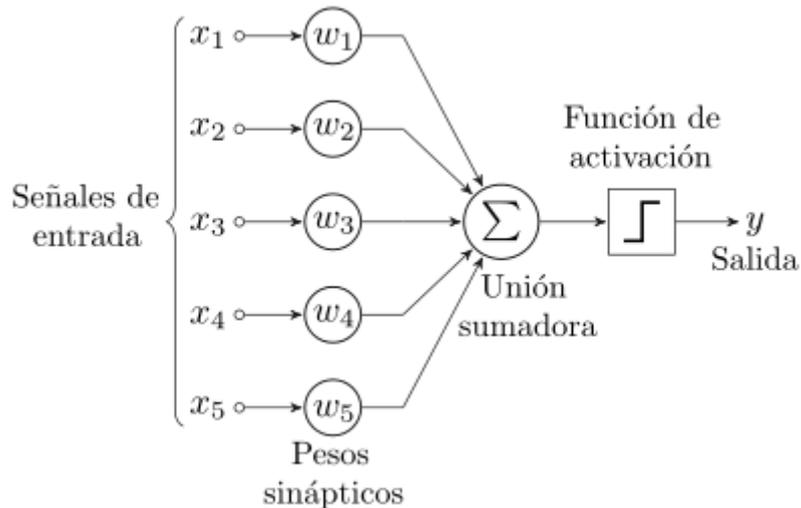


Figura 2. Perceptrón con cinco señales de entrada. Fuente: <https://es.wikipedia.org/wiki/Perceptrón>

Se dispone de varias entradas, cada una de las cuales tiene asociado un peso y existe una unidad lógica de cómputo, que elabora una salida a partir de lo que le llega de cada una de las entradas (unión sumadora). La salida de la unión sumadora es ingresada a una función de activación (comúnmente sigmoide), que es la que entrega la salida del proceso. El aprendizaje se produce cambiando los pesos de las neuronas, de la siguiente manera:

$$w(j)' = w(j) + \alpha(\delta - y) x(j)$$

Donde:

- $x(j)$ es el elemento en la posición j del vector de entrada ($x_1, x_2, x_3, \dots, x_n$).
- $w(j)$ es el elemento en la posición j del vector de pesos ($w_1, w_2, w_3, \dots, w_n$).
- y es la salida de la neurona.
- δ es la salida esperada.
- α es una constante tal que $0 < \alpha < 1$ y sólo aparece en el caso de utilizarse lo que se llama “tasa de aprendizaje”, de lo contrario se hace igual a la unidad, omitiéndose este parámetro.

Existen diferentes tipos de redes neuronales artificiales que se han ido desarrollando a partir de la idea del perceptrón elemental y otros trabajos posteriores. En general, las redes neuronales artificiales consisten asociar varias capas o niveles de neuronas, aunque también existen otros modelos

2.2.1.2. *Procesamiento del Lenguaje Natural*

Conocido comúnmente de forma abreviada como NLP¹¹ de sus siglas en inglés, *Natural Language Processing*, consiste en el estudio de los procesos de interacción entre el lenguaje humano y un ordenador. Trata aspectos como el análisis sintáctico y semántico de las oraciones, la traducción automática. Se considera que de 1950 a principios de los 1990 se desarrolló la NLP simbólica, de los 1990 a los 2010 se produce una revolución, introduciendo el aprendizaje automático y desarrollando el llamado NLP estadístico.

¹¹ https://en.wikipedia.org/wiki/Natural_language_processing

Desde la década de 2010 a la actualidad se ha profundizado más en la NLP neuronal, gracias al desarrollo de las llamadas redes neuronales profundas y el aprendizaje de representación o de características.¹²

2.2.2. Análisis de Sentimientos

El análisis de sentimientos, en inglés, *Sentiment Analysis (SA)*, es el proceso de detección de la polaridad contextual de un texto (Devica et al., 2016) [13]. Determina si el texto dado es positivo, negativo o neutro. También se denomina minería de opiniones, ya que deriva la opinión o actitud de quien se expresa. El SA, visto como un proceso de aprendizaje automático, requiere el uso de un conjunto de entrenamiento cuya calidad es importante de cara a los resultados. El análisis semántico de la oración también aumenta el significado y la precisión del resultado. El etiquetado POS es útil cuando se requiere inferir si un texto (comentario) corresponde al tema relevante buscado. El etiquetado POS es la tarea de etiquetar cada palabra en una oración con su parte gramatical apropiada (sustantivos, verbos, adverbios, adjetivos, pronombres, conjunciones, etc.). La mayor parte del etiquetado de POS se incluye en el etiquetado de POS basado en reglas, el etiquetado de POS estocástico y el etiquetado basado en transformación (Devica et al., 2016) [13].

En general, el análisis de sentimientos se ha investigado principalmente a tres niveles (Joshi et al., 2014) [15].

- **A nivel de documento:** Consiste en clasificar si un documento de opinión, en su conjunto, expresa un sentimiento positivo o negativo, asumiendo la simplificación de que cada documento versa sobre una entidad única.
- **A nivel de oración:** La tarea principal es verificar si cada oración expresa una opinión positiva, negativa o neutral. Este nivel de análisis está íntimamente relacionado con clasificación de subjetividad.
- **A nivel de aspecto:** Se realiza un análisis más detallado. En lugar de mirar construcciones lingüísticas (documentos, párrafos, oraciones, cláusulas o frases), el nivel de aspecto mira directamente la opinión en sí misma.

¹² URL: https://en.wikipedia.org/wiki/Natural_language_processing

2.2.2.1. *Los diferentes enfoques del análisis de sentimientos*

Existen diferentes enfoques a la hora de abordar el análisis de sentimientos. Se trata de diferentes maneras de considerar el problema, desde diferentes disciplinas o planteamientos. A continuación se hace un repaso por los más importantes. Es importante destacar que algunas soluciones pueden pertenecer a más de un enfoque, no tratándose necesariamente de una clasificación mutuamente excluyente.

Enfoque algebraico

Se basa en la suma algebraica de las orientaciones de los términos del texto; en el que cada documento es representado como una “bolsa de palabras”, vector de características $x \in R^{|V|}$, donde las entradas de x son las frecuencias con que las palabras del conjunto vocabulario V aparecen en el documento objeto de estudio. Se busca aprender un vector $x \in R^{|V|}$ tal que el clasificador $orient = signo(c^T x)$ estime con precisión la orientación del documento x . Devolviendo +1 o -1) para documentos que expresan sentimiento positivo o negativo respectivamente sobre el tema de interés (Colbaugh et al., 2010) [11]. A partir de este planteamiento evolucionan diferentes enfoques alternativos.

Enfoque Machine Learning (ML) o Aprendizaje Automático

Plantea la tarea del SA como una tarea de aprendizaje automático. A continuación se detalla un poco más este enfoque, particularizando algunas técnicas vistas en 2.1.1.1

Naïve *Bayes*

El teorema de Bayes, visto en 2.1.1.1:

$$P(X/Y) = P(X) P(Y/X) / P(Y)$$

Puede extrapolarse al análisis de sentimientos como:

$$P(\text{Sentimiento} | \text{Oración}) = \frac{P(\text{Sentimiento}) P(\text{Oración} | \text{Sentimiento})}{P(\text{Oración})}$$

El adjetivo de Naïve (ingenuo) proviene de la asunción de independencia entre los predictores, lo cual no es cierto, pues hay palabras que aparecen juntas más frecuentemente. Sin embargo, realizar dicha asunción simplifica el problema y se siguen obteniendo resultados muy positivos en la práctica, permitiendo considerar $P(\text{Oración} / \text{Sentimiento})$ como el producto de $P(\text{token} / \text{sentimiento})$ a través de todos los tokens de la oración.

Supongamos que tenemos un conjunto de oraciones positivas y negativas formadas por un conjunto de tokens (palabras) que se repiten en la construcción de las oraciones.

La probabilidad a priori de que aparezca un token t cogiendo una frase al azar sería:

$$P(t) = \frac{\text{Cont}(t)}{\text{Cont}(\text{TodosTokens})}$$

Mientras la probabilidad de que aparezca un sentimiento s sería:

$$P(s) = \frac{\text{Cont}(s)}{\text{Cont}(\text{TodosSentimientos})}^{13}$$

Para estimar $P(\text{token} / \text{sentimiento})$ con carácter general, la probabilidad condicionada vendría dada por:

$$P(t | s) = \frac{P(t \cap s)}{P(s)}$$

Donde $P(t | s)$ representa la probabilidad de aparición de una determinada palabra en una oración, dada la pertenencia de dicha oración a una determinada clase de sentimiento. Por ejemplo, la aparición de la palabra “perfecto”, entre otras, en la oración del conjunto de entrenamiento “*Me parece perfecto*”, estando etiquetada dicha oración como positiva. De donde se puede expresar como:

¹³ Típicamente dos sentimientos o clases (positivo y negativo).

$$P(T = t | S = s) = \frac{Cont(T = t, S = s)}{\sum_{i=1}^N Cont(T = t_i, S = s)}$$

Donde:

- $P(T = t | S = s)$ representa a $P(t | s)$.
- $Cont(T = t, S = s)$ representa el número de ocurrencias del token t en oraciones de una determinada clase. Siguiendo con el ejemplo, el número de veces que aparece la palabra “perfecto” en oraciones positivas.
- $\sum_{i=1}^N Cont(T = t_i, S = s)$ representa el número total de ocurrencias de todos los tokens en el conjunto de oraciones de un determinado sentimiento.

O de manera más literal e intuitiva:

$$P(t | s) = \frac{Cont(tokenEnClase)}{Cont(TodosTokensEnClase)}$$

Sin embargo, suele aplicarse el Suavizado de Laplace, *Laplace Smoothing*, que consiste en evitar la probabilidad nula causada por la no existencia de ciertos valores de atributos (palabras) en el conjunto de entrenamiento. Se supone que además de los ejemplos del conjunto de entrenamiento hay k ejemplos “virtuales” en cada clase C por cada posible valor de cada atributo A . (Pang-Ning et al., 2012) [19]

$$P(t|s) = \frac{n(t, s) + k}{|V| + N(s)}$$

Donde $n(t, s)$ es el número de veces que la palabra t aparece en los documentos de entrenamiento etiquetados como de sentimiento s ; $N(s)$ es el número total de palabras en los documentos de entrenamiento que están etiquetadas como de sentimiento s ; $|V|$ es el número de tokens distintos (palabras) y k es un parámetro prefijado. Habitualmente $k = 1$, de donde viene la denominación de “*Add-1 (Laplace) smoothing*”. Con todo, podemos concluir la siguiente expresión:

$$P(token | sentimiento) = \frac{Cont(tokenEnClase) + 1}{Cont(TodosTokensEnClase) + Cont(TodosTokens)}$$

K-NN

El método K-vecinos más cercanos, visto en 2.1.1.1 se aplica al análisis de sentimientos.

Un algoritmo interesante es *two-parse* o dos análisis. Por ejemplo, en Ankita Strivastava et al.) [20], se presenta una propuesta que ofrece buenos resultados para casos en los que los *training datasets* (conjuntos de entrenamiento) tienen un buen equilibrio entre fuertemente, medianamente y débilmente polarizados, en conjuntos de entrenamiento llegándose a alcanzar eficiencias comparables a algoritmos más complejos con conjuntos de entrenamiento significativamente más grandes. La propuesta de A. Strivastava et al. consiste en la utilización de un algoritmo two-parse sobre una modificación de K-NN. Se realizan dos análisis, siendo el resultado del primer análisis la entrada del segundo, incidiendo en este segundo análisis sobre los resultados neutros.

En el segundo análisis se aplica la puntuación (*score*) de la siguiente manera cuando existe coincidencia de alguna de las palabras del texto:

$$Puntuación\ Positiva = \frac{\sum_{i=1}^{i=j} puntuación(palabras\ pos) + \sum_{i=1}^k puntuación(palabras\ neg)}{\sum_{i=1}^{i=j+k} Puntuación\ extrema}$$

En caso de no haber coincidencia de palabras con ninguno de los conjuntos del corpus (positivo ni negativo) en el primer análisis, se asigna una puntuación positiva del 50%.

CNN-LSTM

Dentro del ámbito del aprendizaje automático o la Inteligencia Artificial, se han explorado algoritmos que han demostrado ser eficaces en otros campos, como el reconocimiento de imágenes. Este es el caso de las Redes Neuronales Convolucionales (CNN del inglés *Convolutional Neural Networks*). En Rehman et al., 2019 [21] por ejemplo, se explora un mecanismo híbrido de CNN con LSTM (Long Short Term Memory) un tipo de redes neuronales que aprenden de las secuencias de términos)

Clasificador de Máxima Entropía

Un clasificador de máxima entropía (ME), o clasificador exponencial condicional, está parametrizado por un conjunto de pesos que se utilizan para combinar las características conjuntas que se generan a partir de un conjunto de características mediante una codificación. La codificación asigna cada par de conjunto de características y etiqueta a un vector. Los clasificadores ME pertenecen al conjunto de clasificadores conocidos como clasificadores exponenciales o logarítmicos lineales, porque funcionan extrayendo un conjunto de características de la entrada, combinando linealmente y luego usando esta suma como exponente. Si este método se realiza de manera no supervisada, se utiliza información mutua puntual *Point wise Mutual information (PMI)* para encontrar la co-ocurrencia de una palabra con otras palabras positivas o negativas. El clasificador ME es uno de los modelos que no asume la independencia de características (Tamilselvi et al., 2013) [18]. La incertidumbre es máxima para una distribución uniforme. La medida de la incertidumbre se conoce como entropía.

N-gram Sentiment Analysis. N-gramas

Un n-grama es una subsecuencia de n elementos de una secuencia dada;¹⁴ en este contexto, una secuencia de texto o discurso. Los ítems pueden ser fonemas, sílabas, letras, palabras o pares de bases según la aplicación. Los n-gramas normalmente se recopilan de un corpus de texto o voz. En este caso están considerando la oración como un todo (Šaloun et al., 2013) [16]. Están haciendo uso de cuatro tipos de léxicos, a saber, léxico de frase de sentimiento, léxico de fuerza de sentimiento, léxico con aspectos y léxicos de excepción.

¹⁴ URL: <https://es.wikipedia.org/wiki/N-grama>

Enfoque basado en reglas

Se trata de un aprendizaje supervisado en el que se pueden ir añadiendo palabras y reglas a la base de datos a medida que se van mostrando útiles o necesarias para el análisis. Se definen varias reglas para inferir una opinión. Se *tokeniza* cada oración del documento y se comprueba tanto la presencia de los tokens como el cumplimiento de las reglas (Hutto, C., & Gilbert, E., 2014) [23].

Enfoque basado en el léxico

Estas técnicas trabajan con la asunción de que la polaridad global de un documento (u oración) es la suma de las polaridades individuales de las frases (o palabras). En Klekovkina M and Kotelnikov E.V., 2012 [24] se presenta un método de clasificación de críticas de películas por sentimientos basado en un diccionario y se describe el proceso de creación del mismo. Utilizan el método RF (Man Lan et al., 2009) [25] para calcular el peso de un término, el cual tiene en cuenta la información sobre la distribución del mismo en los textos del *dataset* de entrenamiento, así como la pertenencia de los textos a las clases. El peso del término i -ésimo para la clase C se expresa por la fórmula:

$$RF_i^C = \log_2 \left[2 + \frac{a}{\max(1, b)} \right]$$

Donde a es el número de textos que contienen el término i -ésimo y pertenecen a la clase C y b es la cantidad de textos que contienen el término y no están relacionados con la clase C .

Los pesos de las palabras en el *dataset* se determinaron manualmente, de -5 a -1 para las palabras negativas y de +1 a +5 para las positivas. Creando un diccionario de 260 palabras.

Además se agregaron palabras modificadoras, según las cuales el peso de la palabra que vienen a continuación aumentaba o disminuía.¹⁵ Las palabras modificadoras son adverbios, adjetivos o construcciones con un porcentaje asignado de incremento o

¹⁵ En el caso de la lengua española, el adjetivo suele colocarse detrás del sustantivo, lo cual debería ser tenido en cuenta a la hora de implementar un algoritmo similar.

decremento (mucho, poco, algo, enorme, en realidad, etc.). También se tenía en cuenta la negación para invertir el peso emocional de la palabra evaluada tras la partícula de negación.

Para cada texto de la colección de entrenamiento se calcula su peso, que resulta ser el peso medio de las palabras de evaluación en él incluidas:

$$W_T = \frac{\sum_{i=1}^N W_i}{N}$$

Donde W_T es el peso del texto; W_i es el peso de la palabra i -ésima y N es el número de palabras evaluativas en el texto T .

Enfoques probabilísticos. pLSA y LDA

En Clarizia et al. (2018) [10] se aborda el enfoque **LDA (Asignación latente de Dirichlet)**, en el que cada documento puede ser visto como compuesto por una mezcla de varios temas, similar a la **semántica latente probabilística (pLSA)** (Benítez J.A., 2011) [12], con la excepción de que en LDA la distribución de temas se supone que tiene un antecedente de Dirichlet. Por el uso del enfoque LDA sobre un conjunto de documentos pertenecientes a un mismo dominio de conocimiento, se puede extraer automáticamente un gráfico mixto de términos que contenga un conjunto de pares de palabras ponderadas que resultan discriminatorias para la clasificación de sentimientos.

2.2.3. Análisis de Emociones

Si el análisis de sentimientos es una cuestión binaria en el sentido de detectar expresiones positivas o negativas, el análisis de emociones es más complejo, en tanto en cuanto se centra en cuestiones más profundas relacionadas con las emociones humanas. Se pretenden extraer más matices de las sensibilidades de quien se expresa, considerando las diferentes variaciones de las subjetividades mentales. Por tanto se considera un conjunto más extenso de estados de ánimo en lugar de dos únicas categorías; si bien se pueden clasificar como positivas o negativas. Emociones positivas pueden ser: felicidad, placer, alegría, satisfacción o calma; mientras que pueden considerarse emociones

negativas: miedo, enojo, angustia, frustración, tristeza o infelicidad, por citar algunas de las más significativas.

2.2.3.1. *Emociones básicas*

A la hora de abordar el estudio o análisis de las emociones, no podemos pasar por alto el concepto de emociones básicas.

Sobre las emociones básicas se ha escrito desde la antigüedad. Centrándonos en trabajos de la era científica moderna, Para Izard C.E., 1977 [29] las emociones básicas o discretas son cinco: felicidad, tristeza, enojo, vergüenza y miedo. Eckman y Friesen, 1971 [30] se refieren a seis emociones básicas: **felicidad, tristeza, enojo, sorpresa, asco y miedo**. El modelo de las seis emociones de Ekman ha sido ampliamente aceptado.

2.2.3.2. *Descomposición de las emociones*

Russell J.A., 1980 [26] define un modelo al que denomina *Modelo Circunflejo del Afecto*. En dicho modelo se trabaja con la premisa de que cualquier emoción podría definirse como una combinación de dos componentes: placer y tensión¹⁶, de manera que se podrían trazar unos ejes cartesianos y representar por ejemplo en el eje X el nivel de placer y en el eje Y el de tensión, quedando cada emoción representada por un punto en las coordenadas definidas por dichas componentes.

El trabajo de Russell se apoyaba en Scholberg H., 1952 [27], que introdujo un modelo descriptivo de las expresiones faciales en términos de dos dimensiones. Se definen dos ejes:

pleasantness-unpleasantness (agrado-desagrado) y *attention-rejection* (atención-rechazo).

¹⁶ Russell se refiere a los términos en inglés *pleasure* y *arousal*, que podrían traducirse como “placer” y “**excitación**” respectivamente. Por otro lado, una de las emociones que aparecen en su trabajo es la de *excitement*, cuya traducción más acertada en el contexto que nos afecta sería también la de “excitación”. En aras a una traducción que permita diferenciar entre ambos términos con la mínima pérdida de rigor respecto del trabajo original de Russell, en este trabajo se ha traducido *arousal* como “tensión” y *excitement* como “excitación”, dado que *excitement* representa una emoción de excitación en sentido positivo, mientras que *arousal* se refiere al nivel de excitación en términos generales.

Scholberg a su vez se apoyó en trabajos previos que sugerían una escala circular, como el trabajo no publicado y mencionado por él de Miss Marjory L. Brown.

En la figura 3 se muestra una ilustración original del trabajo de Scholberg [1952]

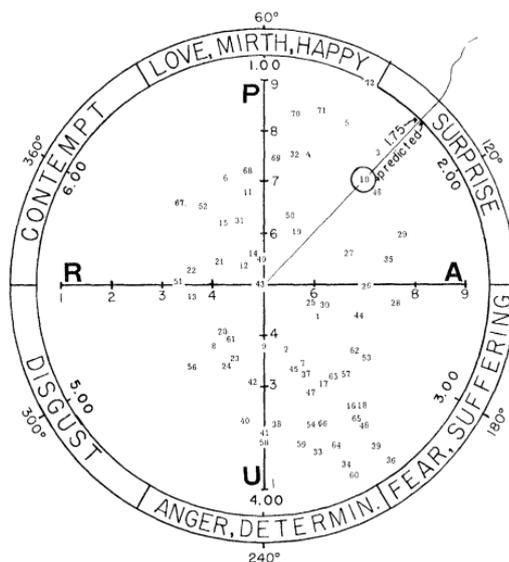


Figura 3. Ilustración original del trabajo de Scholberg, 1952. Representación en dos dimensiones (P-U y A-R) de 72 fotografías de expresiones faciales. El eje P-U representa *pleasantness-unpleasantness* (agrado-desagrado) y el eje A-R *attention-rejection* (atención-rechazo).

Scholberg H, 1954 [28] aborda las *tres dimensiones de las emociones*, donde define una representación tridimensional (cono invertido de la figura 4). En la base del cono se definían ocho regiones (sectores) que representaban ocho emociones básicas: amor-alegría, sorpresa, miedo, sufrimiento, enojo, determinación, asco, desprecio y felicidad. El desplazamiento a lo largo del eje longitudinal del cono representaba el nivel de activación desde dormido (*sleep*) hasta tensionado (*tension*). La base del cono (situada arriba) está inclinada para representar que los sectores de la ira y el miedo pueden alcanzar niveles de activación más altos que el desprecio.

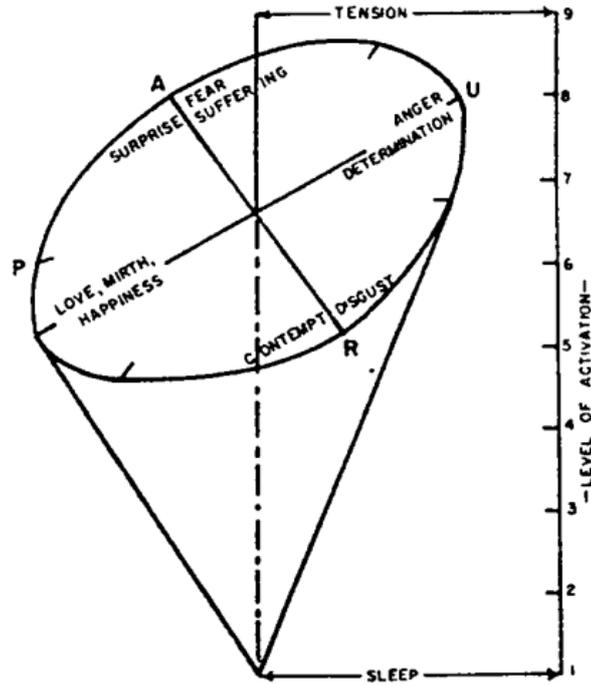


Figura 4. Ilustración original del trabajo de Scholberg, 1954. Sólido que representa las tres dimensiones de las emociones correspondientes a expresiones faciales.

A continuación, en la figura 5, se muestra una adaptación en castellano, realizada para este trabajo de las conclusiones de Russell, J.A, 1980. Se representan algunos de los resultados obtenidos en su trabajo para algunas emociones principales.

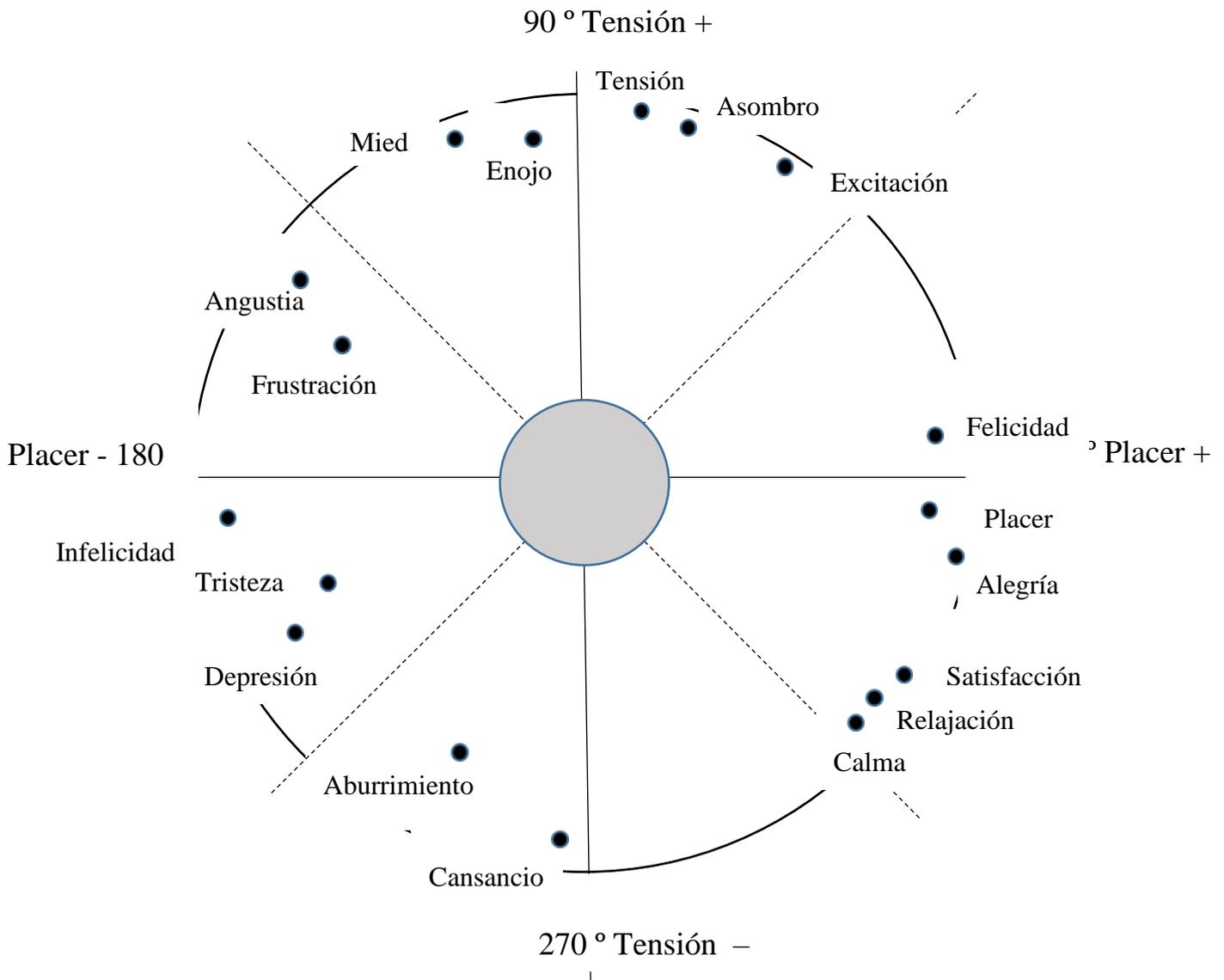


Figura 5. Adaptación para este trabajo de la representación de algunas emociones en dos dimensiones (Russell).

Posner et al. 2005 [31] traslada el modelo circunflejo de las emociones gestuales de Russell, de la psicología al terreno de la neurociencia, presentado un modelo circunflejo simplificado. En este modelo, también se llama al eje de tensión (*arousal*) eje de activación (*activation*) y al eje de placer se le denomina también valencia (*valence*). En este nuevo trabajo, 25 años más tarde, se hace una representación gráfica más simplista de las emociones, situando a todas en el perímetro de un óvalo de manera equidistante entre ellas (Óvalo de Posner de las emociones). Ver la figura 6.



Figura 6. Óvalo de Posner de las emociones

2.2.3.3. *Los diferentes enfoques del análisis de emociones*

Se presentan a continuación, como se ha hecho para la detección de sentimientos, diferentes enfoques abordados para la detección de emociones.

En general, gran parte de lo indicado en el apartado 2.1.1 puede extenderse al análisis de emociones en lo que se refiere a enfoques, salvo por la particularidad de que al estar ante un problema cuya solución es no binaria, algunos métodos o algoritmos no pueden aplicarse directamente, si bien, como hemos visto anteriormente, que existen métodos para descomponer las emociones en dos o tres componentes, transformando el problema de detectar una emoción concreta en el de hallar dichas componentes para un texto de entrada dado.

El enfoque algebraico

De manera análoga a lo explicado para los sentimientos, pueden considerarse los textos como una bolsa de palabras. Disponiendo de un conjunto de palabras (corpus) etiquetadas con una emoción asociada o unas puntuaciones para cada una de ellas, pueden simplemente buscarse las palabras del texto en el corpus e ir sumando las puntuaciones obtenidas para cada emoción. En su planteamiento más simple, este método no tiene en cuenta los análisis sintáctico y semántico, así como las negaciones. Para cada texto se obtiene una puntuación para cada una de las emociones consideradas como básicas (ver 2.1.2.1). La emoción que mayor puntuación obtenga será la que representa al texto en su conjunto.

El enfoque Machine Learning (ML) o Aprendizaje Automático

Los algoritmos explicados o mencionados en el [apartado 2.2.2.1](#), sobre análisis de Sentimientos desde un enfoque de tarea de aprendizaje automático, son igualmente aplicables al análisis de las emociones.

2.3. Análisis de Redes Sociales

El análisis de las redes sociales o en inglés, Social Network Analysis (SNA)¹⁷ es el proceso de investigación de estructuras sociales mediante el uso de redes y teoría de grafos (Otte et al., 2002) [35]. Caracteriza las estructuras en red en términos de nodos (actores individuales) y las aristas, (relaciones o interacciones) que los conectan y que en el ámbito de este TFM son definidas por los mensajes que se envían unos usuarios a otros.

¹⁷ https://en.wikipedia.org/wiki/Social_network_analysis

2.4. La explosión de la IA

Recientemente, hemos asistido a la aparición en escena de soluciones muy potentes dentro del ámbito de la Inteligencia Artificial (IA) que se ofrecen al gran público para el análisis de datos y abordaje de un gran número de problemas. Gracias al incremento de la capacidad de almacenamiento y procesamiento, así como a la ingente cantidad de información disponible para el entrenamiento de algoritmos, unido al refinamiento y mejora de los algoritmos mismos, las herramientas de IA han pasado a formar parte de la cotidianeidad de un gran número de personas. Herramientas como el chatbot ChatGPT¹⁸, de OpenAI han irrumpido en el mercado con fuerza.

En este contexto, soluciones de IA se ofrecen a empresas y particulares a través de la web para gran número de aplicaciones. Por ejemplo, para el análisis cualitativo de la información, existen herramientas como Atlas.ti¹⁹, que pueden utilizarse para el análisis de foros²⁰.

¹⁸ <https://es.wikipedia.org/wiki/ChatGPT>

¹⁹ <https://atlasti.com/es>

²⁰

https://www.researchgate.net/publication/259557742_Analisis_de_foros_virtuales_con_empleo_del_programa_Atlasti

Capítulo 3

Trabajo previo

A continuación, se explica someramente el trabajo previo que ha servido de punto de partida para el desarrollo del trabajo actual. Trabajo realizado por Andrea Rey Presas, bajo la dirección del Dr. Antonio Rodríguez Anaya y el Dr. Jesús González Boticario en febrero de 2021. Se hace un análisis atendiendo a diferentes aspectos, como son: interfaz, ayuda, dinámica de funcionamiento o aspectos no cubiertos (oportunidades de mejora).

3.1. Interfaz

En la figura 7 se puede ver la interfaz de la página principal de la aplicación del trabajo anterior. Cabe mencionar que se trata de una interfaz muy sencilla con tan solo dos botones que dan acceso al *análisis de sentimientos* o al *análisis de la red* respectivamente.

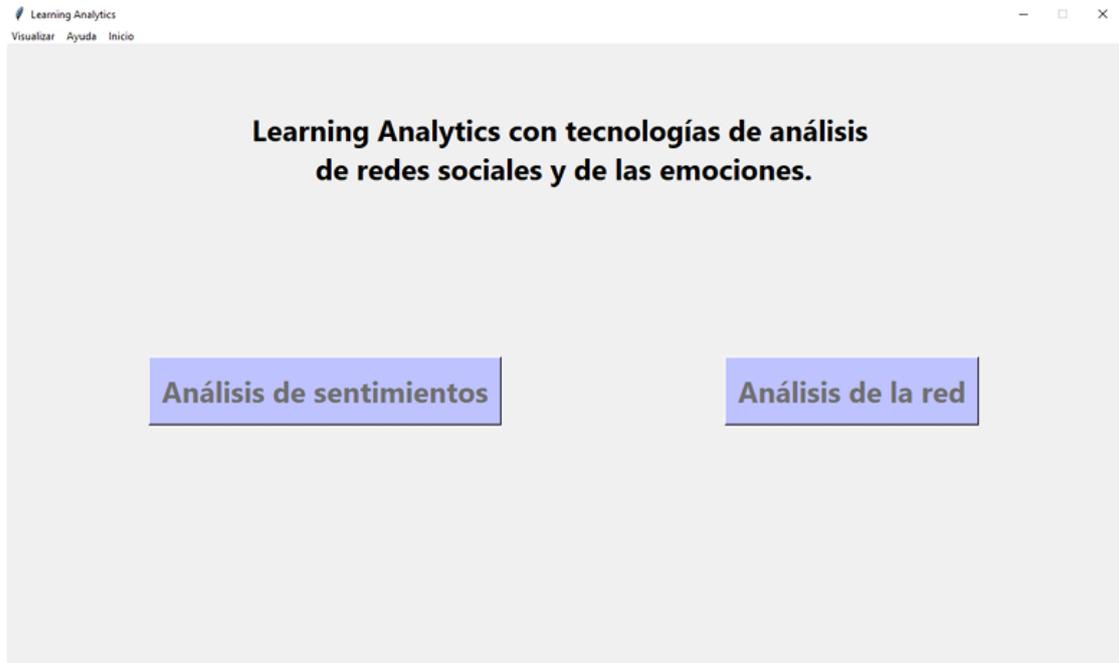


Figura 7. Página principal de la aplicación del trabajo previo.

Una vez se pulsa el botón de Análisis de sentimientos se pasa a la pantalla de la figura 8.



Figura 8. Pantalla de análisis de sentimientos de la aplicación andel trabajo anterior.

El resultado del análisis de sentimientos se presenta siguiendo el diseño mostrado en la figura 9.



Figura 9. Página de resultados de sentimientos de la aplicación del trabajo anterior.

Respecto del análisis de redes, la interfaz se muestra en la figura 10.

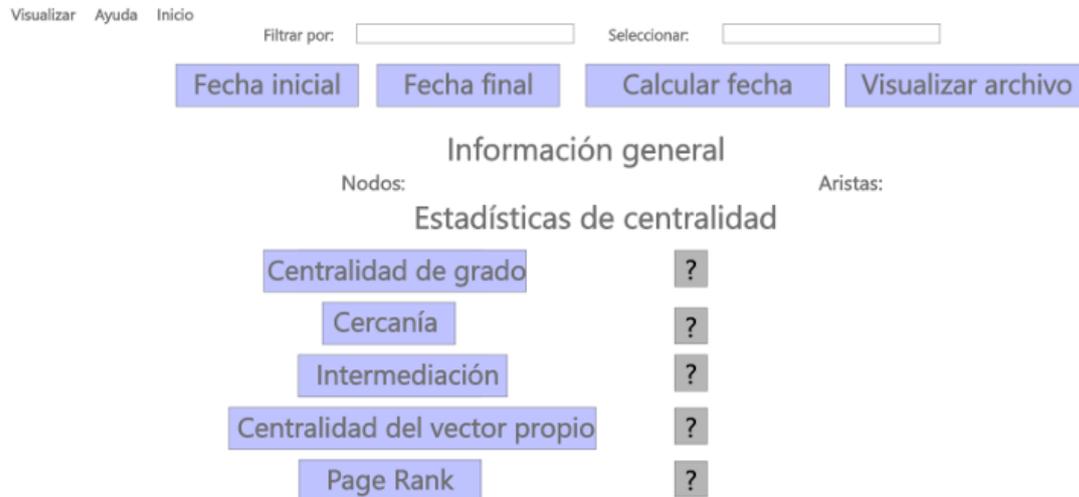


Figura 10. Interfaz del análisis de redes de la aplicación del trabajo anterior.

Teniendo el resultado del análisis de red el aspecto mostrado en la figura 11. En

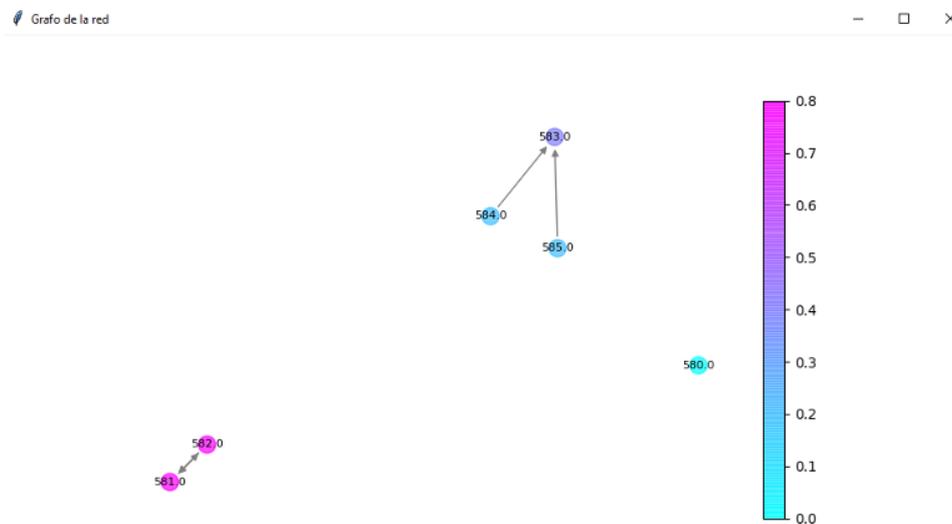


Figura 11. Análisis de red (Estadístico grado) de la aplicación del trabajo anterior.

Es posible visualizar los valores de los diferentes estadísticos (medidas de centralidad) para los nodos en forma de tabla. Ver figura 12.

NODO	GRADO	GRADO_ENTRADA	GRADO_SALIDA	CENTRALIDAD_G	CERCANIA	INTERMEDIAION	VECTOR_PROPIO	PAGE_RANK	INIATIVA	MOOLARIDAD
1	582.00	0	0	0	0	0	8.4e-08	0.053	1	0
2	582.00	4	1	3	0.8	0.2	0.067	0.71	0.35	1
3	581.00	4	3	1	0.8	0.2	0.067	0.71	0.35	0
4	583.00	2	2	0	0.4	0.4	0.067	4e-06	0.14	2
5	584.00	1	0	1	0.2	0	0.033	8.4e-08	0.053	0
8	585.00	1	0	1	0.2	0	0.033	8.4e-08	0.053	0

Figura 12. Valores estadísticos de los nodos

Respecto del análisis de sentimientos y emociones se muestra un gráfico de tarta con los porcentajes de mensajes positivos y negativos y un ranking de usuarios por mensajes positivos y negativos. Si elegimos analizar las emociones se muestra un gráfico temporal en el que se representa por cada mes la proporción de mensajes con cada emoción identificada. Ver figura 13.

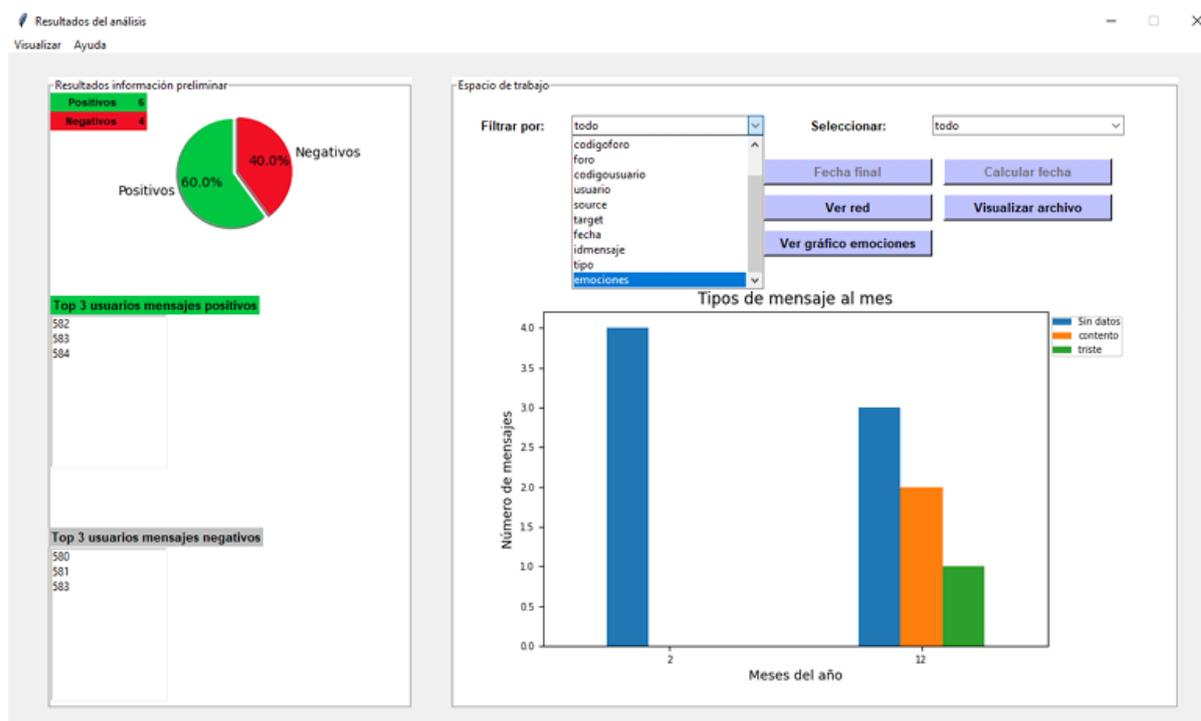


Figura 13. Análisis de sentimientos y emociones de la aplicación anterior.

La aplicación también permite realizar la operación de *clustering* (agrupamiento de usuarios), representándose en los ejes X e Y respectivamente las variables *Cercanía* y *Grado de Salida*. Ver figura 14.

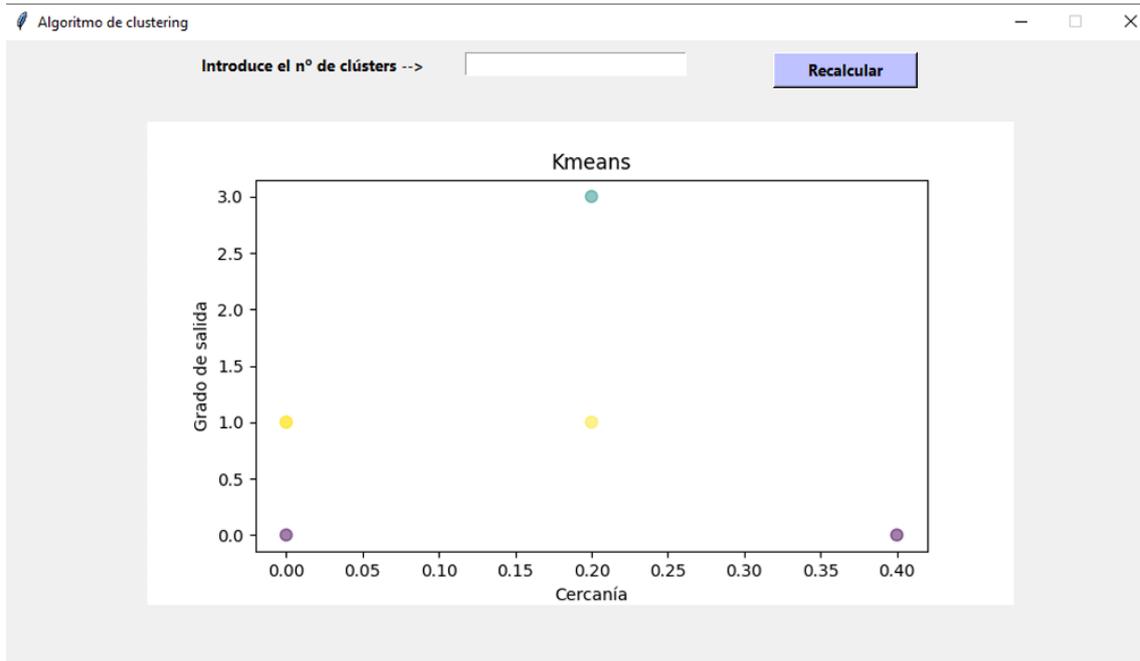


Figura 14. Algoritmo de *clustering* de la aplicación anterior.

Finalmente cabe destacar el análisis de la red atendiendo a sentimientos (positivos o negativos) que se muestra en la figura 15.

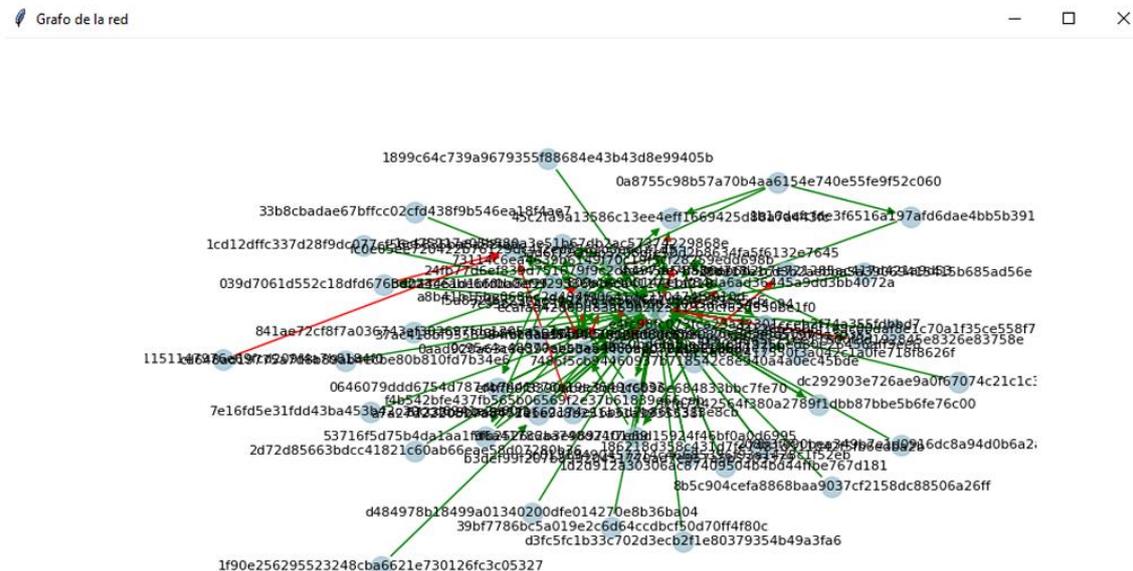


Figura 15. Análisis de red de sentimientos de la aplicación anterior.

3.2. Ayuda

La ayuda que ofrece la aplicación consiste en:

- Mostrar el formato requerido para los archivos de datos y corpus de sentimientos.



Figura 16. Ayuda contextual sobre el formato de los archivos.

3.3. Dinámica de Funcionamiento

La aplicación se compone de varias ventanas que se van abriendo de manera independiente en lo que a visualización se refiere. Por ejemplo, cuando elegimos ver los datos de los foros se nos abre una nueva ventana. Lo mismo sucede cuando se dibuja un gráfico de red o de clustering. Esto se ha considerado un aspecto a mejorar, constituyendo uno de los objetivos la creación de una aplicación que se muestre como una integración de funcionalidades dentro de una única ventana principal de la aplicación, utilizando la creación de nuevas ventanas sólo en casos concretos que se verán más adelante en la sección correspondiente a la nueva aplicación.

3.4. Aspectos no cubiertos (oportunidades de mejora)

3.4.1. Persistencia de resultados

Un aspecto no cubierto en el trabajo anterior que se puede considerar deseable es la posibilidad de crear y guardar proyectos para poder abrirlos con posterioridad.

3.4.2. Generación de informes

Se ha considerado un ejercicio interesante tanto a nivel de elemento constitutivo del trabajo actual, en términos evaluables, como a nivel práctico, la posibilidad de generar informes de manera automática por la aplicación.

3.4.3. Anonimización de datos

Se ha planteado también como una tarea interesante la posibilidad de que la aplicación cree una tabla en la que haga corresponder a cada identificador de usuario un identificador alternativo o alias. Esto cumple un doble propósito: por un lado se puede anonimizar la información que se muestra en los gráficos etiquetados y por otro lado se pueden acortar los identificadores para mejorar la legibilidad e interpretabilidad de los mismos. Basta observar la figura 15 en la que la longitud de los identificadores prácticamente imposibilita dicha interpretabilidad.

Capítulo 4

Trabajo realizado

La parte práctica de este TFM ha consistido en la realización de una aplicación de escritorio desarrollada en Python.

Se ha partido como se ha explicado con anterioridad del trabajo previamente realizado como base y se ha realizado una profunda revisión tanto estética como funcional para obtener un resultado que consiga por un lado cumplir con el principio de aprovechamiento y reutilización que se pedía del trabajo previo, pero por otro lado se consiga una aplicación original y funcional que realice tareas comparables a las de otras herramientas orientadas a diferentes finalidades, como puede ser el análisis de redes sociales (Gephi)²¹ o el análisis de sentimientos (Linguakit)²² por citar algunas. Además, se ha realizado una aplicación con considerable flexibilidad en lo que se refiere al formato de los datos de entrada (admite diferentes formatos de archivo y no requiere unas cabeceras de columnas específicas, sino que son igualmente flexibles) y con la funcionalidad de generar informes automáticos en PDF.

El trabajo práctico se ha apoyado tanto en el TFM anteriormente mencionado, como en la aplicación mencionada anteriormente ([Álvaro Rodrigo y otros](#)).

Las herramientas de desarrolla utilizadas han sido Visual Studio Code y PyCharm.

Se ha utilizado un entorno virtual en **Python** mediante la herramienta de creación de **entornos virtuales virtualenv**²³. De este modo, la distribución de la aplicación, tanto para su seguimiento por parte del tutor de este TFM, como para la evaluación por parte

²¹ <https://gephi.org/>

²² <https://linguakit.com/es/analisis-completo>

²³ <https://pypi.org/project/virtualenv/>

de un pequeño conjunto de usuarios de prueba, se simplifica gracias a la propia naturaleza y dinámica de los entornos virtuales. Esto es debido a que se evitan incompatibilidades con paquetes previamente instalados en los entornos de cada usuario. Además basta con distribuir el código fuente (archivos con extensión *.py*), los archivos auxiliares necesarios (corpus, imágenes, ficheros de ayuda, etc.) y el archivo denominado *requirements.txt*, a partir del cual se puede replicar un entorno virtual idéntico al utilizado para el desarrollo, no sólo con los mismos paquetes y dependencias necesarios, sino con las versiones idénticas de los mismos.

La simple ejecución del comando `pip list` desde el entorno virtual, muestra todos los paquetes instalados en el mismo y la ejecución del comando `pip freeze > requirements.txt` crea el fichero *requirements.txt*. Ver el apartado de [Distribución e Instalación de la aplicación](#).

4.1. Aprovechamiento del trabajo anterior

Un objetivo de este trabajo es la reutilización y aprovechamiento parcial del trabajo previo ([OBJ3](#)). En este sentido se ha reutilizado:

- Parte del código de entrenamiento para el algoritmo Naive Bayes utilizado en el análisis de sentimientos. Se ha rediseñado al utilizarse otra librería diferente a la del trabajo anterior. En este trabajo se hace uso de *naiveBayesClassifier*²⁴ mientras que en el trabajo previo se utiliza el clasificador *NaiveBayesClassifier* de *textblob.classifiers*.
- Parte del código de creación de los grafos de red.
- Parte del código del filtrado por fechas.
- Parte del código de apertura de ficheros.

Partir de estas estructuras ha constituido una gran ayuda y ahorro de tiempo, centrándose en la mejora funcional y estética y en la corrección de pequeños errores. Lo

²⁴ <https://pypi.org/project/naiveBayesClassifier/>

que ha permitido muy probablemente obtener unos mejores resultados atendiendo a la relación coste/beneficio.

Además de lo anterior, se ha aprovechado parte del código facilitado por el Equipo Docente, correspondiente a la aplicación desarrollada y presentada por Álvaro Rodrigo, José Luis Fernández, Jorge Pérez, Ismael Iglesias, Víctor Fresno, Aitor Díaz, Francisco Javier Sánchez y Roberto Centeno en la edición XI de las Jornadas de Investigación en Innovación Docente de la UNED. Avances en la Formación híbrida y en línea para una educación post-pandemia. Durante los días 11 a 13 de mayo de 2021.

De la aplicación anterior se ha aprovechado de manera parcial con algunas modificaciones lo siguiente:

- Algoritmo para calcular Región de Russell para el análisis de emociones.
- Algoritmo para parseo de foros en formato TXT.

4.2. Aportaciones significativas sobre el trabajo anterior

Aunque en las secciones subsiguientes se podrá ver en más detalle, se van a mencionar las mejoras más significativas realizadas en este trabajo partiendo de la idea y trabajo anterior. Estas son entre otras:

- Evolución hacia un diseño de interfaz integrado, en el que los resultados van apareciendo dentro de una única ventana de aplicación estructurándose en diferentes pantallas.
- Una de las líneas de continuación mencionadas por la autora del trabajo anterior consistía en la realización del análisis de sentimientos y emociones de manera nativa en lengua española, puesto que en dicho trabajo se utilizaban sendos análisis en lengua inglesa, debiéndose traducir todos los mensajes al inglés antes

de realizar el análisis. Esto planteaba un problema de rendimiento por un lado y de posible desvirtuación a nivel semántico en el proceso de traducción. En este sentido se han incorporado dos análisis de sentimientos que utilizan la lengua española, como son PySentimiento y Sentyment Análisis Spanish además de implementar el análisis de sentimientos Naybe Bayes con el que se estaban realizando pruebas en el trabajo anterior. En lo que se refiere al análisis de emociones se utilizan dos análisis de emociones que también se verán a continuación en más detalle y que utilizan corpus con palabras en lengua española.

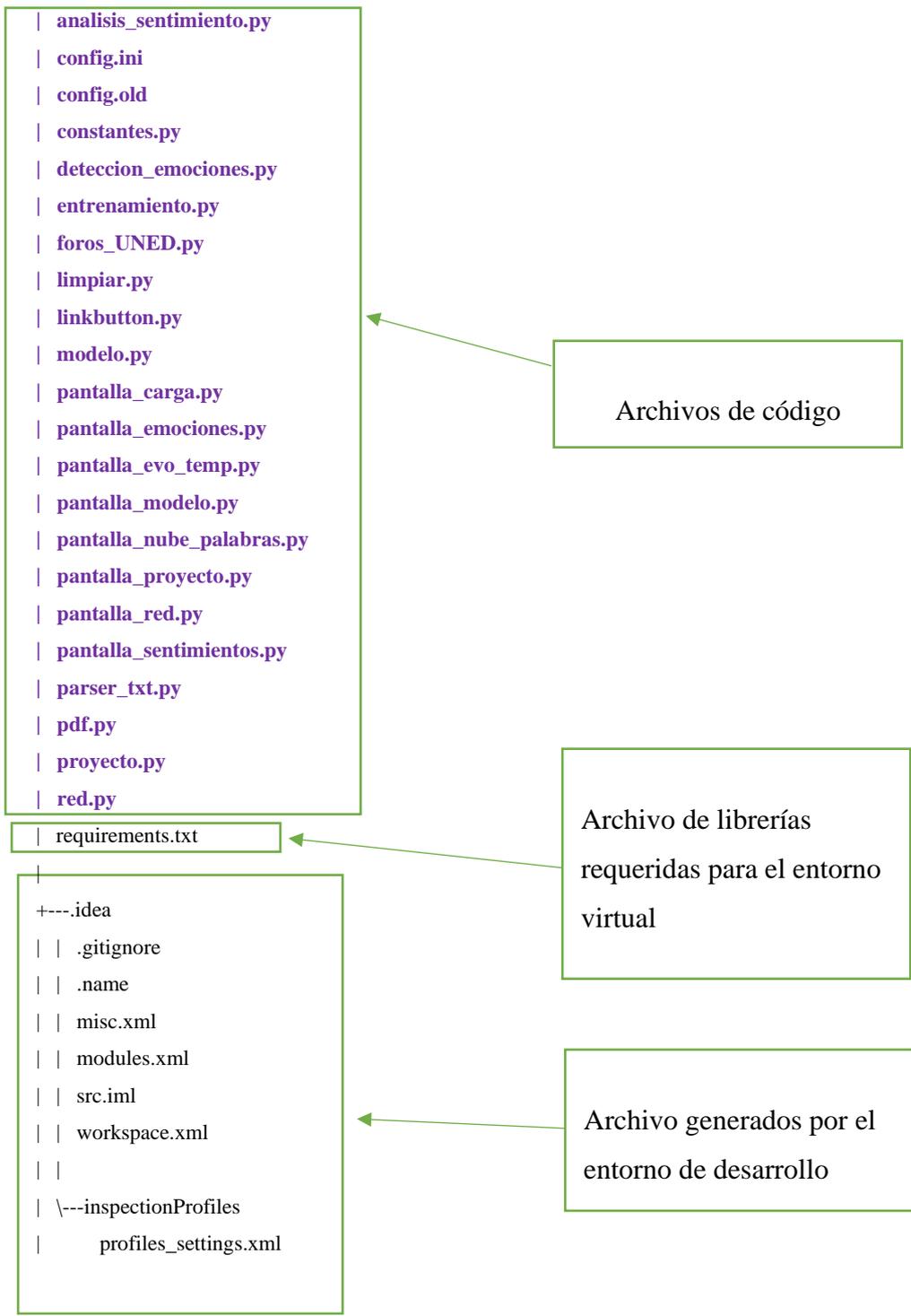
- Flexibilidad en el formato y tipo de ficheros para procesar datos de foros.
- Creación de nube de palabras con las palabras más utilizadas en los foros.
- Mejora de la funcionalidad de representación de redes.
- Mejora en la funcionalidad de *clustering* (agrupamiento).
- Gráficos de usuarios.
- Posibilidad de exportar a Excel los resultados de los análisis.
- Módulo de generación de informes automáticos con las tablas y los gráficos generados durante la sesión de trabajo.
- Posibilidad de guardar proyectos y abrirlos con posterioridad.

4.3. Estructura de la aplicación creada

4.3.1. Estructura de ficheros

Respecto a la estructura de ficheros, la aplicación cuenta con los siguientes directorios y ficheros:

C:.



| Project_Default.xml
|
|

```
+---archivos  
| | 62011037_2018.txt  
| | 62011037_2018.xlsx  
| | 62011037_2018_reducido.xlsx  
| | AA-I-foro-19-20-anon.csv  
| | AA-I-Foros.xlsx  
| | AA-I-Foros_editado.xlsx  
| | archivo.csv  
| | archivoForo.csv  
| | archivoForo.xlsx  
| | archivo_editado.csv  
| |
```

Archivo generados por el entorno de desarrollo

```
+---Ayuda  
| \---on. line  
| ayuda_emoc.pdf  
| ayuda_mod_inform.pdf  
| ayuda_red.pdf  
| ayuda_sent.pdf  
| ayuda_temp.pdf  
| como_usar.pdf  
| formato_1.pdf  
| formato_2.pdf  
| formato_nb.pdf  
| informe_ejemplo.pdf  
|
```

Archivos de ayuda

```
+---Encuesta  
| \---on. line  
| encuesta_ux.docx  
|
```

Carpetas y archivo para la encuesta de Experiencia de Usuario

```
+---imagenes  
| analisis.png  
| angry_amarillo.png  
| angry_rojo.png  
| asco.png  
| ayuda.png  
| AyudaProyecto_1.png  
| AyudaProyecto_2.png  
| AyudaProyecto_3.png  
| AyudaProyecto_4.png  
| AyudaProyecto_5-6.png
```

Imágenes de la aplicación

- | cargar_csv.png
- | cargar_foros.png
- | cargar_txt.png
- | clustering.png
- | corpus.png
- | csv.png
- | dia_semana.png
- | document.png
- | documento.png
- | documentos.png
- | ejemplo1.jpg
- | enfadado.png
- | escuela_uned.gif
- | escuela_uned.jpg
- | etsiuned.png
- | excel.png
- | feliz.png
- | home.png
- | icono.png
- | informe.png
- | modelo.png
- | noemoc.png
- | nube.png
- | nube2.png
- | plan.png
- | procesando.gif
- | procesando_peq.gif
- | proyecto.png
- | ranking.png
- | red.png
- | relajado.png
- | sent0.png
- | sent1.png
- | sent2.png
- | sent3.png
- | sent4.png
- | sentiment_icon.png
- | sentiment_icon_old.png
- | settings.png
- | tabla.png
- | tabla_datos.webp
- | temeroso.png
- | triste.png
- | usuarios.png

Imágenes de la aplicación

```
+---Proyectos
| |
| \---Prueba ver. Alpha 1.62 archivoForo.csv
| | Prueba ver. Alpha 1.62 archivoForo.csv.txt
| |
| +---Datos
| | +---Corpus
| | | +---Emociones
| | | | 13428_2016_768_MOESM1_ESM.xls
| | | | Spanish-es-NRC-VAD-Lexicon.txt
| | | |
| | | \---Sentimientos
| | | | mi_corpus.csv
| | | |
| | | \---Foros
| | | | archivoForo.csv
| | | |
| +---Graficos_Clustering
| +---Graficos_Emociones
| +---Graficos_Nube
| +---Graficos_Red
| +---Graficos_Sentimientos
| +---Graficos_Temporales
| +---Graficos_Usuarios
| | Usuarios 45c2fa9a13586c13ee4eff1669425d88a7a443fc-u3.png
| | Usuarios ecafaa42d80d83a69b2f2311930cf62d550be1f0-u2.png
| |
| \---Infor. mes
| | informe_1.html
| | informe_modelo.html
| |
| +---Excel
| \---PDF
| | informe_grafico.pdf
```

Proyecto de ejemplo.

```
+---recursos
| 13428_2016_768_MOESM1_ESM.xls
| mi_corpus.csv
| Spanish-es-NRC-VAD-Lexicon.txt
```

Carpetas y archivo para la encuesta de Experiencia de Usuario

4.4. Interfaz de la aplicación y funcionamiento

La interfaz de la aplicación desarrollada consiste en:

- Una **barra superior de menús convencional** [1] con los menús: Archivo; Configuración; Preferencias y Ayuda.
- Una **barra de menú lateral** [2] para acceder a los diferentes módulos en los que está estructurada la aplicación.
- Un **área principal** [3] en la que se muestra el módulo actualmente en uso. En la **pantalla de Inicio**, que es la que se muestra inmediatamente después de abrirse la aplicación, se dispone de unas **pequeñas instrucciones** de uso e **información de las funcionalidades** disponibles [4]. También aparece en la parte central, bajo las instrucciones básicas, una **columna de botones** (Baker et al., 2009) [5] mediante los que se puede acceder a los diferentes módulos exactamente igual que se hace desde la barra lateral izquierda. Finalmente, en la parte inferior izquierda se presentan unos **enlaces a sitios web** [6] de interés relacionados con la aplicación.

Esta estructura se puede ver en la figura 17.

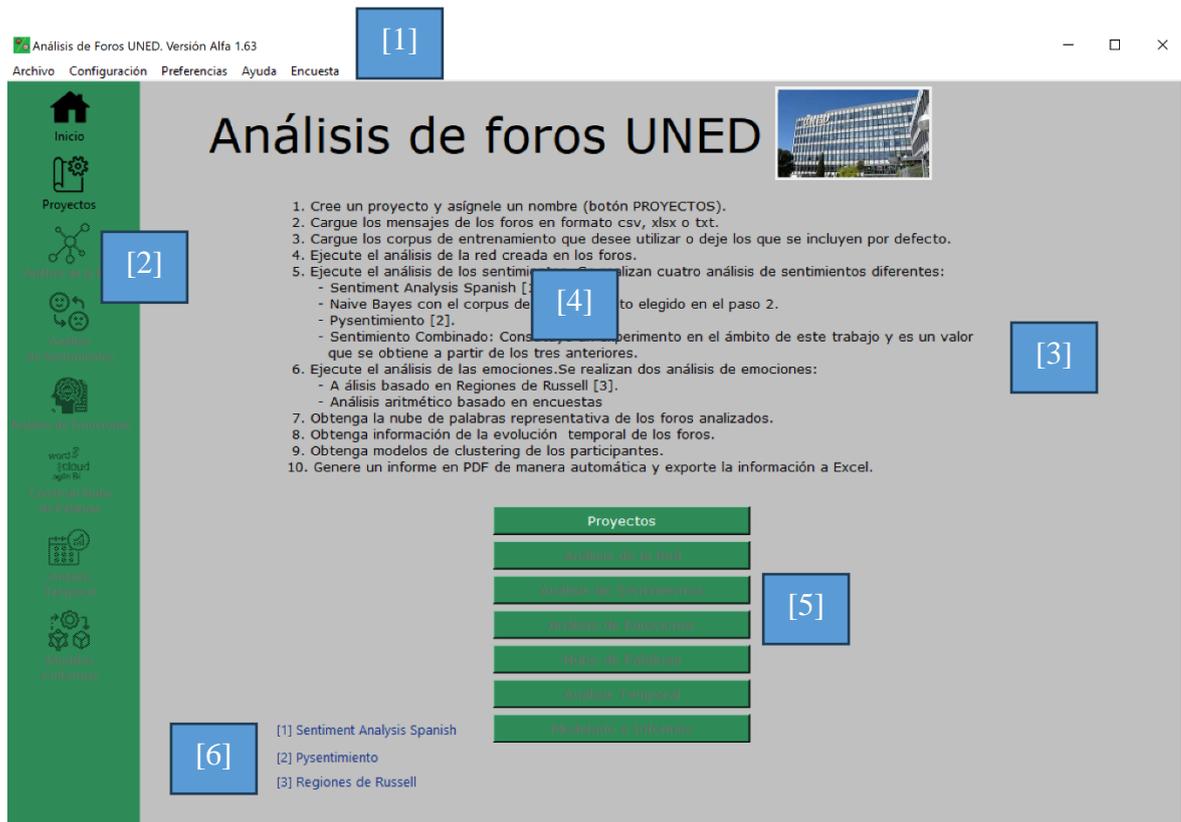


Figura 17. Aspecto de la interfaz de la aplicación inmediatamente después de abrirse.

4.4.1. Preferencias de Tema

Se dispone de la posibilidad de poder configurar dos temas de interfaz que se han identificado como Verde UNED y Azul RELAX. Para cambiar de tema se debe acceder al menú superior Preferencias/Tema/ (Ver figura 18) y para que el cambio de tema surta efecto es necesario reiniciar la aplicación. El tema Verde UNED tiene el aspecto de la figura 17, mientras que el tema Azul RELAX se puede ver en la figura 19.

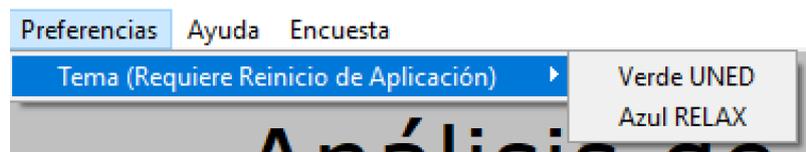


Figura 18. Preferencias de Tema.

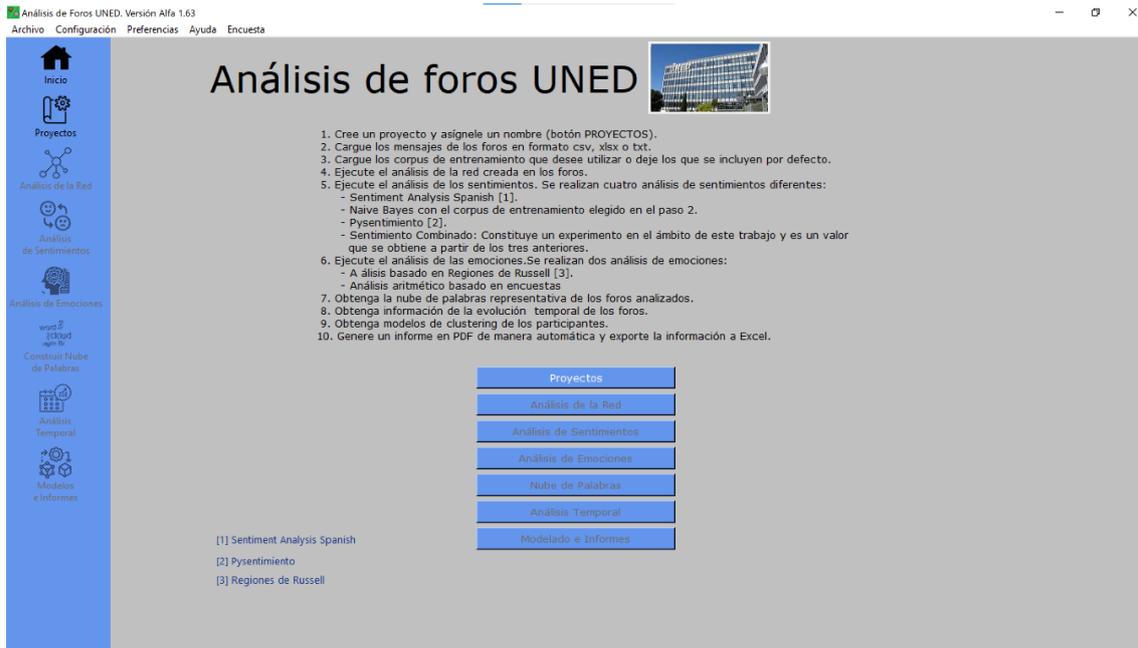


Figura 19. Aspecto de la interfaz con el tema Azul RELAX.

4.4.2. Configuración

En la barra de menús superior se dispone del menú *Configuración* que en principio sólo tiene la opción de *Editar config.ini* mediante la cual se abre dicho archivo de configuración para poder editarlo. Esto se muestra en la figura 20.

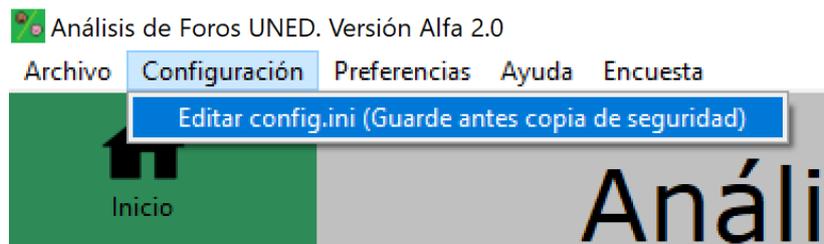


Figura 20. Menú de *Configuración*.

Es muy importante tener en cuenta que cualquier error en el archivo *config.ini* puede tener consecuencias fatales en el funcionamiento de la aplicación, por lo que es importante disponer de una copia de seguridad del archivo *config.ini*. El contenido de este archivo se puede ver en la figura 21.

```

config.ini: Bloc de notas
Archivo Edición Formato Ver Ayuda
[[info]]
autor = Félix Adame Toledano

[colores]
bgalzap = silver
bgsidebar = seagreen
sidebar = seagreen
bgbotones = seagreen
bgbotones2 = seagreen
bgbotones3 = #BEE4D2
bgbotones4 = #BEE4D2
fondoppal = silver

[DimPantalla]
widthmultip = 9
widthdiv = 11
heightmultip = 3
heightdiv = 5

[menuLateral]
coefAncho = 0.1
coefAlto = 1.0

[NBnum]
pos = 0.75
neg = 1e-1

[PYSnum]
neu = 0.5
pos = 1
neg = 1e-2

[Combina]
coefNB = 1.0
coefPyS = 1.0
coefSentSpNorm = 1.5
pot = 0.5
coefTot = 0.01
    
```

Figura 21. Archivo *config.ini*.

4.4.3. Creación de un proyecto

Al iniciar la aplicación, sólo están disponibles la pantalla de Inicio y el Módulo de Proyectos. Así mismo se puede acceder a las opciones de los menús de la barra superior. Si accedemos al módulo de Proyectos se nos presentará la pantalla de la figura 22.

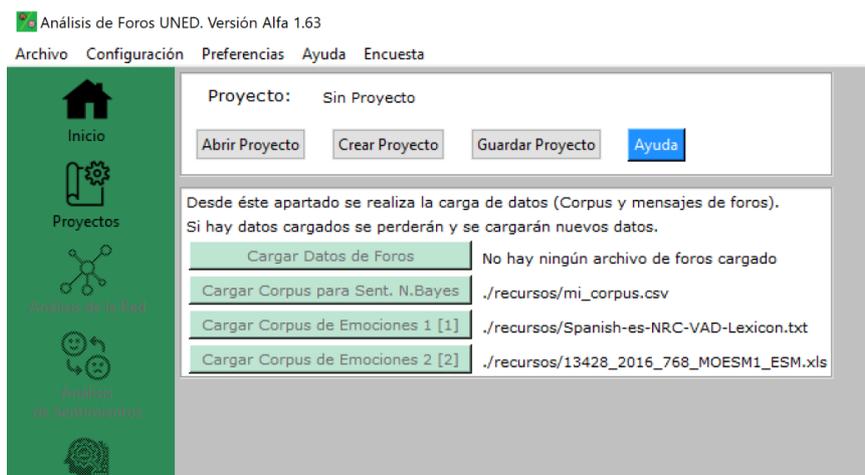


Figura 22. Módulo de Proyectos al acceder al mismo tras abrir la aplicación.

Desde esta interfaz podemos abrir un proyecto previamente guardado, crear un nuevo proyecto o guardar un proyecto en uso. Se dispone de un botón de **Ayuda contextual** para la creación de proyectos (destacado con fondo azul). Si pulsamos el botón de Ayuda de esta pantalla se nos muestra una ventana como la que se muestra en la figura 23.

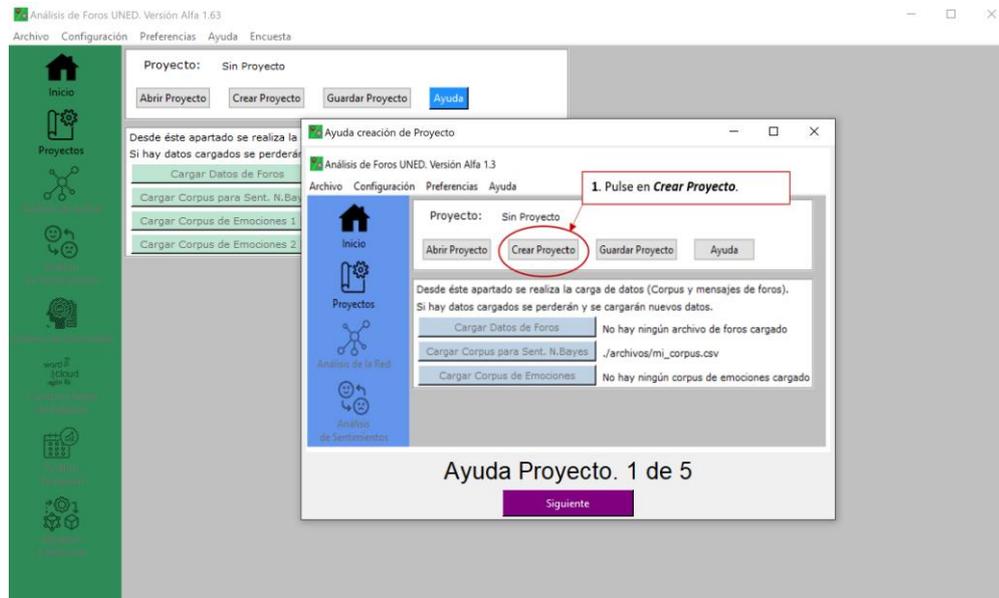


Figura 23. Ayuda contextual para la creación de proyectos.

Dentro de la ayuda de creación de proyecto se dispone de un botón con la inscripción **Siguiente** que al pulsar va avanzando por el proceso de creación de proyecto. Se trata de una secuencia cíclica de cinco imágenes. En la figura 24 se ven los cuatro pasos siguientes de esta ayuda.

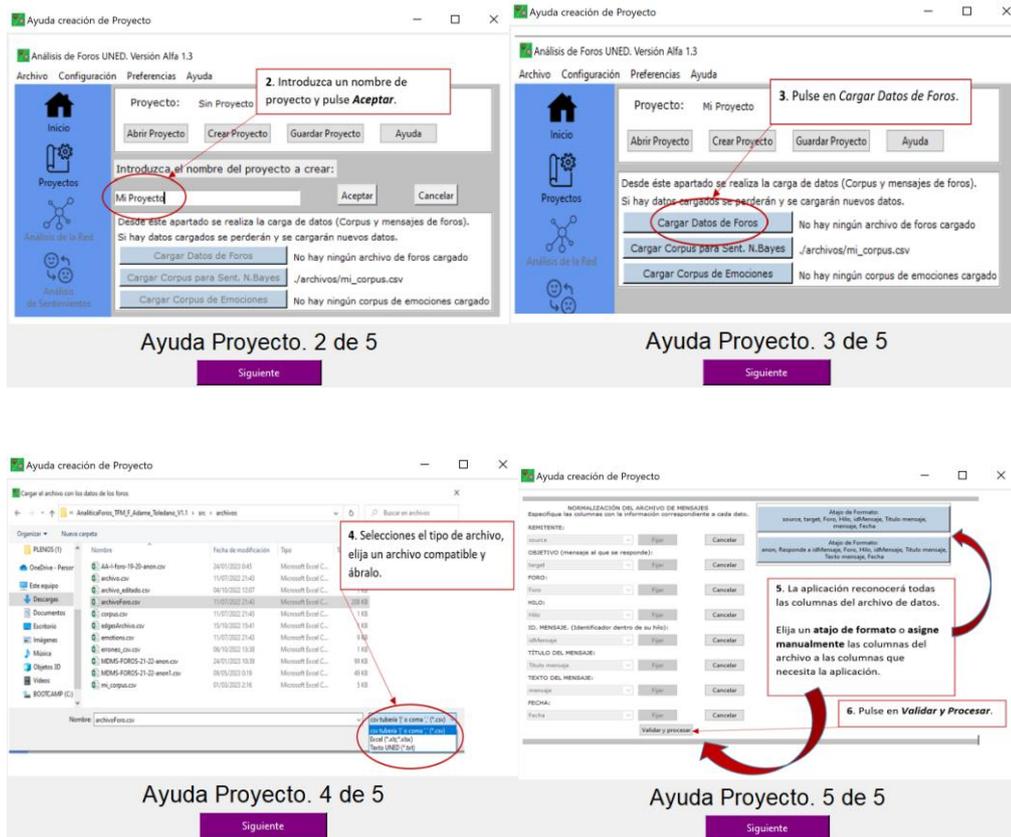


Figura 24. Resto de pasos de la ayuda de creación de proyecto.

Como se indica en la propia ayuda, lo primero que hay que hacer es darle un nombre al proyecto. Es importante indicar en este punto, que inmediatamente después de asignar un nombre de proyecto, se crearán automáticamente dentro de la ruta ./Proyectos relativa a la ubicación de los archivos de aplicación, una carpeta con el nombre asignado al proyecto así como la estructura de subcarpetas que se muestra en la figura 25.

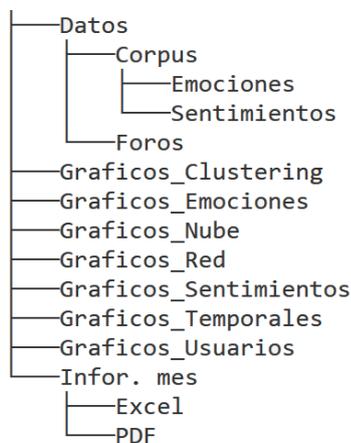


Figura 25. Estructura de directorios de un proyecto.

Sin embargo, el archivo de proyecto propiamente dicho no se puede guardar hasta que no se haya realizado el procesamiento de datos. Supongamos que queremos crear el proyecto denominado *Mi Proyecto*. En la figura 26 se muestra cómo inmediatamente después de asignar un nombre de proyecto se habilita el botón de *Cargar Datos de Foros*.



Figura 26. Habilitación del botón de carga de datos de foros.

Por tanto, resulta obvio que el siguiente paso es la carga de datos de foros.

Respecto de la carga de datos es importante destacar que se ha dotado a la aplicación de una flexibilidad considerable a la hora procesar archivos con datos de foros.

Si en la aplicación del trabajo anterior las columnas debían tener un nombre específico y el formato de archivo debía ser *.csv* con valores separados por el carácter “|”, en esta aplicación se admiten los siguientes formatos:

- Archivos csv de valores separados por “|”
- Archivos csv de valores separados por “,”
- Archivos csv de valores separados por “;”
- Archivos xls o xlsx
- Archivos txt con el formato utilizado en la aplicación suministrada y presentada en el programa de Redes de Innovación Docente.

Además de la flexibilidad anterior respecto del tipo de archivo, no se requiere ningún nombre específico para las columnas o encabezados de los archivos. Esto es gracias y a costa de un proceso denominado de *Normalización* del archivo.

Cuando pulsamos en *Cargar Datos de Foros* se muestra un cuadro de diálogo de selección de archivo, como se muestra en la figura 27.

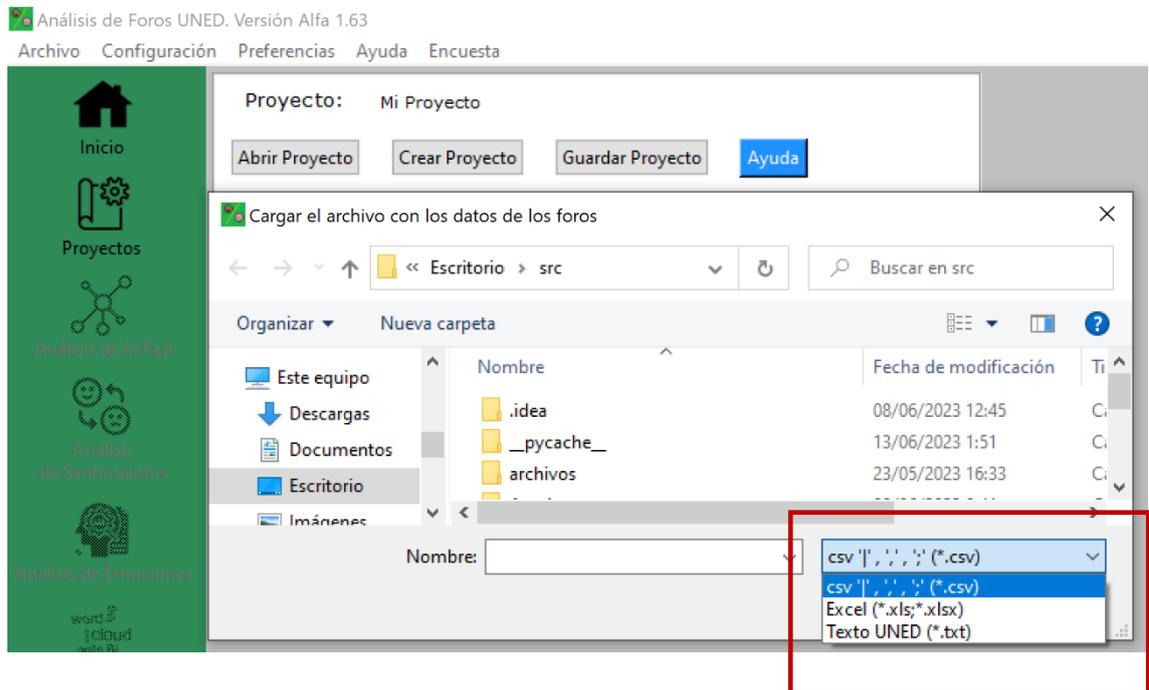


Figura 27. Cuadro de diálogo de apertura de archivo de datos de foros.

En este cuadro de diálogo se puede, mediante el desplegable de tipo de archivo, elegir cualquiera de los formatos disponibles, según se ve en el recuadro de color rojo de la figura 27.

Una vez abierto (cargado) un archivo de datos, se amplía la pantalla del módulo de Proyectos, incluyendo una nueva sección de *Normalización de Archivos de Mensajes*, adoptando el aspecto de la figura 28.

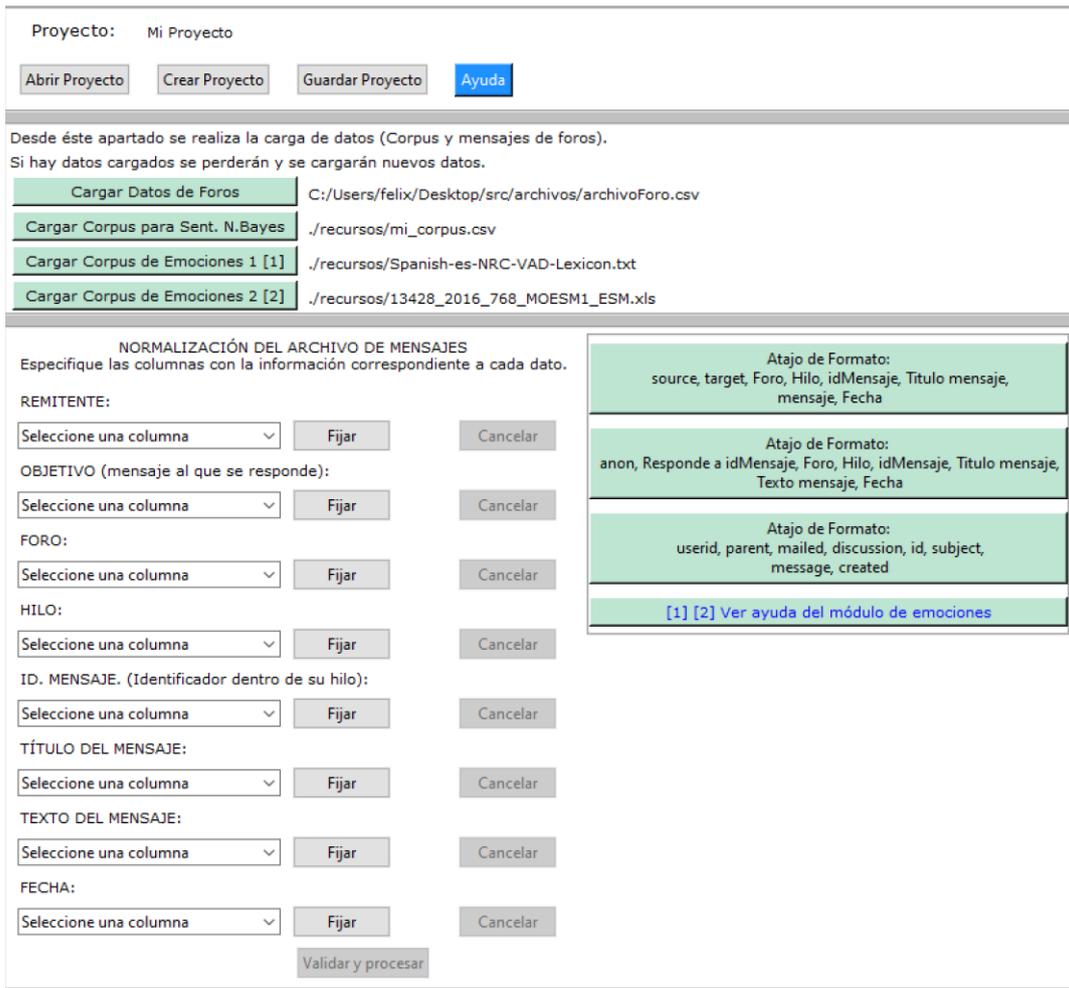


Figura 28. Pantalla de Proyectos con la sección de normalización de archivos.

El proceso de normalización consiste en crear una asociación entre las columnas del archivo y las columnas que se requieren para el procesamiento de los datos. Dicha asociación se puede realizar de dos maneras:

- Normalización *manual*, asociando columna a columna. Cuando estemos seguros de que queremos asignar una columna podemos pulsar el botón *Fijar* asociado a la misma, de modo que esta columna dejará de aparecer en el resto de desplegables. Podemos deshacer una asignación o asociación de columna pulsando el botón *Cancelar* asociado a la columna en cuestión. Una vez fijadas todas las columnas se habilita el botón *Validar y procesar* situado en la parte inferior de la sección de normalización de archivo.
- Normalización *automática* mediante *atajos de formato*. Existen cuatro atajos de formato en la versión 3.0 Beta de la aplicación.

Durante la fase de desarrollo y pruebas posteriores se han ido incorporando formatos de archivos, por lo que existen los atajos de formatos que se muestran y que corresponden con los formatos utilizados. Aún así se mantiene la opción de normalización manual para poder realizar pruebas con otros archivos diferentes.

No es posible aún guardar el archivo de proyecto, obteniendo el mensaje que se muestra en la figura 29 en caso de intentarlo.

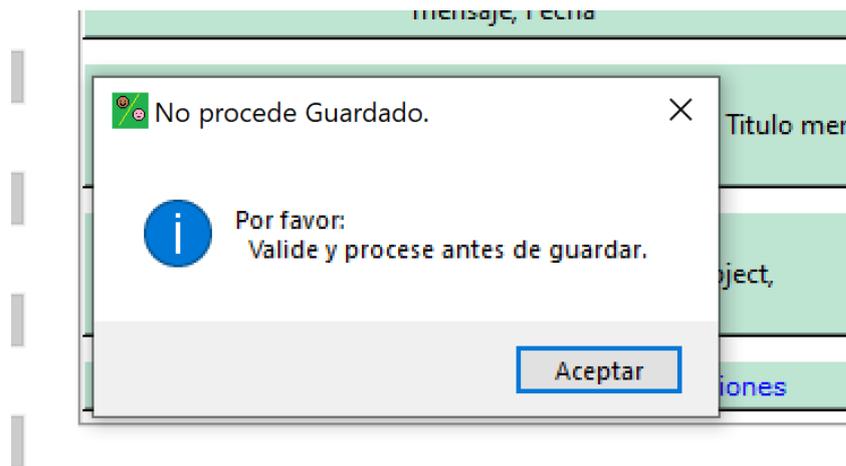


Figura 29. Mensaje indicando que aún no procede el guardado de proyecto.

En este punto es también posible, antes de realizar la normalización, seleccionar archivos diferentes a los proporcionados por defecto como corpus de sentimientos y emociones para los diferentes análisis, lo que se realiza mediante los botones de Carga de Corpus correspondientes, que llevan a cuadros de diálogo de selección de archivos.

Una vez realizada la asociación de columnas (normalización) ya se de manera manual o mediante alguno de los atajos de formato y habilitado el botón *Validar y procesar*, podemos pulsar dicho botón, lo que nos lleva a una pantalla de carga que se muestra en la figura 30.

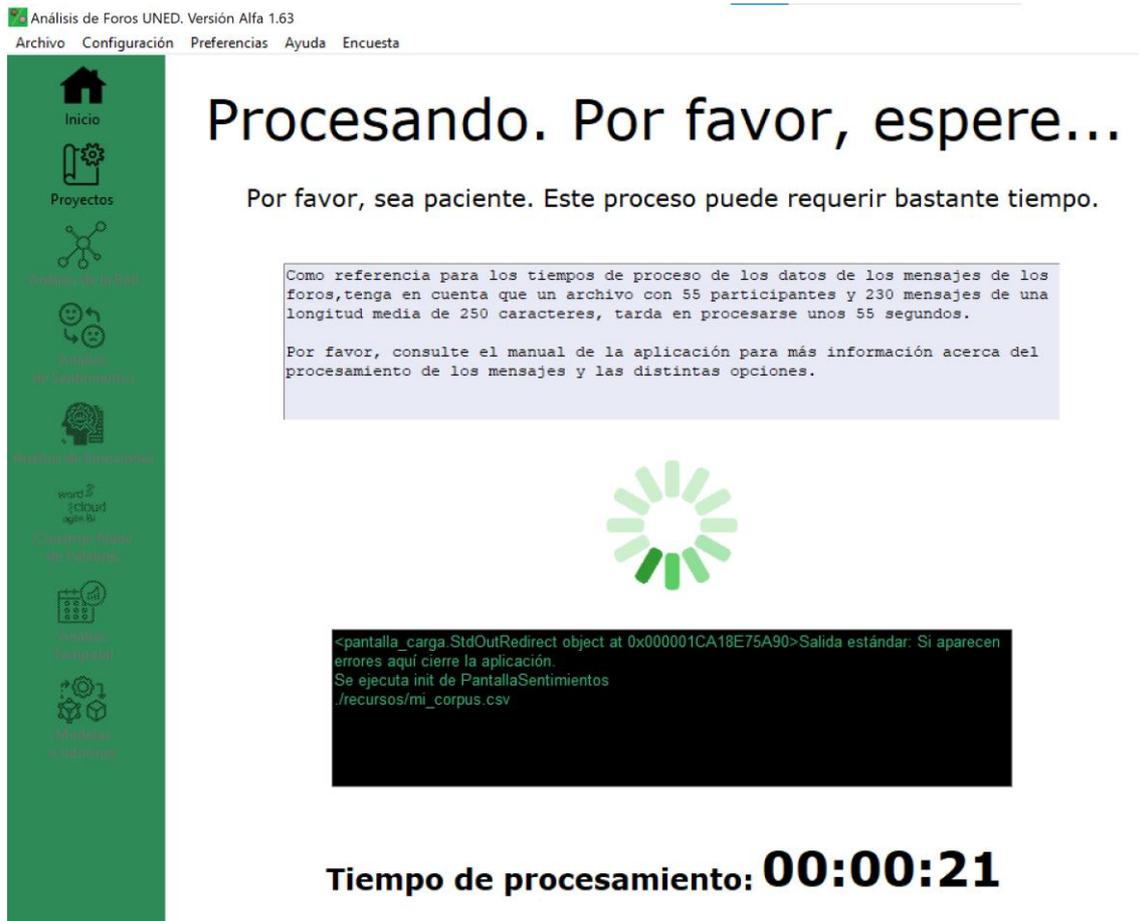


Figura 30. Pantalla de procesamiento de datos.

Esta pantalla de procesamiento se muestra tanto en la creación como en la carga de proyectos. Se muestra un **cronómetro** para tener una referencia del tiempo de procesamiento y poder realizar comparaciones y análisis de funcionamiento de la aplicación y una **ventana en la que se visualiza la salida estándar y de errores**. Si se produce un error durante este proceso el cronómetro no se detendrá, por lo que **deberá vigilarse si aparece algún error**, lo que significará que debemos abortar la ejecución de la instancia de la aplicación. Si esta ventana produce comportamiento no deseado se pueden comentar las líneas 47, 48 y 52 del código del archivo `pantalla_carga.py`.

Una vez finalizado el preprocesamiento de los datos, veremos la interfaz con el aspecto de la figura 31.

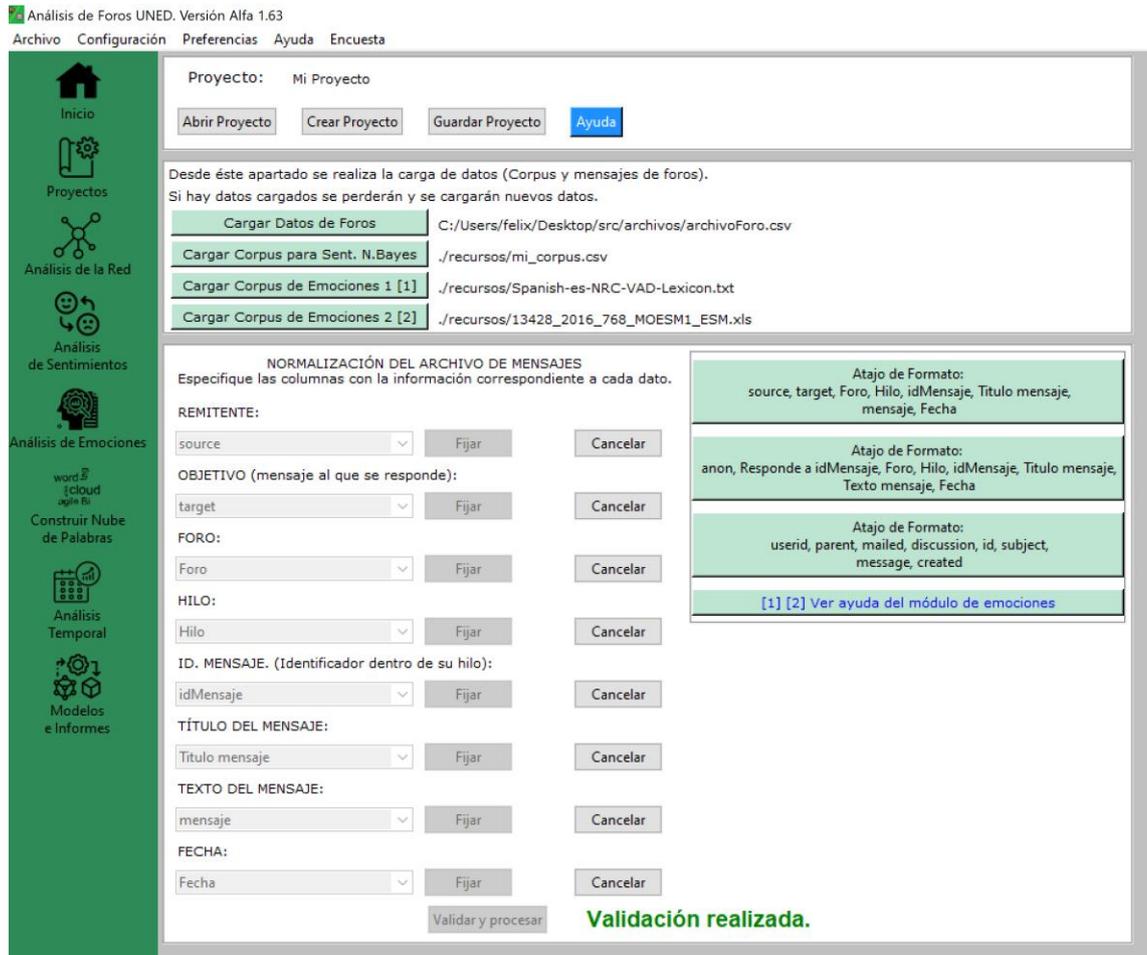


Figura 31. Interfaz de la aplicación una vez realizado el preprocesamiento de los datos de los foros.

En esta vista se puede observar que se han habilitado los botones laterales correspondientes a los diferentes módulos que ya están accesibles. A partir de este momento también es posible guardar el archivo de proyecto pulsando el botón *Guardar Proyecto*, lo que dará como resultado el mensaje que se muestra en la figura 32.

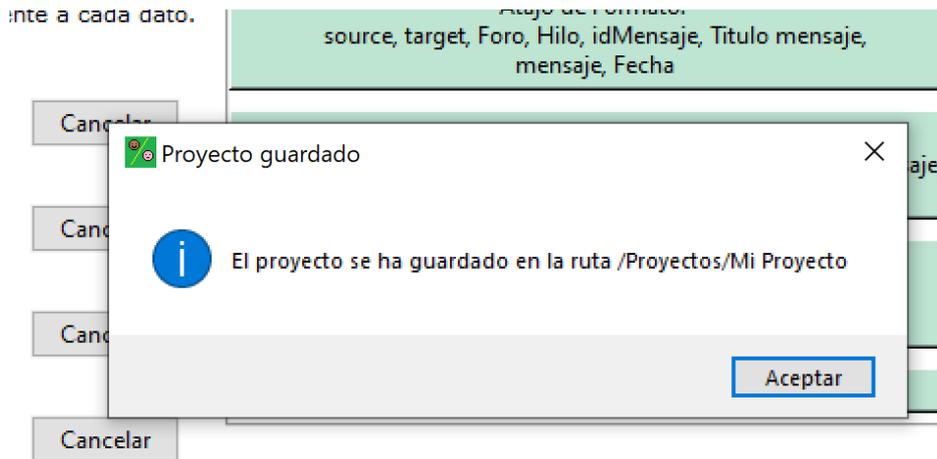


Figura 32. Cuadro de aviso informando del guardado del proyecto.

Así mismo, se guardará un archivo con el nombre asignado al proyecto y la extensión .txt en la ruta: *./Proyectos/Mi Proyecto/Mi Proyecto.txt*, asumiendo que el nombre de proyecto es *Mi Proyecto*.

4.4.4. Análisis de la Red

Al pulsar el botón de Análisis de Red se irá al módulo del mismo nombre y se mostrará la pantalla de la figura 33.

Analisis de Foros UNED, Versión Beta 3.0

Archivo Configuración Preferencias Ayuda Encuesta

Red del proyecto proy 1

Nodos: 28 Aristas: 148

Estadísticas de la red

Centralidad de Grado Cercanía Intermediación Centralidad de Vector Propio Page Rank Iniciativa Modularidad Sentimientos Ver Valores de los Nodos

Etiquetas de gráfico: Alias Sentimiento: N.Bayes (Corpus Usuario) Ayuda Exportar Red

Filtrar por: todo Seleccionar: TODO Visualizar archivo Fecha inicial Fecha final Calcular fecha

	NODO	ALIAS	GRADO	GRADO	GRADO	CENTRAL	CERCANIA	INTERME	VECTOR	PAGE R	INICIATIV	MODULA
1	Prof-130	u1	0	0	0	0	0	0	4.4e-13	0.0063	1	0
2	Estu-115	u2	26	16	10	0.96	0.64	0.14	0.4	0.11	6	1
3	Estu-56	u3	16	7	9	0.59	0.5	0.073	0.25	0.046	1	2
4	Estu-39	u4	21	13	8	0.78	0.58	0.13	0.27	0.076	4	5
5	Estu-14	u5	8	5	3	0.3	0.47	0.064	0.18	0.042	2	1
6	Estu-103	u6	7	4	3	0.26	0.45	0.065	0.13	0.031	3	1
7	Estu-21	u7	16	6	10	0.59	0.49	0.08	0.2	0.041	2	2
8	Coor-36	u8	39	18	21	1.44	0.7	0.4	0.42	0.11	19	1
9	Estu-40	u9	10	5	5	0.37	0.48	0.068	0.18	0.038	2	1
10	Estu-159	u10	10	5	5	0.37	0.47	0.066	0.18	0.035	2	1
11	Estu-1	u11	16	9	7	0.59	0.54	0.11	0.26	0.058	6	2
12	Estu-100	u12	2	1	1	0.074	0.41	0.064	0.05	0.011	0	5
13	Estu-140	u13	12	5	7	0.44	0.45	0.067	0.15	0.034	5	1
14	Estu-66	u14	8	4	4	0.3	0.46	0.068	0.13	0.032	0	1
15	Estu-143	u15	10	5	5	0.37	0.44	0.064	0.15	0.028	0	2
16	Estu-92	u16	3	3	0	0.11	0.48	0.034	0.12	0.024	2	5
17	Estu-113	u17	29	16	13	1.07	0.63	0.19	0.37	0.094	6	5
18	Estu-122	u18	2	1	1	0.074	0.41	0.063	0.05	0.011	0	1
19	Estu-141	u19	16	8	8	0.59	0.49	0.076	0.22	0.04	5	2

28 rows x 12 columns

Figura 33. Vista inicial del módulo de Análisis de Red.

En esta vista inicial se puede ver en la parte superior el nombre del proyecto y el número de nodos y aristas que conforman la red. Se muestran todos los nodos de la red (usuarios) junto con los estadísticos de medidas de centralidad y modularidad para cada uno de ellos. También se ha incluido una columna denominada *ALIAS* en la que se puede ver un alias asignado automáticamente a cada usuario. Como se verá más adelante, este alias, además de proporcionar un medio para anonimizar los datos, sirve para mejorar la visualización de los grafos de red.

Las acciones disponibles en este módulo además de la primera vista de *Valores de los Nodos*, son:

- Visualización de los grafos correspondientes a las diferentes estadísticas. Junto a cada botón de estadística, a su derecha, hay un botón de ayuda mediante el cual se muestra una breve descripción de la misma. Estos botones de ayuda es algo que se ha mantenido prácticamente igual que en la aplicación anterior, si bien la disposición es totalmente diferente.
- Filtrado por cualquier variable o por fecha para ver la red parcial correspondiente a los datos una vez aplicado dicho filtro.
- Visualización de los datos correspondientes al filtro aplicado o bien de todos los datos.
- Exportación a Excel de los datos de la red.

A modo de ejemplo, se reproduce a continuación el resultado de la ejecución de Centralidad de Grado en la figura 34.

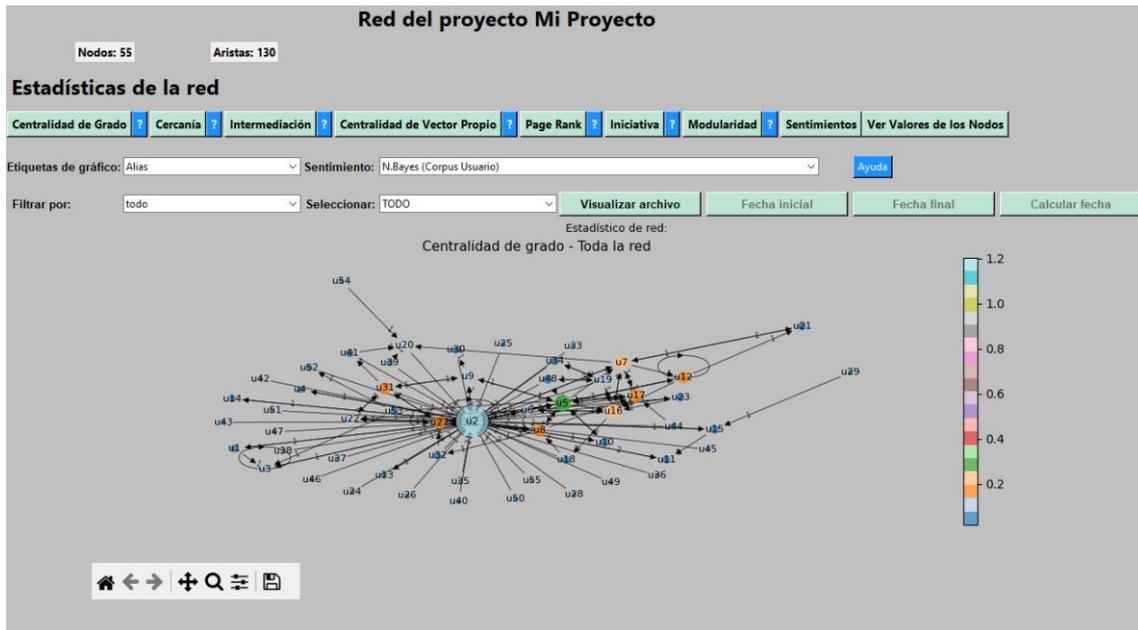


Figura 34. Grafo de la red según la Centralidad de Grado.

Las **mejoras incorporadas** en esta funcionalidad son las siguientes:

- Variación del tamaño de los nodos en función del grado de los mismos.
- Mejora del mapa de color, utilizando colores discretos en lugar de un degradado.
- Incorporación de la barra de herramientas de gráfico, mediante la cual se puede realizar zoom, mover, ajustar la ventana del gráfico y guardar el mismo en cualquier ubicación.
- Posibilidad de elegir las etiquetas para los nodos, pudiendo asignar como etiquetas: los identificadores de usuario, los alias asignados automáticamente por la aplicación o no mostrar etiquetas. Esto se hace mediante el desplegable incorporado a tal efecto con el texto a su izquierda que reza Etiquetas de *gráfico*.
- Incorporación de las etiquetas de arista para indicar el número de mensajes.

En la figura 35 se muestra el mismo gráfico una vez aplicado zoom y cambiadas las etiquetas a identificadores de usuarios, ya que por defecto se utilizan los alias.

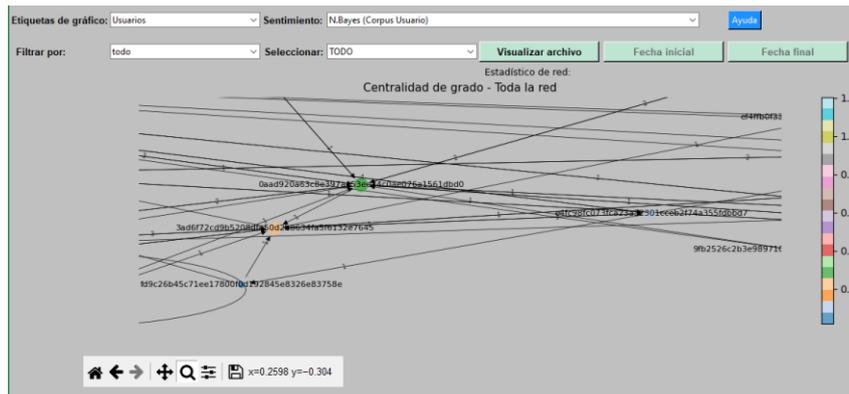


Figura 35. Red según Centralidad de Grado con etiquetas de usuarios y zoom aplicado.

Para poder apreciar la diferencia, en la figura 36 se ve este grafo con nombres de usuarios sin zoom.

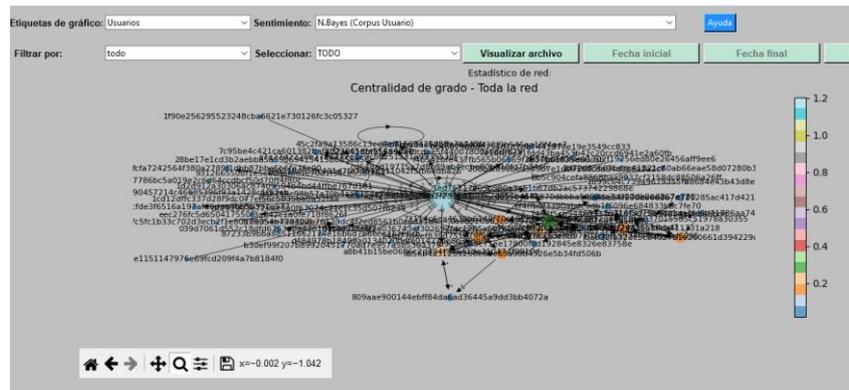


Figura 36. Mismo grafo de la figura 35 pero sin zoom.

4.4.4.1. Redes de Sentimientos

Como en el trabajo predecesor, se han mantenido los diferentes estadísticos de red, pero se ha añadido aquí la red de sentimientos, que en el trabajo anterior aparecía en otra pantalla a parte.

Respecto a la red de sentimientos, se ha incluido un desplegable para poder elegir entre dos diferentes algoritmos para representar la red. *N. Bayes* y *PySentimiento*.

N. Bayes considera mensajes positivos o negativos, mientras que *PySentimiento* incluye la categoría de mensajes neutros. Tanto en un algoritmo como en el otro, los

mensajes positivos se representan en verde mientras que los negativos se representan en rojo (figura 37).

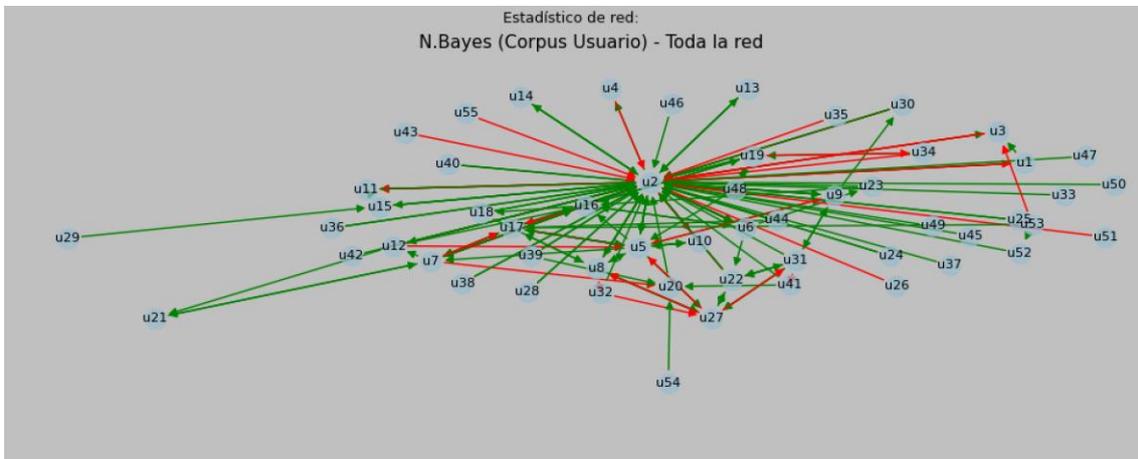


Figura 37. Red de Sentimientos según *Naive Bayes*

Para los mensajes neutros en el caso de Py Sentimiento, se ha elegido el color negro para representarlos (figura 38).

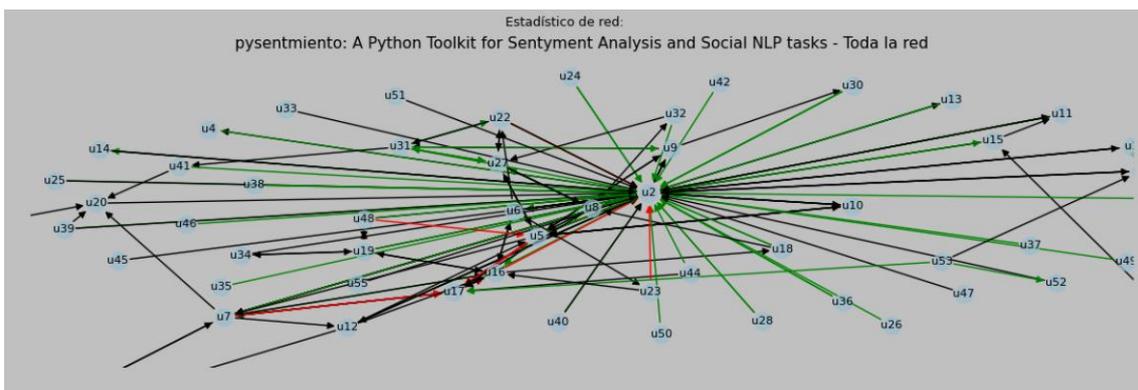


Figura 38. Red de Sentimientos según el algoritmo *PySentimiento*.

4.4.4.2. Modularidad

Pulsando el botón *Modularidad* de la pantalla de *Análisis de Redes* se dibuja un grafo de la red distinguiendo los nodos por colores según su pertenencia a supuestas comunidades (figura 39). Estas comunidades son las mismas que se indican en la columna *MODULARIDAD* de la tabla mostrada al pulsar en *Ver Valores de los Nodos*. El tamaño de los nodos se corresponde con el *GRADO*.

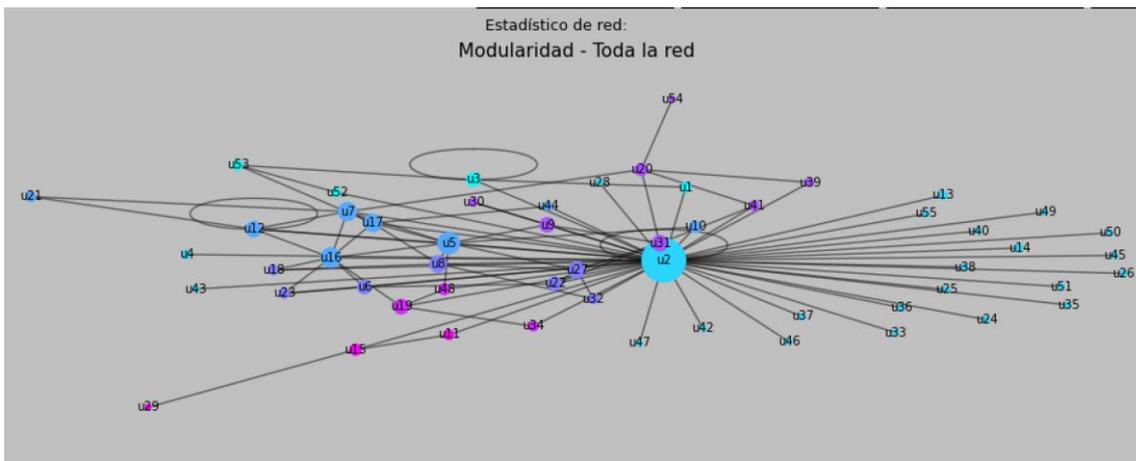


Figura 39. Modularidad de la red.

Es posible aplicar filtros atendiendo a cualquier variable (columna). Para ello basta con elegir un campo del desplegable *Filtrar por:* y seleccionar un valor del desplegable *Seleccionar:*

Se puede realizar también un filtrado por fechas para mostrar la subred que se genera en un determinado intervalo de tiempo. Para el filtrado por fechas hay que elegir el campo asignado a la fecha (Puede ser *Fecha*, *Date*, *Created*, etc.). Una vez hecho esto se habilitará el botón de *Fecha inicial*, que al pulsar mostrará un calendario en el que seleccionar la fecha inicial (figura 40). Una vez seleccionada la fecha inicial se habilita el botón *Fecha final*, el cual al pulsar muestra otro calendario para elegir la fecha de fin de intervalo.

Filtrar por:	Fecha	Seleccionar:	fecha	Visualizar archivo	Fecha inicial	Fecha final	Calcular fecha						
1	u5	Sábado	31/10/2020	40	28	29	30	1	2	3	4	da y videoconferencia p puntos buenas gracias por la correccion i	
2	u17	Lunes	30/11/2020	41	5	6	7	8	9	10	11	da y videoconferencia p puntos buenas tardes lo primero muchas	
3	u30	Lunes	30/11/2020	42	12	13	14	15	16	17	18	da y videoconferencia p puntos buenas tardes de acuerdo a la vida	
4	u40	Lunes	30/11/2020	43	19	20	21	22	23	24	25	da y videoconferencia p puntos buenas la conferencia me parece	
5	u7	Lunes	30/11/2020	44	26	27	28	29	30	31	1	da y videoconferencia p puntos buenas tardes muchas gracias pc	
6	u6	Lunes	30/11/2020	45	2	3	4	5	6	7	8	videoconferencia para buenas tardes la tarea 1 ha sido evaluad	
7	u2	Lunes	30/11/2020	20:35:51	Re: Tarea 1 evaluada y videoconferencia p puntos buenas tardes tambien creo que e							27	buenos dias paola en el paper menciona
8	u28	Lunes	30/11/2020	19:13:50	Re: problemas con la página de oreilly buenas tardes al principio del curso segu							27	buenos dias paola en el paper menciona
9	u31	Viernes	30/10/2020	14:56:51	Re: Significado de cada columna muchas gracias francisco coincido contigo							27	buenos dias paola en el paper menciona
10	u10	Viernes	30/10/2020	14:53:07	Re: Cuál lenguaje, R o Python, es mejor e buenas tardes os remito un enlace intere:							27	buenos dias paola en el paper menciona
11	u6	Viernes	30/10/2020	00:06:04	Re: Optimización de hyperparámetros hola luis gracias por el dato ayuda mucho							27	buenos dias paola en el paper menciona
12	u8	Martes	29/12/2020	09:47:36	Re: Número de búsquedas 2							27	buenos dias paola en el paper menciona
13	u18	Martes	29/12/2020	09:47:36	Re: Número de búsquedas 2							27	buenos dias paola en el paper menciona

Figura 40. Filtrado por fecha en la pantalla de *Análisis de Red*.

Cuando tenemos un filtro aplicado y generamos un grafo de red, en el título del grafo se puede ver el filtro que tenemos aplicado. Así, en la figura 41 se puede ver que se ha aplicado un filtro de fecha con el criterio de intervalo entre el 1 de octubre de 2020 y el 31 de octubre de 2020. Si no hay filtros aplicados se indicará “*Toda la red*”.

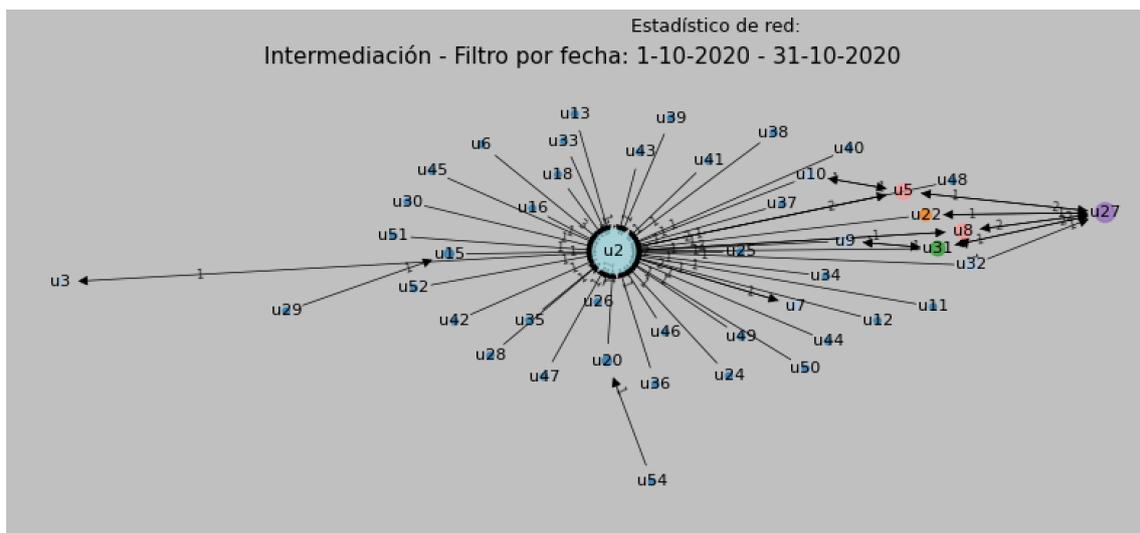


Figura 41. Grafo de red en un intervalo de tiempo.

4.4.5. Análisis de Sentimientos

Cuando pulsamos el botón de Análisis de Sentimientos accedemos a la pantalla de la figura 42, donde se muestra una tabla con lo que en el ámbito de este trabajo se ha denominado:

4.4.5.1. Sentimiento medio combinado (por cada usuario).

Se muestra para cada usuario: Un icono representativo de su sentimiento medio combinado, Su alias (generado automáticamente por la aplicación), su identificador de

usuario y el valor numérico para su *Sentimiento medio combinado*²⁵ que se calcula con la fórmula configurable:

$$Sent\ Comb = \frac{coefNB * NBnum + coefPyS * PySnum + coefSentSpNorm * Sent.Norm^{pot}}{3} * coefTot$$

Por defecto: coefNB = 1.0; coefPyS = 1.0; coefSentSpNorm = 1.5; pot = 0.5; coefTot = 0.01

Los coeficientes anteriores se configuran en el archivo *config.ini*.

Teniendo además que:

$$0.1 \leq NBnum \leq 0.75; PySnum \in \{1, 0.5, 0.01\}; 0 \leq Sent.Norm \leq 1$$

Por lo que $0 \leq Sent.Comb \leq 1.125$, siendo 0 la representación del peor sentimiento posible y en el otro extremo, 1.125 el mejor sentimiento posible.

En la parte derecha se puede ver una leyenda de los iconos que se utilizan en esta vista. Para ocultar esta leyenda se puede pulsar la tecla *Mostrar/Ocultar Leyenda*.

²⁵ La fórmula aplicada, así como los coeficientes aplicados, se han alcanzado mediante múltiples pruebas. Los parámetros son configurables mediante el archivo **config.ini**.

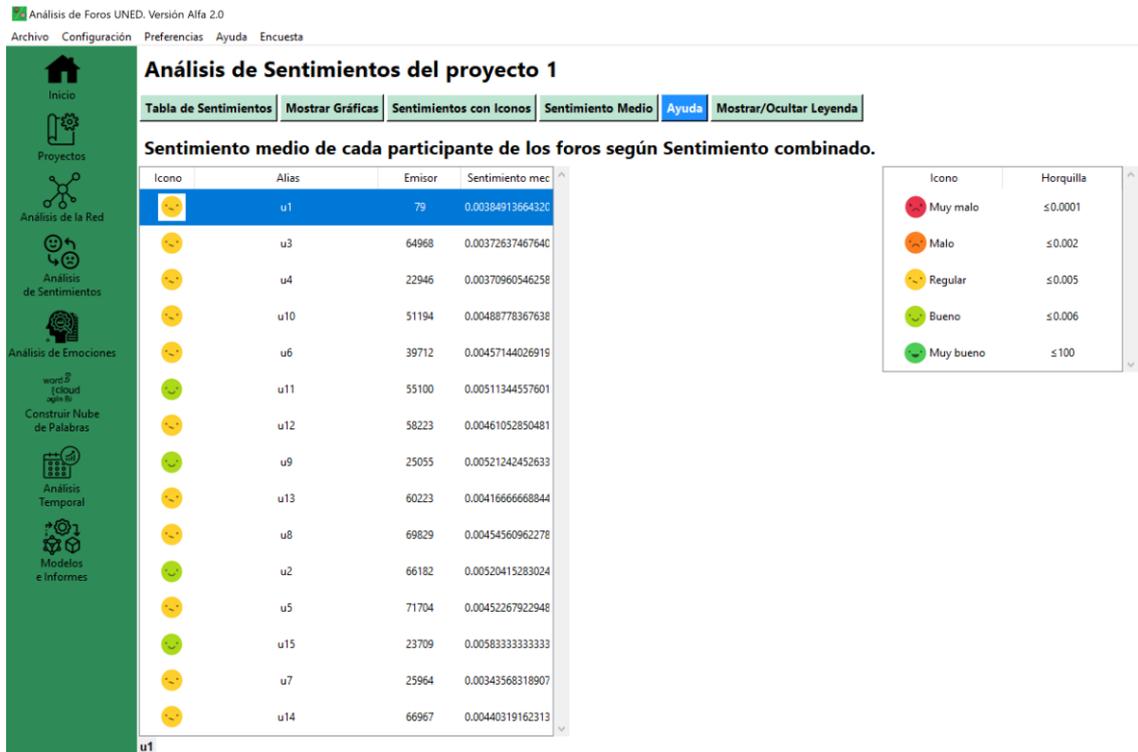


Figura 42. Tabla de *Sentimiento medio combinado* en el módulo de Análisis de Sentimientos.

La clasificación realizada en la aplicación para traducir el sentimiento combinado medio de los participantes a un literal se realiza mediante las particiones definidas en el siguiente fragmento de código del archivo pantalla_sentimientos.py (a partir de la línea 711). Estas particiones se han realizado por experimentación, arbitrariamente y podrían ser otras editando el código:

```
# Sentimiento medio combinado literal
for val in medias_sent_comb:
    if val <= 0.0001 or val == 0.0:
        medias_comb_lit.append('Muy malo')
    elif val <= 0.002:
        medias_comb_lit.append('Malo')
    elif val <= 0.005:
        medias_comb_lit.append('Regular')
    elif val <= 0.006:
        medias_comb_lit.append('Bueno')
    elif val <= 100:
        medias_comb_lit.append('Muy bueno')
```

Esto se traduce en una conversión no lineal de acuerdo al siguiente gráfico de la figura 43:

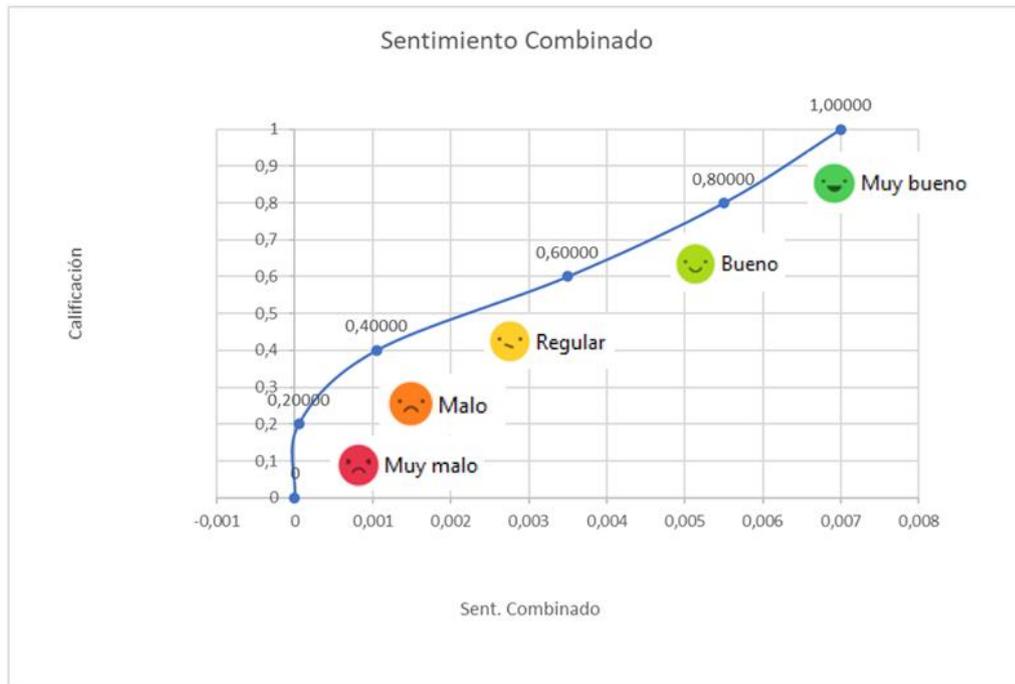


Figura 43. Correspondencia entre el Sentimiento Combinado y su traducción literal (Particiones).

4.4.5.2. *Tabla de Sentimientos*

Pulsando el botón *Tabla de Sentimientos* del módulo de *Análisis de Sentimientos* accedemos a una tabla como la de la figura 44.

Análisis de Sentimientos del proyecto 1

Tabla de Sentimientos | Mostrar Gráficas | Sentimientos con Iconos | Sentimiento Medio | Ayuda

Tabla de mensajes --> Desplazar las barras de scroll inferior y lateral para ver toda la información.

	Emisor	Alias	Destinatario	Sentim. Analys. Sp.	Sentim. Analys. Sp. Norm.	Sent. N. B	PySentim	Sent. Combinad	Mensaje	Foro	Hilo
30	23025	u9	79	3e-10	3e-10	pos	POS	0.0042	buenas tardes	1	11
37	79	u1	22946	3.1e-08	3.1e-08	pos	NEU	0.0042	hola pedro	1	74
38	79	u1	71704	2.2e-25	2.2e-25	pos	NEU	0.0042	buenos dias adrian	1	11
39	69829	u8	79	0.012	0.012	pos	NEU	0.0047	hola	1	11
40	22946	u4	79	0.47	0.47	neg	NEU	0.0054	no desde el principio bueno	1	74
41	25055	u9	79	1.5e-70	1.5e-70	pos	NEU	0.0042	siguiendo el acercamiento e	1	10
42	66182	u2	79	1.00	1.00	neg	NEU	0.007	analysis of moodle analytics	1	10
43	39712	u6	79	8.5e-36	8.5e-36	pos	NEU	0.0042	buenas tardes envio mi anal	1	10
44	39712	u6	Inicia el h	7e-13	7e-13	pos	NEU	0.0042	buenos dias antonio	1	14
45	25964	u7	23709	4e-16	4e-16	pos	NEU	0.0042	hola a todos para conocer	1	51
46	71704	u5	79	1.6e-112	0	pos	NEU	0.0042	epistemologia que estamos	1	10
47	79	u1	39712	5.4e-05	5.4e-05	neg	NEU	0.002	buenos dias maria	1	14
48	66967	u14	79	6.4e-16	6.4e-16	pos	NEU	0.0042	buenas tardes	1	51
49	66967	u14	Inicia el h	2.6e-11	2.6e-11	pos	NEU	0.0042	buenas tardes	1	14
50	71704	u5	Inicia el h	1.1e-26	1.1e-26	pos	NEU	0.0042	hola	1	14
51	66182	u2	25055	4.2e-07	4.2e-07	pos	NEU	0.0042	hola miguel	1	10
52	69829	u8	Inicia el h	1.8e-21	1.8e-21	pos	NEU	0.0042	hola	1	15
53	69829	u8	71704	1.3e-10	1.3e-10	pos	NEU	0.0042	buenas noches	1	14
54	69829	u8	66967	3.4e-10	3.4e-10	pos	NEU	0.0042	buenas noches	1	14

122 rows x 11 columns

hola

a mi me resulta muy util la pagina https://www3schools.com/python como recurso para empezar desde cero pero tambien como fuente de consulta rapida con ejemplos para implementar cosas concretas pudiendolos usar como base para adaptarlo luego con nuestras necesidades especificas ademas a la hora de estar programando suele venir muy bien tener a mano hojas de resumen como las disponibles en https://www.tutoria1foret.com/python/cheatsheet.pdf un saludo

Figura 44. Tabla de Sentimientos.

La información de las diferentes columnas de la tabla de sentimientos se puede ver al seleccionar cada celda de la tabla en la parte inferior de la pantalla y es la siguiente:

- **Emisor:** Nombre del emisor del mensaje de cada fila.
- **Alias:** Alias asignado automáticamente por la aplicación para acortar los nombres de usuario y anonimizarlos.
- **Destinatario:** Destinatario del mensaje representado en cada fila.
- **Sentim. Analys. Sp.:** Valor devuelto por el algoritmo implementado en la librería [sentiment-analysis-spanish](#).
- **Sentim. Analys. Sp. Norm.:** Valor anterior normalizado [descrito más arriba](#). Se muestra con fondo verde.
- **Sent. N. Bayes:** Valor devuelto por el algoritmo Naive Bayes utilizando el corpus . Este algoritmo devuelve *neg* para mensajes negativos y *pos* para mensajes clasificados como positivos. Se muestra con fondo marrón.
- **PySentimiento:** Valor devuelto por el algoritmo implementado en la librería [PySentimiento](#). Este algoritmo devuelve tres valores posibles: *NEG*, *POS* y *NEU* para los mensajes clasificados como negativos, positivos y neutros respectivamente. Se muestra con fondo azul.
- **Sent. Combinado:** Valor de [Sent. Comb.](#) para el mensaje representado en esa fila. Se muestra con fondo amarillo.
- **Mensaje:** Se transcribe el mensaje representado en esa fila. Al seleccionar un mensaje se puede ver su contenido íntegro en la parte inferior de la pantalla.
- **Foro:** Nombre del foro donde se publicó el mensaje.
- **Hilo:** Hilo en el que se publicó el mensaje dentro del foro.

Las columnas que contienen los valores relevantes del análisis son las que se muestran con fondos de color.

- **Sentim. Analys. Sp. Norm.:** verde.
- **Sent. N. Bayes:** marrón.
- **PySentimiento:** azul.
- **Sent. Combinado:** amarillo.

Además de lo indicado anteriormente, los valores numéricos máximos de cada columna se muestran con fondo de color verde claro.

4.4.5.3. Gráficos de Sentimientos

Al pulsar el botón Mostrar Gráficos se muestran los gráficos de la pantalla de la figura 45.

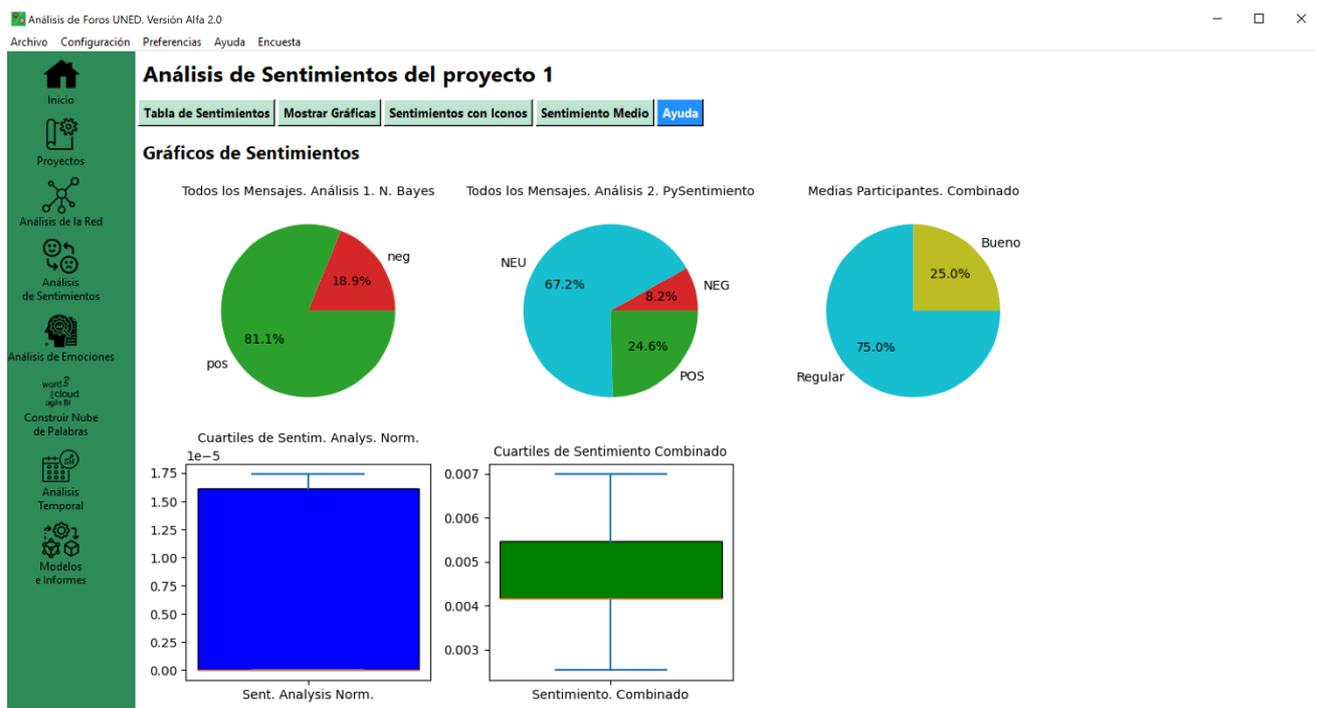


Figura 45. Gráficos de Sentimientos.

En esta pantalla se muestran tres gráficos de tarta y dos gráficos de caja.

Gráficos de tarta

- Todos los Mensajes Análisis 1. N. Bayes
 - o Se representa la proporción de mensajes positivos y negativos según este algoritmo.
- Todos los Mensajes. Análisis 2. PySentimiento
 - o Se representa la proporción de mensajes positivos, negativos y neutros según esta librería.

Gráficos de caja

- Cuartiles de Sentym. Analys. Norm.
 - o Se muestra un gráfico de caja para esta variable donde se pueden ver los cuartiles, incluida la mediana.
- Cuartiles de Sentimiento Combinado
 - o Se muestra un gráfico de caja para esta variable donde se pueden ver los cuartiles, incluida la mediana.

4.4.5.4. Sentimientos mediante iconos y diferentes análisis

Al pulsar el botón *Sentimientos con Iconos* se muestra la tabla de la figura 46, la cual visualiza el sentimiento de cada mensaje.

Icono	Alias	Emisor	Destinatario	Sent. Analysis Sp	Sentimiento NB	PySentimiento	Sent. Combinado	Mensaje	Titulo	Foro	Hilo
	u5	71704	69829	5.60032120028535	pos	POS	0.00583333333333	hola inma	Re: Propuesta Mo	1	14717
	u4	22946	79	1.27701675827268	pos	NEU	0.00416666666666	muchas gracias p analizaremos el si	Re: Acercamiento	1	10981
	u4	22946	25055	6.90037476460522	pos	NEG	0.00257486738054	comentario sobre me resulta interes	Re: Acercamiento	1	10981
	u4	22946	71704	1.20367739899153	pos	NEU	0.00416840136937	comentario sobre tambien creo que	Re: Acercamiento	1	10981
	u5	71704	Inicia el hilo	1.31257192100502	neg	NEG	0.00036666666666	hola	Tarea Tema 4 y fo	1	24137
	u8	69829	71704	1.22386062645726	pos	NEU	0.00416684195240	hola adrian pues: un saludo inma	Re: Tarea Tema 4:	1	24137
	u8	69829	Inicia el hilo	3.24546606914601	pos	NEU	0.00416951510749	buenas tardes	Examen de febrer	1	24316
	u8	69829	79	4.93275238738135	pos	POS	0.00583333333333	hola adrian pues: como alternativa	Re: Respuesta a la	1	19664
	u7	25964	79	0.00040011019661	pos	NEU	0.00426668044025	hola miguel	Re: Respuesta a la	1	19664
	u5	71704	25055	2.09861481997514	pos	NEU	0.004166666668957	hola miguel	Re: Acercamiento	1	10981

TEXTO DEL MENSAJE:
hola inma
muchas gracias por tu comentario y perdona por no responder antes pero se me habia pasado
en primer lugar me parece muy importante tener en cuenta la llamada de atencion que haces sobre la parte de proteccion de datos relativa a la propuesta de aplicacion ya que probablemente es uno de los aspectos mas delicados de esta propuesta al basarse en el uso intensivo de datos personales muy sensibles como la localizacion y las preferencias por distintos tipos de ocio y actividades
en cuanto a incluir la sugerencia de compra de distintos productos creo que seria una funcionalidad totalmente diferente de la que se propone y muchos usuarios podrian ver esto como simple publicidad y provocarlos un fuerte rechazo con lo que en mi opinion no seria adecuado incluir esta funcionalidad en la misma aplicacion
sin embargo me parece que seria una idea muy interesante para una aplicacion independiente ya que de esta modo los usuarios realmente interesados en oferta

Figura 46. Tabla de análisis de sentimientos de cada mensaje con iconos.

La información representada en esta vista es exactamente la misma que se muestra en la [tabla de sentimientos](#) y la existencia de ambas se debe cuestiones técnicas de desarrollo, en tanto en cuanto en esta se muestra un objeto *Treeview* que permite introducir iconos de manera relativamente sencilla, cosa que no sucede con un objeto *Table*, que es el que se visualiza en otro caso. Por contrapartida resulta más sencillo capturar el contenido de una celda para representar su contenido en otro campo. Es por ello que en esta vista con iconos, sólo se pueden seleccionar filas y no celdas. Al seleccionar una fila se representa en la ventana de texto inferior el contenido íntegro del mensaje correspondiente a dicha fila.

4.4.6. Análisis de Emociones

Esta aplicación realiza dos análisis de emociones; uno basado en las conocidas como regiones de Russell y otro basado en la asignación a palabras de una puntuación para diferentes emociones primarias. Ambos son métodos estadísticos en el sentido de que parten de estudios previos en los que se ha pedido a un número relativamente alto de participantes que etiqueten (puntuén) un gran número de términos. Además, se pueden considerar métodos aritméticos simples, en el sentido de que se realizan operaciones elementales sobre las puntuaciones de los términos que aparecen en los mensajes (adición). No obstante lo anterior, al primer análisis se le ha denominado *de regiones de Russell* y al segundo *análisis aritmético*.

Mediante el botón *Análisis de Emociones* se accede al módulo de análisis de emociones, cuya primera vista es la de la figura 47, que muestra las Emociones promedio de cada participante según Regiones de Russell.

4.4.6.1. Emociones promedio de cada participante según Regiones de Russell.

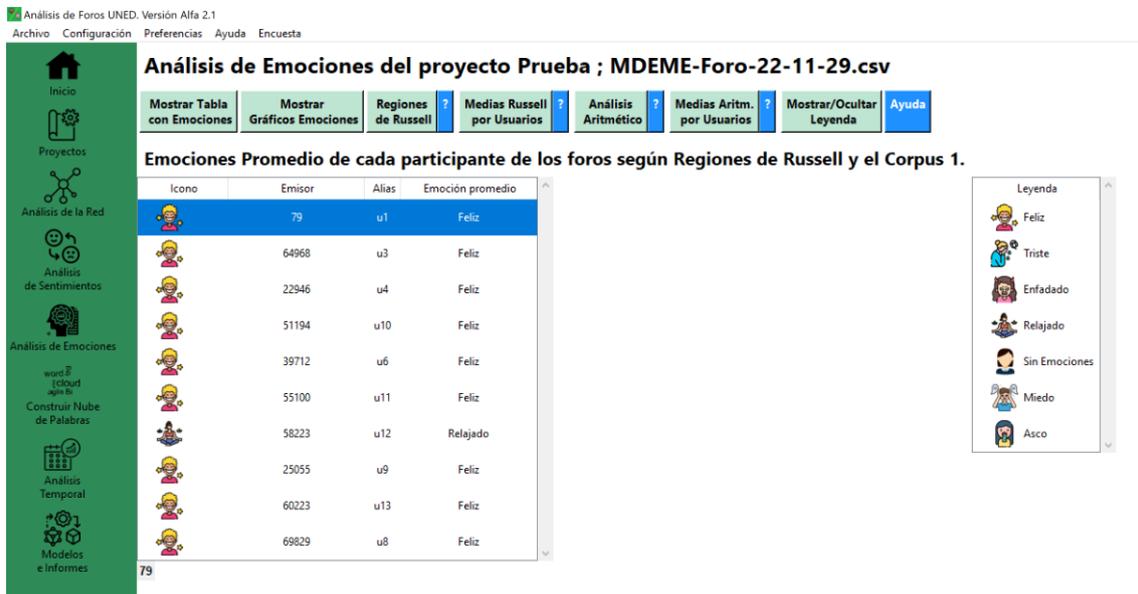


Figura 47. Emociones promedio.

En esta pantalla se puede ver una tabla en la que se puede hacer *scroll* y en la que se muestra para cada usuario un icono representativo de la emoción. El identificador de usuario se muestra en la columna Emisor. En la columna Alias se representa el alias asignado automáticamente por la aplicación y en la columna Emoción Promedio aparecen las emociones promedio de cada participante de manera literal.

Estas emociones pueden ser: Feliz, Triste, Enfadado, Relajado, Sin Emoción, Miedo y Asco (Rechazo o Desprecio).

Al seleccionar una fila se muestra en la parte inferior de la pantalla el nombre de usuario completo, dado que puede ser que no quepa en el ancho por defecto de la columna, si bien el ancho de las columnas se puede redimensionar.

Para el cálculo de estas emociones se utiliza un método aritmético de bolsa de palabras y un corpus referido como corpus 1 que puede ser cambiado por el usuario. Por defecto se utiliza como corpus el archivo ubicado en la ruta:

./recursos/ Spanish-es-NRC-VAD-Lexicon.txt

El formato de este archivo se muestra en la figura 48.

Word	Spanish-es	Valence	Arousal	Dominance
aaaaaaah	aaaaaaah	0.479	0.606	0.291
aaaah	NO TRANSLATION	0.520	0.636	0.282
aardvark	cerdo hormiguero	0.427	0.490	0.437
aback	detrás	0.385	0.407	0.288
abacus	ábaco	0.510	0.276	0.485
abalone	abulón	0.500	0.480	0.412
abandon	abandonar	0.052	0.519	0.245
abandoned	abandonado	0.046	0.481	0.130
abandonment	abandono	0.128	0.430	0.202
abashed	avergonzado	0.177	0.644	0.307
abate	disminuir	0.255	0.696	0.604
abatement	disminución	0.388	0.338	0.336
abba	Abba	0.562	0.500	0.480
abbey	abadía	0.580	0.367	0.444
abbot	NO TRANSLATION	0.427	0.321	0.483
abbreviate	abreviar	0.531	0.375	0.330
abbreviation	abreviatura	0.469	0.306	0.345
abdomen	abdomen	0.469	0.462	0.471
abdominal	abdominal	0.490	0.456	0.445
abduct	secuestrar	0.173	0.720	0.615
abduction	secuestro	0.062	0.990	0.673
aberrant	aberrante	0.146	0.765	0.431
aberration	aberración	0.125	0.816	0.417
abeyance	suspensión	0.330	0.510	0.292
abhor	aborrecer	0.125	0.602	0.349
abhorrence	aborrecimiento	0.167	0.684	0.420
abhorrent	aborrecible	0.229	0.750	0.474
abide	acatar	0.635	0.354	0.705
abiding	permanencia	0.796	0.327	0.750
ability	capacidad	0.875	0.510	0.816

Figura 48. Contenido parcial del archivo utilizado como corpus por defecto para el cálculo de las emociones según Regiones de Russell.

4.4.6.2. Tabla de Emociones

Mediante el botón *Mostrar Tabla con Emociones* del módulo de emociones, accedemos a una tabla como la que se muestra en la figura 49.

Análisis de Emociones del proyecto Prueba ; MDEME-Foro-22-11-29.csv

Mostrar Tabla con Emociones | Mostrar Gráficos Emociones | Regiones de Russell | Medias Russell por Usuarios | Análisis Aritmético | Medias Aritm. por Usuarios | Mostrar/Ocultar Leyenda | Ayuda

Tabla de mensajes -----> Desplazar la barra inferior para ver toda la información ----->

	Emisor	Alias	Destinatario	Emoción Reg	Felicidad	Enojo	Tristeza	Miedo	Asco	Dominante	Mensaje	Foro	Hilo
1	79	u1	Inicia el hilo	Feliz	32.23	16.83	19.43	21.93	16.80	Felicidad	buenos dias	1	5193
2	64968	u3	79	Feliz	10.67	6.05	6.29	6.86	6.21	Felicidad	hola a todos	1	5193
3	22946	u4	79	Feliz	2.87	2.37	2.43	2.37	2.27	Felicidad	hola	1	5193
4	51194	u10	79	Feliz	1.33	1.10	1.17	1.07	1.07	Felicidad	hola	1	5193
5	39712	u6	51194	Feliz	3.70	2.77	2.47	2.80	2.47	Felicidad	hola a todos	1	5193
6	55100	u11	79	Feliz	23.90	12.40	13.80	14.50	12.67	Felicidad	buenos dias	1	5193
7	58223	u12	55100	Feliz	12.40	6.70	7.93	8.10	6.53	Felicidad	buenos dias	1	5193
8	25055	u9	79	Feliz	17.33	7.47	7.90	9.17	7.53	Felicidad	hola a todos	1	5193
9	60223	u13	79	Feliz	5.07	2.13	2.37	2.57	2.47	Felicidad	buenas me p	1	5193
10	79	u1	Inicia el hilo	Feliz	4.63	1.43	1.30	1.83	1.17	Felicidad	buenas tarde	1	7470
11	55100	u11	79	Ninguna	0	0	0	0	0	N/A	buenas tarde	1	7470
12	25055	u9	79	Ninguna	0	0	0	0	0	N/A	hola	1	7470
13	69829	u8	79	Feliz	9.57	3.63	4.53	4.33	3.37	Felicidad	hola	1	5193
14	69829	u8	79	Ninguna	0	0	0	0	0	N/A	hola	1	7470
15	51194	u10	79	Triste	1.23	2.33	3.10	3.43	1.63	Miedo	hola	1	7470
16	64968	u3	79	Triste	1.87	1.40	1.40	1.70	1.37	Felicidad	hola	1	7470
17	66182	u2	79	Ninguna	0	0	0	0	0	N/A	buenas tarde	1	7470
18	66182	u2	79	Feliz	10.03	5.03	5.33	6.23	4.77	Felicidad	hola a todos	1	5193
19	39712	u6	79	Triste	2.30	1.53	2.07	2.13	1.23	Felicidad	buenas tarde	1	7470

Figura 49. Tabla de emociones.

En esta tabla se representan todos los mensajes contenidos en el archivo cargado. Para cada mensaje se puede ver el Emisor, el alias asignado automáticamente a dicho emisor, el usuario destinatario del mensaje, la emoción según la región de Russell (primer análisis de emociones con fondo color mostaza), y las componentes del segundo análisis de emociones junto con su dominante (columna de fondo verde).

A continuación de las columnas mencionadas se muestran sendas columnas con el contenido del mensaje, el Foro en el que se publicó y el Hilo. Si seleccionamos una celda de la tabla su contenido se muestra en la parte inferior de la pantalla. Ver figura 50.

17	66182	u2	79	Ninguna	0	0	0	0	0	N/A	buenas tarde	1	7470
18	66182	u2	79	Feliz	10.03	5.03	5.33	6.23	4.77	Felicidad	hola a todos	1	5193
19	39712	u6	79	Triste	2.30	1.53	2.07	2.13	1.23	Felicidad	buenas tarde	1	7470

122 rows x 13 columns

buenas tardes
a mi me viene mejor el martes pero si se decide el lunes creo que podre cambiar lo que tengo si lo se con tiempo
muchas gracias por organizarlo saludos

Figura 50. Contenido de un mensaje al pie de la tabla de emociones.

4.4.6.3. Gráficos de Emociones

Al pulsar el botón Mostrar Gráficos Emociones se muestra una pantalla como la que se ve en la figura 51.

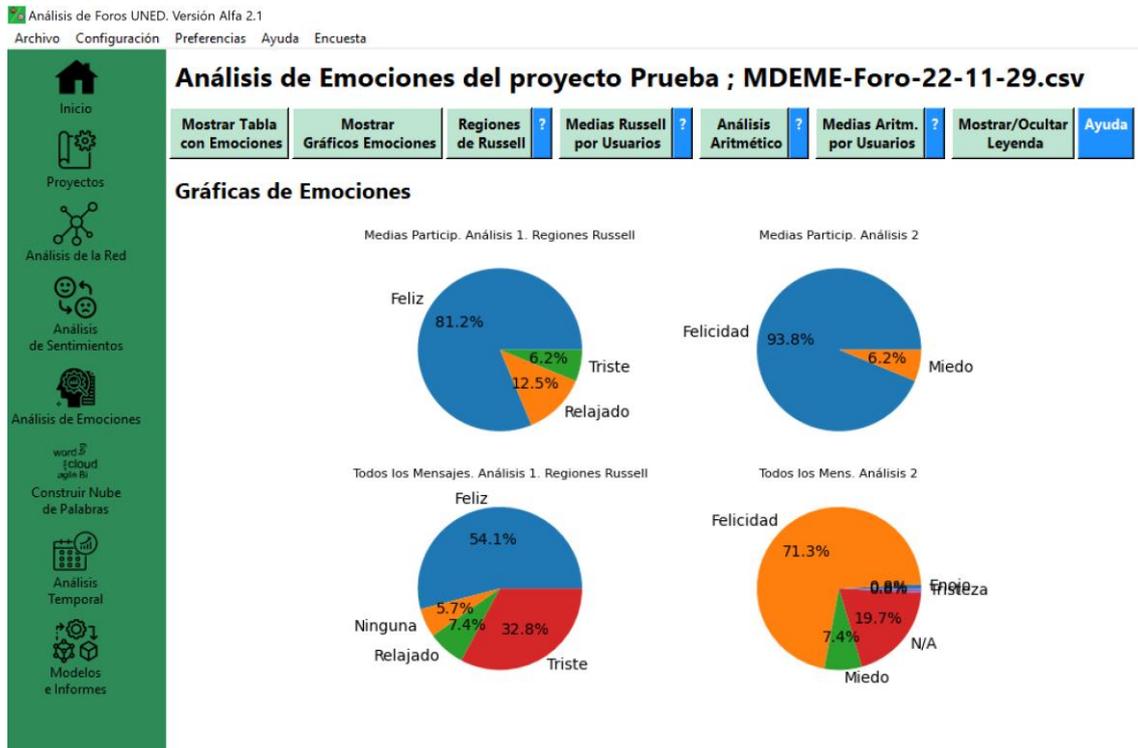


Figura 51. Gráficos de Emociones.

En esta pantalla se muestran cuatro gráficos de tarta:

- Medias de los Participantes según el análisis de Regiones de Russell.
- Medias de los Participantes según el análisis puramente aritmético.
- Proporción de cada región de Russell entre todos los mensajes de los foros.
- Proporción de cada emoción dominante según el análisis 2 entre todos los mensajes de los foros.

4.4.6.4. *Región de Russell para cada mensaje*

Al pulsar el botón **Regiones de Russell** se muestra una pantalla como la que se ve en la figura 52. La región de Russell se halla a partir de un corpus de palabras a las que se le ha asignado un valor de excitación y un valor de valencia (figura 48), si bien puede utilizarse otro archivo con el mismo formato que elija el usuario. Se busca cada palabra del mensaje en el corpus y se realizan los sumatorios de valencia y excitación para cada mensaje. La región de Russell se halla a partir de dichos valores de valencia y excitación.

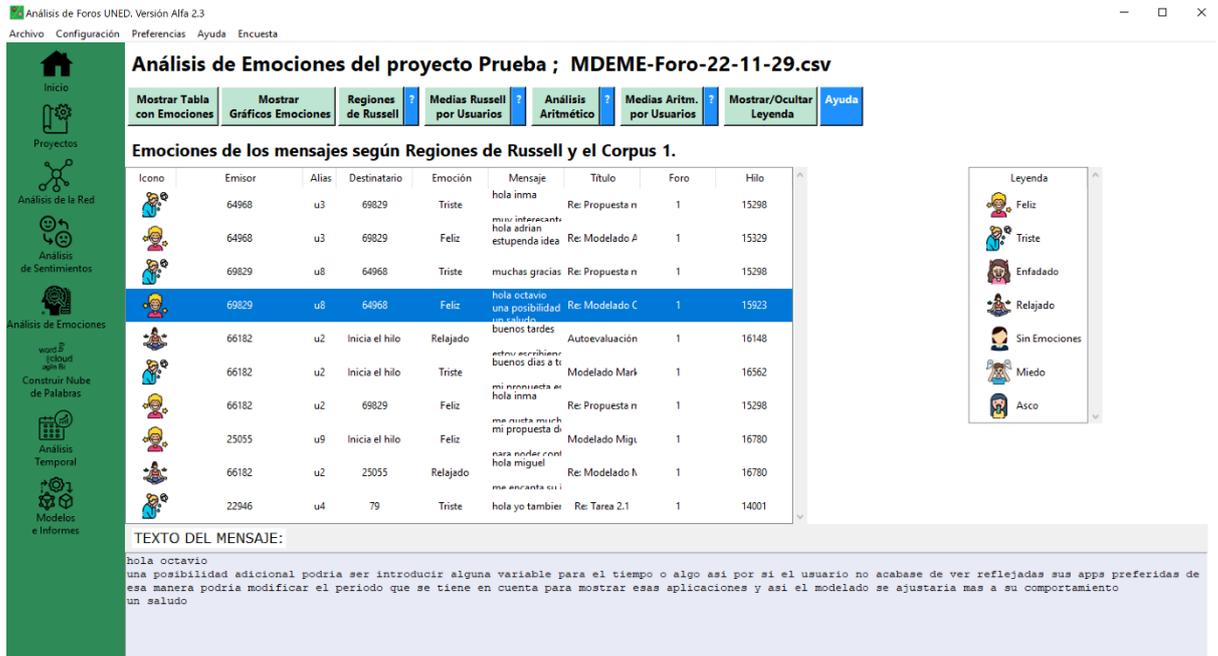


Figura 52. Análisis de regiones de Russell de cada mensaje.

El método utilizado en esta aplicación es adoptado de la aplicación presentada en *Redes de Innovación Docente*. Es muy simple y se basa en la función arco-tangente. Dicho método se reproduce a continuación.

```
def getRusselRegion(self, arousal, valence):
    """ Function to get the Russell Region from arousal and valence """

    russel = {
        270: C_RELAX,
        180: C_SAD,
        90: C_ANGRY,
        0: C_HAPPY
    }

    a = math.atan2(arousal - 5, valence - 5)
    mydegrees = math.degrees(a)
    deg = mydegrees if mydegrees > 0 else mydegrees + 360
    # print ("Grado:",deg)
    result = ''
    for x in russel.keys():
        if deg > x:
            result = russel[x]
            break
    if valence == 0:
        result = C_NONE
    # print ("Resultado:",result)
    return result
```

4.4.6.5. Análisis Aritmético

Este análisis de emociones se realiza a partir de un corpus de términos etiquetados (puntuados) obtenido del trabajo de Ferré et al, 2017 [34]. Consiste en una lista de términos para cada uno de los cuales se les ha asignado una puntuación asociada a cada una de cinco emociones consideradas primarias (felicidad, enfado, tristeza, miedo y rechazo). Para cada mensaje se buscan sus términos en la lista y se lleva la cuenta de cada emoción primaria. Al terminar el procesamiento se muestra la puntuación obtenida para cada mensaje y cada una de las cinco emociones. La emoción que obtiene mayor puntuación se considera la **dominante**, aunque pueden consultarse las puntuaciones de todas las emociones. Este análisis se muestra en la figura 53.

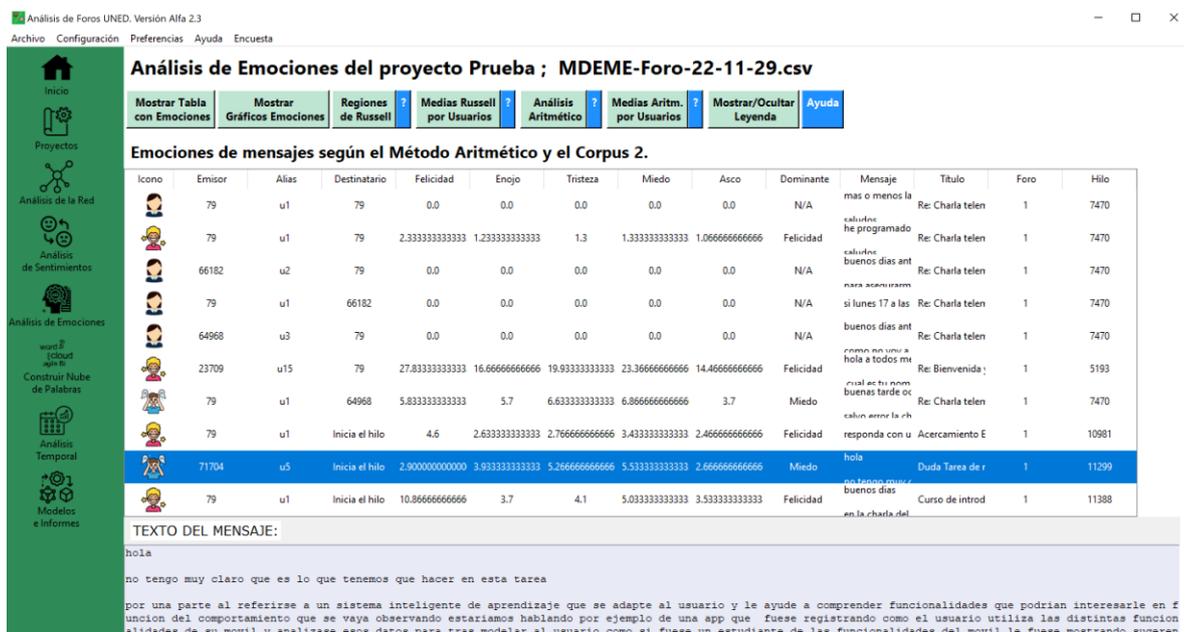


Figura 53. Análisis llamado *aritmético*.

4.4.6.6. Medias Aritméticas por Usuarios

Se trata de obtener las emociones del método descrito en el punto anterior para todos los mensajes emitidos por cada usuario y después hallar las medias aritméticas de cada emoción primaria. De igual modo al descrito anteriormente, se indica la emoción dominante. En la figura 54 se puede ver la tabla de dichas medias y las emociones dominantes para cada usuario.

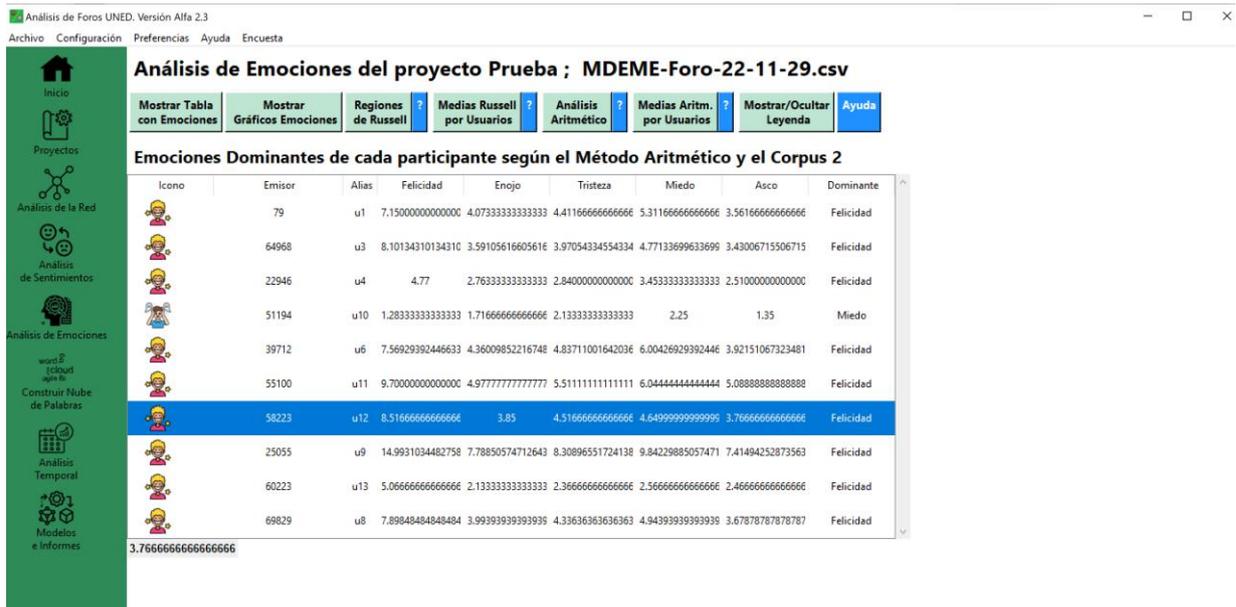


Figura 54. Emociones medias para los participantes según el segundo método aritmético.

4.4.7. Nube de Palabras

La nube de palabras es una manera muy intuitiva de mostrar a golpe de vista cuáles son las palabras más utilizadas en un documento (en este caso en los foros), cuya utilidad está contrastada.

La aplicación dispone de la posibilidad de crear una nube de palabras a través del botón *Construir Nube de Palabras*. Ver figura 55.

Esta funcionalidad no existía en el trabajo anterior y se ha tomado la idea del trabajo de referencia indicado al inicio, aunque se utiliza otro código y es un recurso ampliamente utilizado.

Para esta funcionalidad se utiliza la librería *wordcloud*. La nube de palabras se construye teniendo en cuenta ciertos parámetros y una lista de palabras no relevantes (*stopwords*). El método encargado de construir la nube de palabras es `crearNubePalabras` que crea un objeto `WordCloud` y lo muestra en un nuevo *frame* de la clase `PantallaNubePalabras`.

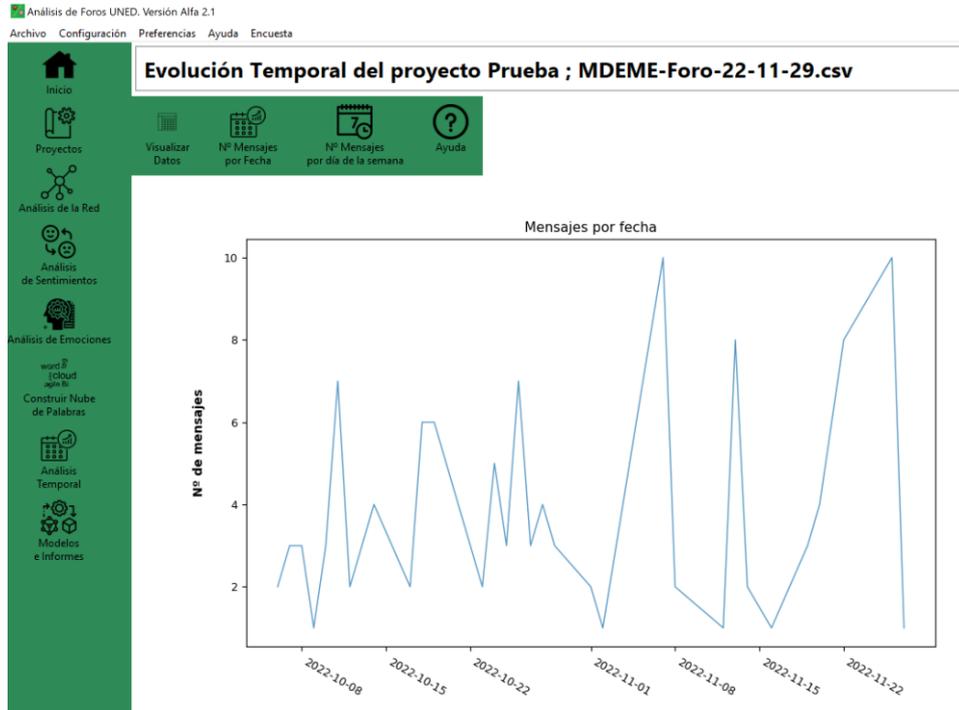


Figura 56. Mensajes por fecha en el módulo de análisis temporal.

4.4.8.2. *Número de mensajes por día de la semana*

Dentro del módulo de Análisis Temporal, pulsando el botón de *Nº Mensajes por día de la semana* se accede a una gráfica como la de la figura 57.

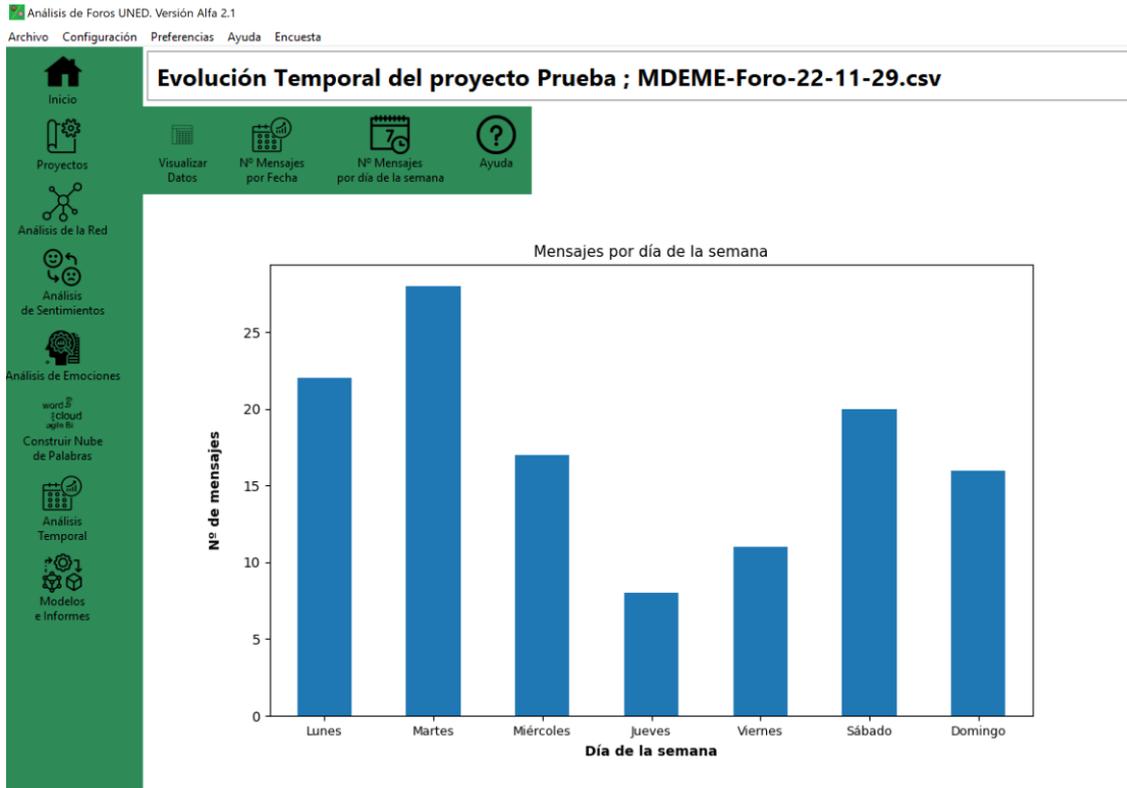


Figura 57. Mensajes por día de la semana.

4.4.8.3. Visualizar Datos temporales

Si dentro del módulo de Análisis Temporal pulsamos el icono del menú superior con la inscripción *Visualizar Datos* accedemos a una tabla como la de la figura 58, en la que se ve el número de mensajes para cada fecha y el día de la semana de dicha fecha. Estos son los datos a partir de los que se muestran las dos gráficas vistas en los apartados anteriores.

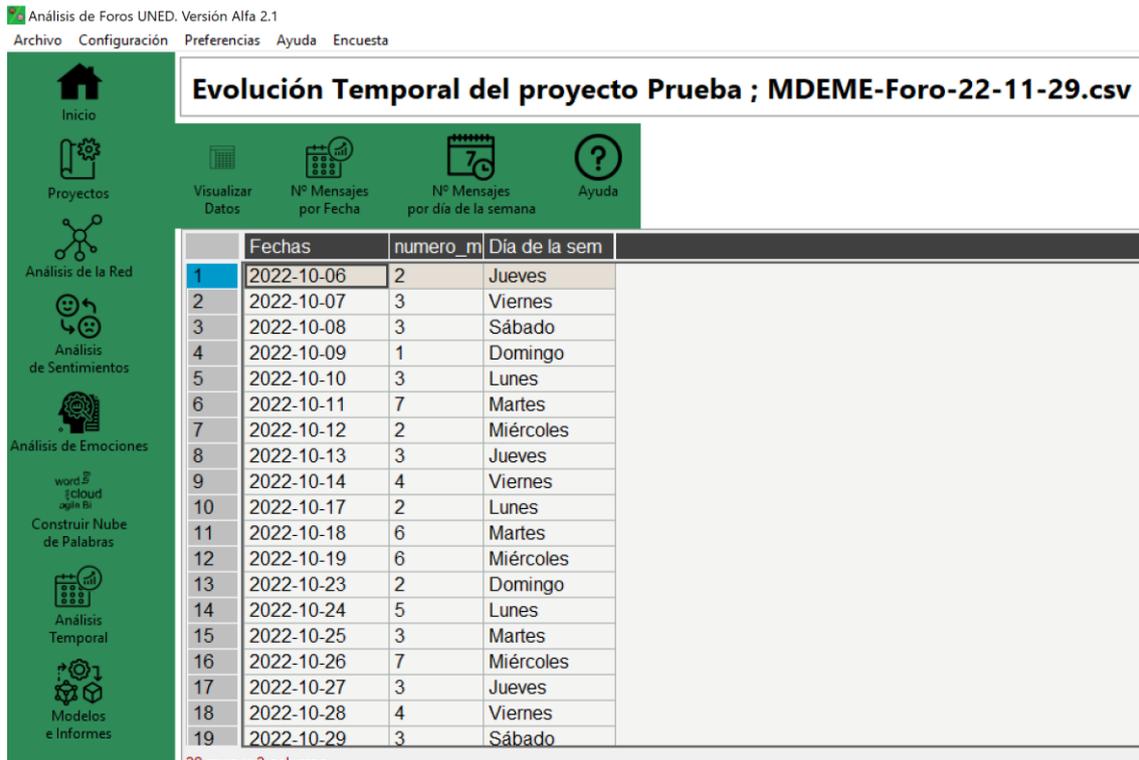


Figura 58. Tabla de datos temporales.

4.4.9. Modelos e Informes

Pulsando el botón *Modelos e Informes* se accede al módulo del mismo nombre. La pantalla “Modelos e Informes” constituye una buena parte del trabajo realizado en la parte práctica de este TFM.

Al entrar al módulo se muestra en primer lugar la pantalla que se muestra en la figura 55, donde se ve una tabla con toda la información disponible sobre los participantes de los foros. Desde este módulo también se pueden realizar gráficas de agrupamiento (*clustering*) y **gráficas de usuarios**, así como **exportar la información de los usuarios a Excel y generar informes automáticamente**.

4.4.9.1. Datos de los participantes

Dentro del módulo de Modelos e Informes, pulsando el botón *Visualizar Datos* accedemos a una tabla como la de la figura 59.

Analisis de Foros UNED, Versión Alfa 3.0
 Archivo Configuración Preferencias Ayuda Encuesta

Modelos del proyecto Prueba ; MDEME-Foro-22-11-29.csv
 Mensajes: 122 Participantes: 16

Visualizan Datos Algoritmo de Clustering Gráficos de Usuarios Generar Informe PDF Exportar a Excel Ayuda

Información de cada participante en los foros --> Desplace la barra inferior para ver todas las columnas. -->

	Participan	Alias Usua	N Mensaj	Hilos Inic	Resp. A O	A sí mism	A Sus Hil	A Otros Hi	Grado	Grado Ent	Grado Sal	Centralid	Cercanía	Int
1	79	u1	20	5	13	7	10	10	25	16	9	1.67	1.00	0.0
2	66182	u2	11	2	9	2	4	7	7	3	4	0.47	0.56	0.0
3	64968	u3	6	1	5	1	1	5	5	3	2	0.33	0.56	0.0
4	22946	u4	10	1	9	1	2	8	7	4	3	0.47	0.58	0.0
5	71704	u5	23	3	20	3	5	18	14	5	9	0.93	0.6	0.0
6	39712	u6	7	2	5	2	3	4	6	3	3	0.4	0.56	0.0
7	25964	u7	5	1	4	1	1	4	6	3	3	0.4	0.56	0.0
8	69829	u8	22	2	20	2	6	16	13	5	8	0.87	0.6	0.0
9	25055	u9	5	1	4	1	1	4	5	4	1	0.33	0.58	0.0
10	51194	u10	2	0	2	0	0	2	2	1	1	0.13	0.38	0.0
11	55100	u11	3	0	3	0	0	3	2	1	1	0.13	0.067	0.0
12	58223	u12	2	0	2	0	0	2	2	0	2	0.13	0	0.0
13	60223	u13	1	0	1	0	0	1	1	0	1	0.067	0	0.0
14	66967	u14	3	1	2	1	2	1	4	2	2	0.27	0.42	0.0
15	23709	u15	1	0	1	0	0	1	2	1	1	0.13	0.38	0.0
16	23778	u16	1	0	1	0	0	1	1	0	1	0.067	0	0.0

16 rows x 32 columns

Figura 59. Tabla con información de los participantes de los foros.

En esta tabla se pueden ver las diferentes variables que se evalúan, calculan y visualizan en el resto de módulos. Cada fila corresponde a un participante. La primera columna muestra el identificador de usuario. La segunda columna muestra el alias asignado automáticamente por la aplicación a cada usuario. A continuación aparece la siguiente información:

Columnas de la 3ª a la 8ª

Se muestra información relativa al número de mensajes publicados.

- *N.Mensajes* (Número total de mensajes)
- *Hilos iniciados* (Número de hilos iniciados)
- *Resp. A Otros Autores* (Número de mensajes enviados como respuesta a mensajes de otros autores)
- *A sí mismos + inicios* (Suma del número de mensajes publicados en respuesta a sus propios mensajes más mensajes iniciando hilos)
- *A sus Hilos* (Número de mensajes publicados en hilos iniciados por el mismo usuario)
- *A Otros Hilos* (Número de mensajes publicados en hilos no iniciados por el propio usuario)

Columnas de la 9ª a la 18ª

Se muestra información relativa a medidas de centralidad y otras estadísticas de la red. Ver la ayuda del módulo de análisis de Red.

- *Grado*
- *Grado Ent.*
- *Grado Sal.*
- *Centralidad Grado*
- *Cercanía*
- *Intermediación*
- *Cent. Vect. Propio*
- *Page Rank*
- *Iniciativa*
- *Modularidad*

Columnas de la 19ª a la 23ª

Se muestra información relativa a al análisis de sentimientos. Ver la ayuda del módulo de análisis de sentimientos.

- *Sent. Analysis*
- *Sentimiento N.B*
- *Py Sentimiento*
- *Sent. Comb*
- *Sent. Comb Lit.*

Columnas de la 24ª a la 32ª

Se muestra información relativa al análisis de emociones. Ver ayuda del módulo de análisis de sentimientos.

- *Valencia Media*
- *Excit. Media*
- *Región Russell*
- *Felicidad*
- *Enfado*
- *Tristeza*
- *Miedo*
- *Disgusto*
- *Dominante*

4.4.9.2. Agrupamiento (Clustering)

Pulsando el botón *Algoritmo de* visualización de un gráfico similar al de la



Clustering se accede a la figura 60.

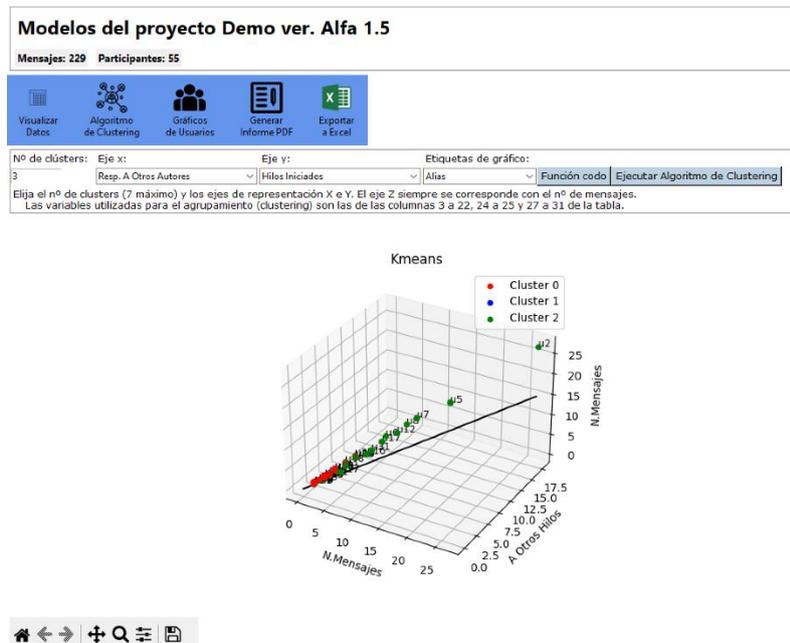


Figura 60. Gráfico de agrupamiento de usuarios (clustering).

Al acceder a la función de agrupamiento se genera un gráfico de agrupamiento inicial. En esta ejecución inicial del algoritmo de *clustering* se agrupan los usuarios como pertenecientes a tres categorías o grupos. El algoritmo utilizado es K-medias (*Kmeans*)²⁶.

Se trata de una representación espacial en la que el eje X se corresponde con el número de mensajes publicados como respuestas a otros autores; el eje Y se corresponde con el número de hilos iniciados y el eje Z con el número de mensajes.

Las variables utilizadas para ejecutar el algoritmo de *clustering* son las columnas 3 a 22, 24 a 25 y 27 a 31 de la tabla.

²⁶ https://scikit-learn.org/stable/modules/generated/sklearn.cluster.k_means.html

Configuración del agrupamiento

En la parte superior, bajo los botones del módulo se encuentran unos campos de configuración. Mediante esta configuración se puede elegir:

- Número grupos (*clusters*). Pueden ser 7 como máximo.
- Eje X de la representación espacial.
- Eje Y de la representación espacial.
- Etiquetas a mostrar junto a cada punto (usuarios, alias generados automáticamente por la aplicación o ninguna etiqueta).

Una vez configurado el algoritmo del agrupamiento puede ejecutarse y dibujar el gráfico pulsando el botón *Ejecutar Algoritmo de Clustering*.

Función codo

Dentro del apartado de *clustering*, si se pulsa el botón *Función codo* se muestra un gráfico como el de la figura 61. El número de clusters idóneo según este método coincide con la zona que se asemeja a un codo. En este caso estaría entre 2 y 3.

Nº de clústers:	Eje x:	Eje y:	Etiquetas de gráfico:	
3	Resp. A Otros Autores	Hilos Iniciados	Alias	Función codo Ejecutar Algoritmo de Clustering

Elija el nº de clusters (7 máximo) y los ejes de representación X e Y. El eje Z siempre se corresponde con el nº de mensajes.
Las variables utilizadas para el agrupamiento (clustering) son las de las columnas 3 a 22, 24 a 25 y 27 a 31 de la tabla.

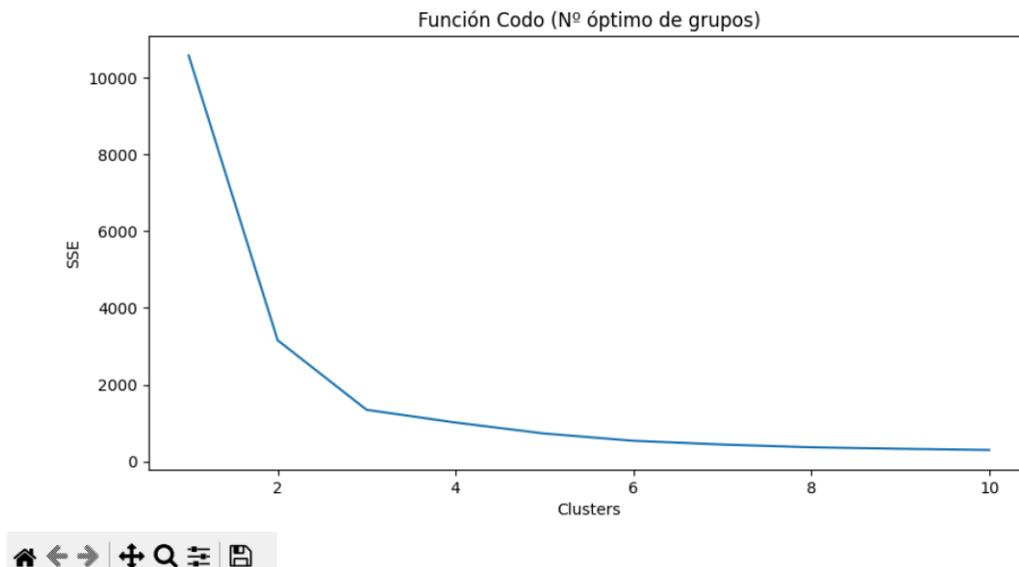


Figura 61. Función codo que facilita la elección del número de grupos (clusters).

4.4.9.3. Gráficos de Usuarios

Al pulsar el botón *Gráficos de Usuarios* se accede a una pantalla como la que se ve en la figura 62.

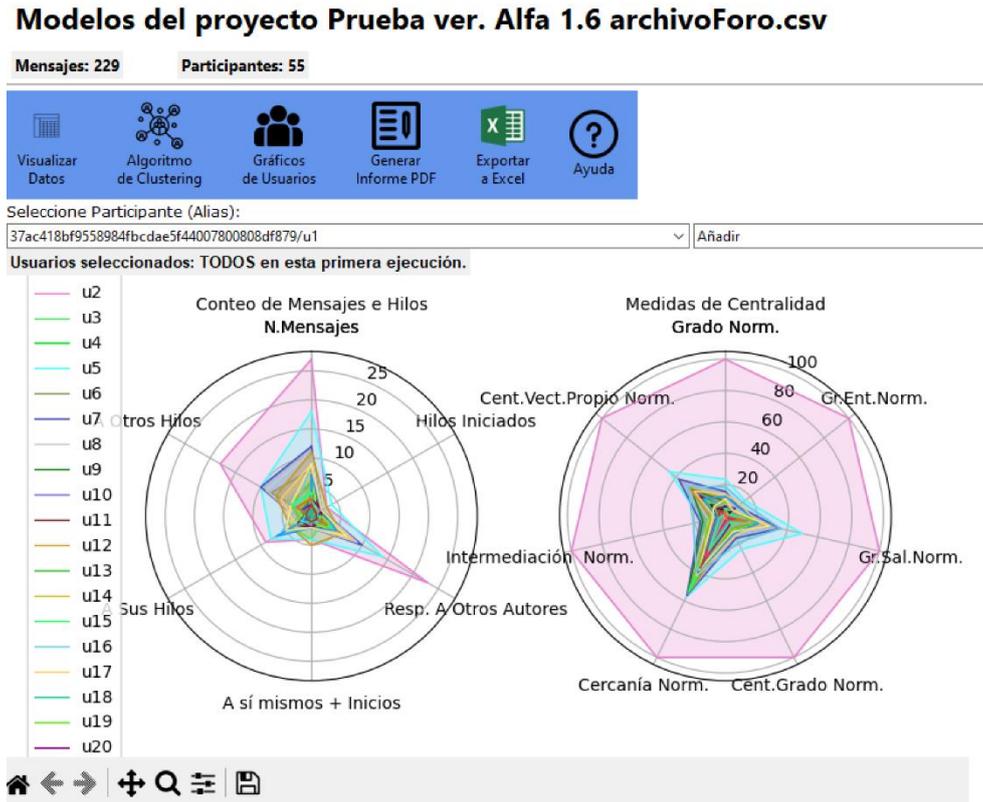


Figura 62. Selección de usuario para dibujar su gráfico.

Lo que se muestra son dos gráficos de tipo radar. En el de la izquierda se representan variables relacionadas con el número de mensajes e hilos iniciados, mientras que en el de la derecha se representan las medidas de centralidad. [Ver el documento de Ayuda del módulo de Red.](#) Los valores de este segundo gráfico se encuentran normalizados mediante el siguiente código:

```
# Nueva escala de 0 a 100.
new_max = 100
new_min = 0
new_range = new_max - new_min
```

```
# Hacer transformación lineal en cada variable para cambiar
el valor
# to [0, 100].
for factor in factors:
    max_val = df[factor].max()
    min_val = df[factor].min()
    val_range = max_val - min_val
    df[factor + ' Norm'] = df[factor].apply(
        lambda x: ((x - min_val) * new_range) / val_range)
+ new_min)
```

Donde *factors* es un array que contiene todas las variables y *factor* representa a cada una de esas variables (medidas de centralidad).

En la primera visualización se representan todos los usuarios. Mediante la rueda del ratón podemos hacer *scroll* para desplazarnos por la leyenda cuando no se puede ver completa por contener muchos usuarios.

Mediante el desplegable de la izquierda podemos seleccionar cualquier usuario para añadirlo o eliminarlo en las gráficas de tipo radar. Una vez se seleccione un usuario, podemos seleccionar en el desplegable de la derecha una de las opciones (*Añadir* o *Quitar*). Al seleccionar cualquiera de estas opciones se dibujarán automáticamente los dos gráficos de tipo radar. La primera vez debe realizarse la adición de cualquier usuario para borrar los gráficos iniciales de todos los usuarios y dibujar los gráficos de ese primer usuario. A partir de ese momento podemos ir añadiendo o quitando usuarios y los gráficos se irán actualizando automáticamente. Ver figura 63.

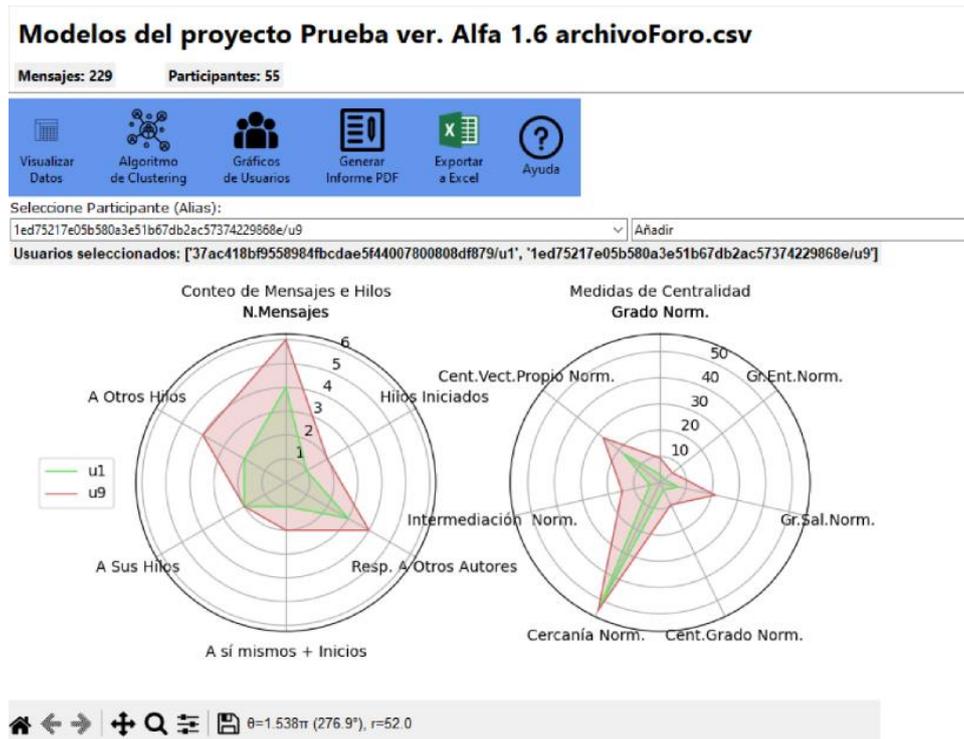


Figura 63. Gráficos de radar de dos usuarios.

En la versión Beta 3.0 de la aplicación se representan:

VARIABLES ASOCIADAS A LAS ESTADÍSTICAS DE NÚMERO DE MENSAJES

- N. Mensajes
- Hilos Iniciados
- Resp. A Otros Autores
- A sí mismos + Inicios
- A Sus Hilos
- A Otros Hilos

VARIABLES DE LAS MEDIDAS DE CENTRALIDAD (NORMALIZADAS)

- Grado
- Grado de Entrada
- Grado de Salida
- Centralidad de Grado
- Cercanía
- Intermediación
- Centralidad de Vector Propio

4.5. Distribución e Instalación de la aplicación

La aplicación se ha desarrollado en un entorno virtual de Python mediante `virtualenv` utilizando el entorno de desarrollo PyCharm. Desde la consola de comandos del entorno virtual se ha creado el archivo `requirements.txt` mediante la instrucción:

```
pip freeze > requirements.txt
```

Esto crea el fichero *requirements.txt*. dentro de la carpeta del proyecto.

Esto facilita la distribución de la aplicación, ya que simplemente es necesario distribuir una copia de la estructura de directorios y archivos (tanto código fuente Python como imágenes, archivos de ayuda, conjuntos de entrenamiento, etc.) junto con el mencionado archivo *requirements.txt*.

Para disponer de la aplicación, la persona que recibe estos datos deberá crear un entorno virtual e instalar todas las librerías indicadas en el archivo *requirements.txt*. Esta tarea se puede realizar de manera muy sencilla utilizando un entorno de desarrollo, como por ejemplo el propio PyCharm (aunque no es imprescindible).

Para generar el entorno virtual e instalar todas las dependencias (librerías) haciendo uso de PyCharm, se debe proceder como se indica a continuación.

Primero debe instalarse PyCharm, ya sea la versión gratuita (Community) o Professional dirigiéndose a la dirección <https://www.jetbrains.com/pycharm/> y descargándose el archivo ejecutable de instalación correspondiente.

Antes de abrir PyCharm colocaremos en una carpeta toda la estructura de archivos suministrada. Si por ejemplo se ha suministrado un archivo *TFM_F_Adame_Toledano_Ver_Beta_3.0.zip* con toda la estructura de ficheros de la aplicación, deberemos descomprimir el contenido de dicho archivo en una carpeta, por ejemplo, llamada *Analítica_Foros*. En la figura 60 se ha optado por llamar a la carpeta con el mismo nombre del archivo (*TFM_F_Adame_Toledano_Ver_Beta_3.0*).

Abrimos PyCharm y vamos a File/New Project accediendo a la pantalla que se muestra en la figura 64. Debemos elegir como ubicación para el nuevo proyecto la carpeta del paso anterior. También elegiremos la opción *New environment using* y en el desplegable elegiremos *Virtualenv* y como ubicación para el entorno virtual también elegiremos la misma ruta de la carpeta anterior.

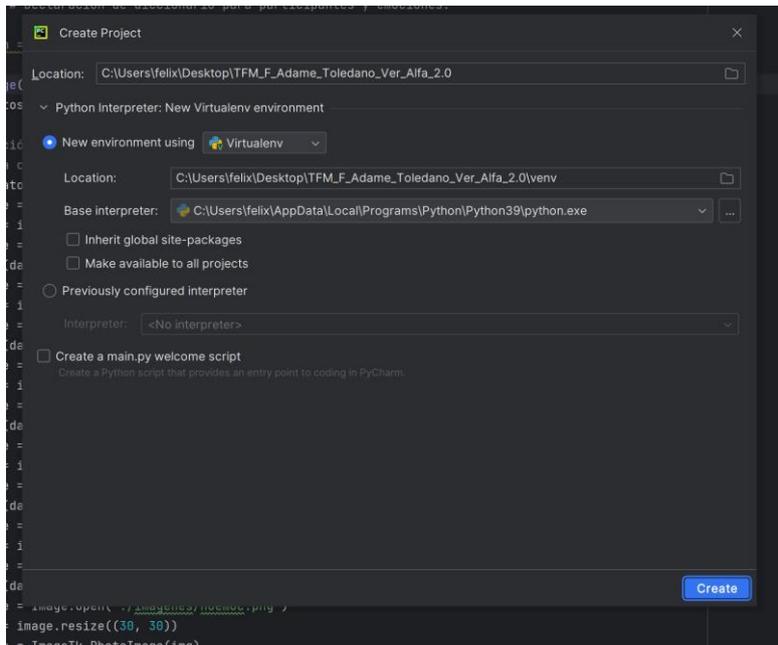


Figura 64. Creación de un nuevo proyecto.

Pulsamos en *Create* y cuando salga el aviso de la figura 65 pulsamos en *Create from Existing Sources*.

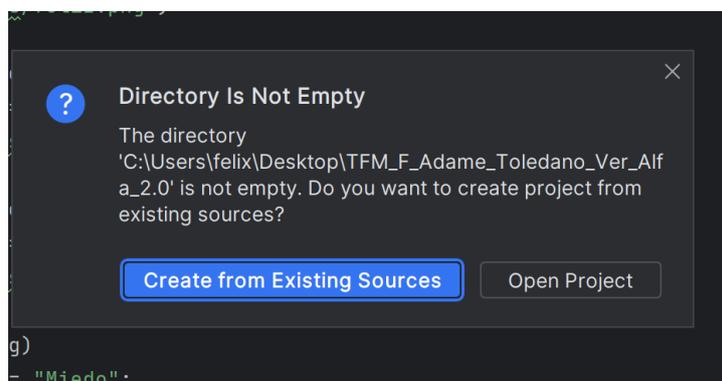


Figura 65. Creación de nuevo proyecto utilizando el código fuente existente.

A continuación es posible, en caso de que tengamos abierto otro proyecto en PyCharm, que se nos pregunte si queremos abrir el nuevo proyecto en la misma o en otra ventana (figura 66).

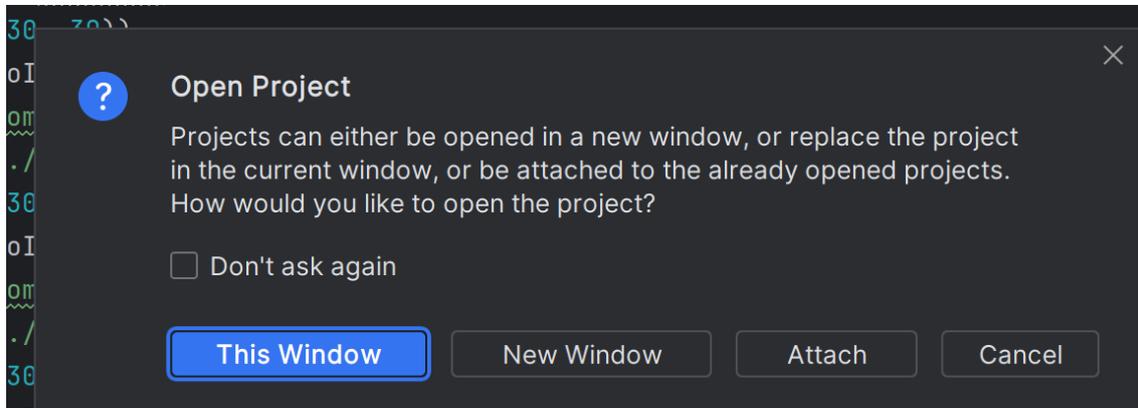


Figura 66. Apertura del proyecto que acabamos de crear.

Una vez abierto el proyecto PyCharm reconocerá la existencia del archivo *requirements.txt* y se nos preguntará si deseamos instalar los paquetes requeridos, indicados en este archivo (figura 67).

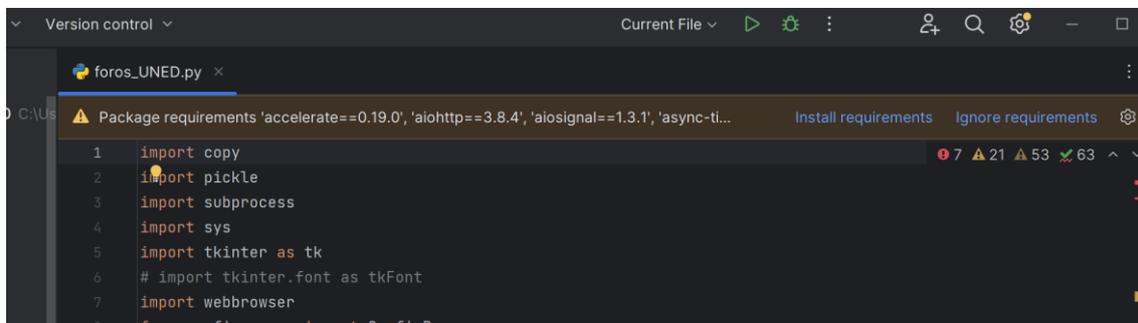


Figura 67. Reconocimiento del archivo *requirements.txt* por el entorno de desarrollo PyCharm.

Deberemos pulsar en *install requirements* tras lo que se nos mostrará el cuadro de la figura 68. En este cuadro es posible desmarcar paquetes que no queramos que se instalen. En principio debe dejarse todo marcado y pulsar en *Install*.

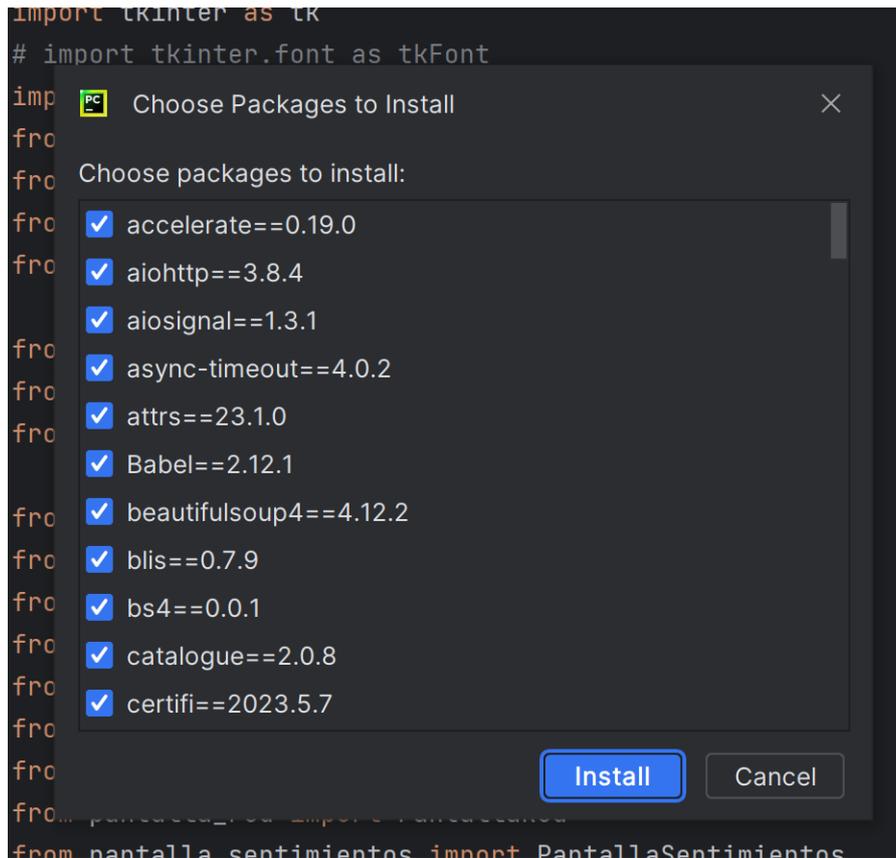


Figura 68. Paquetes que se instalarán por estar recogidos en *requirements.txt*.

En cuanto pulsamos en *Install* podemos ver en la parte inferior de la pantalla cómo comienzan a instalarse los paquetes (figura 69).

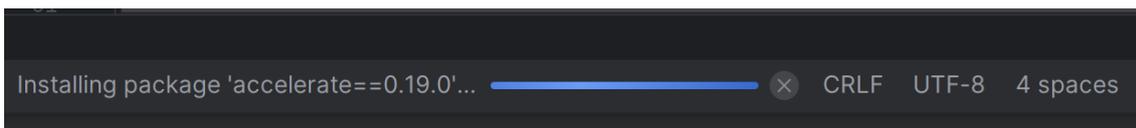


Figura 69. Indicación de los paquetes que se están instalando en el entorno virtual.

El último paso que hay que realizar es la descarga del corpus utilizado por el módulo *Sentiment Analysis Spanish*. Para ello deberemos ir a la consola del propio PyCharm (o a una consola de comandos mediante el comando `cmd` de Windows o bien abriendo un terminal de Linux en caso de hacerlo manualmente); y desde la consola escribir el comando:

```
Python -m spacy download es_core_news_sm
```

Como se indica en la figura 70.

```

Terminal Local x + v
Windows PowerShell
Copyright (C) Microsoft Corporation. Todos los derechos reservados.

Prueba la nueva tecnología PowerShell multiplataforma https://aka.ms/pscore6

(venv) PS C:\Users\felix\Desktop\TFM_F_Adame_Toledano_Ver_Alfa_2.0> python -m spacy download es_core_news_sm
    
```

Figura 70. Descarga del corpus para análisis de sentimientos en español.

Se descargarán los paquetes necesarios indicándose en la consola. Es posible que se nos indique también que hay una nueva versión disponible del instalador *pip* (figura 71).

```

Terminal Local x + v
.6.0,>=3.5.0->es-core-news-sm==3.5.0) (8.1.3)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\felix\desktop\tfm_f_adame_toledano_ver_alfa_2.0\venv\lib\site-packages\MarkupSafe-2.0.1-py3.10-win-amd64.whl from https://pypi.org/simple/markupsafe/ (2.0.1)
Requirement already satisfied: spacy-core-news-sm==3.5.0 in c:\users\felix\desktop\tfm_f_adame_toledano_ver_alfa_2.0\venv\lib\site-packages\spacy-core-news-sm-3.5.0-py3.10-win-amd64.whl from https://pypi.org/simple/spacy-core-news-sm/ (3.5.0)
Installing collected packages: es-core-news-sm
Successfully installed es-core-news-sm-3.5.0

[notice] A new release of pip available: 22.3.1 -> 23.1.2
[notice] To update, run: python.exe -m pip install --upgrade pip
✓ Download and installation successful
You can now load the package via spacy.load('es_core_news_sm')
(venv) PS C:\Users\felix\Desktop\TFM_F_Adame_Toledano_Ver_Alfa_2.0>
    
```

Figura 71. Instalación de es-core-news-sm e indicación de que existe una nueva versión de *pip*.

Además de lo anterior, la primera vez que se ejecute la aplicación, se descargará código adicional, como se puede ver en la figura 72.

```

Salida estándar: Si aparecen errores aquí cierre la aplicación.
Se ha calculado la fecha desde un timestamp
Se ejecuta init de PantallaSentimientos
./recursos/mi_corpus.csv
Downloading (...)lve/main/config.json: 100%|██████████| 925/925 [00:00<00:00, 306kB/s]
C:\Users\felix\Desktop\TFM_F_Adame_Toledano_Ver_Alfa_2.2\venv\lib\site-packages\huggingface_hub\file_download.py:133:
To support symlinks on Windows, you either need to activate Developer Mode or to run Python as an administrator. In or
warnings.warn(message)
Downloading pytorch_model.bin: 5%|█          | 21.0M/435M [00:13<04:04, 1.70MB/s]

```

Figura 72. Descarga de contenido adicional.

4.6. Diseño de la aplicación desarrollada

En esta sección se detalla el diseño de la aplicación atendiendo a varios aspectos. En primer lugar, se expone un diagrama de casos de uso, seguido de un diagrama de clases y una representación basada en UML del diseño. Para la realización de estos diseños se ha utilizado la aplicación Astah²⁷.

4.6.1. Diagrama de Casos de Uso de la aplicación

En la figura 73 se puede ver una representación de los principales casos de uso definidos en la aplicación.

²⁷ <https://astah.net/>

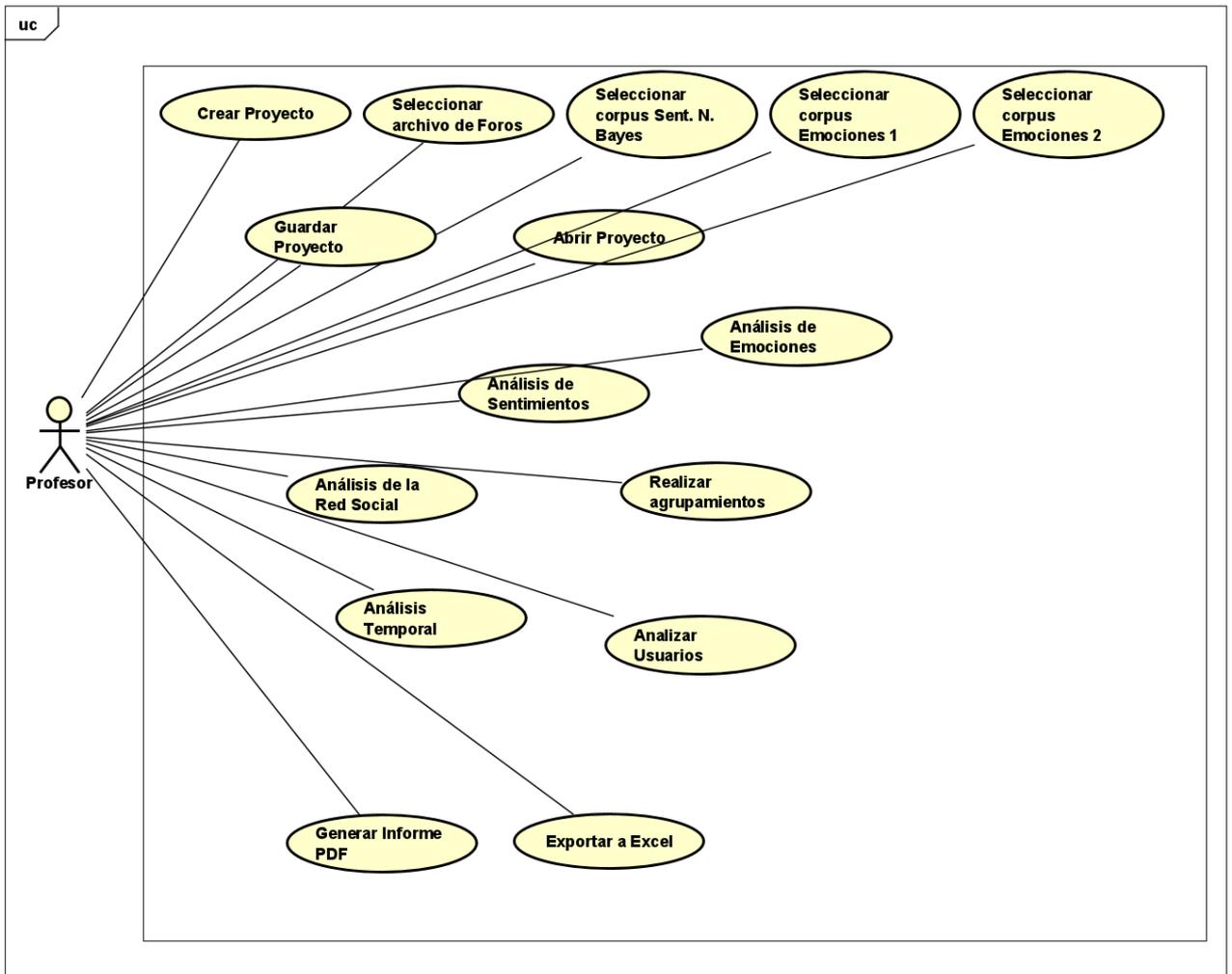


Figura 73. Diagrama de casos de uso.

4.6.2. Diagrama de Clases

La figura 74 representa un esquema muy simplificado de las clases definidas en el proyecto y de las relaciones entre las mismas.

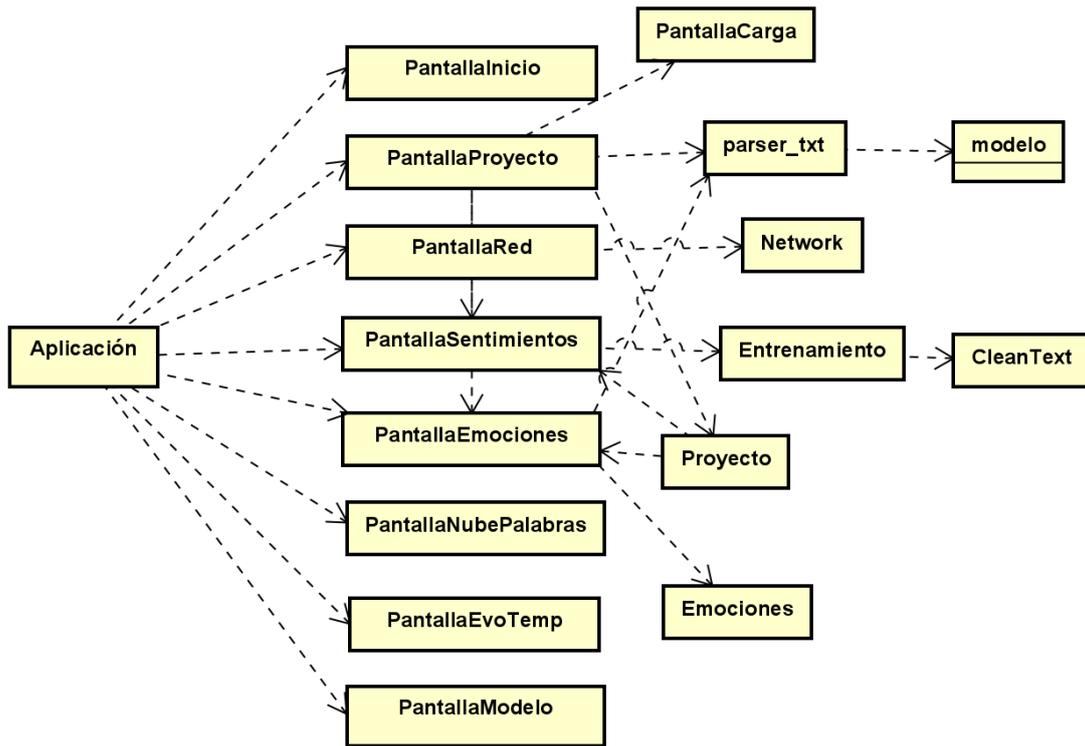


Figura 74. Diagrama de clases.

4.6.3. Modelado UML

En la figura 75 se aprecia una representación siguiendo el estándar UML de las principales clases definidas en la aplicación junto con sus atributos y métodos más significativos o relevantes.

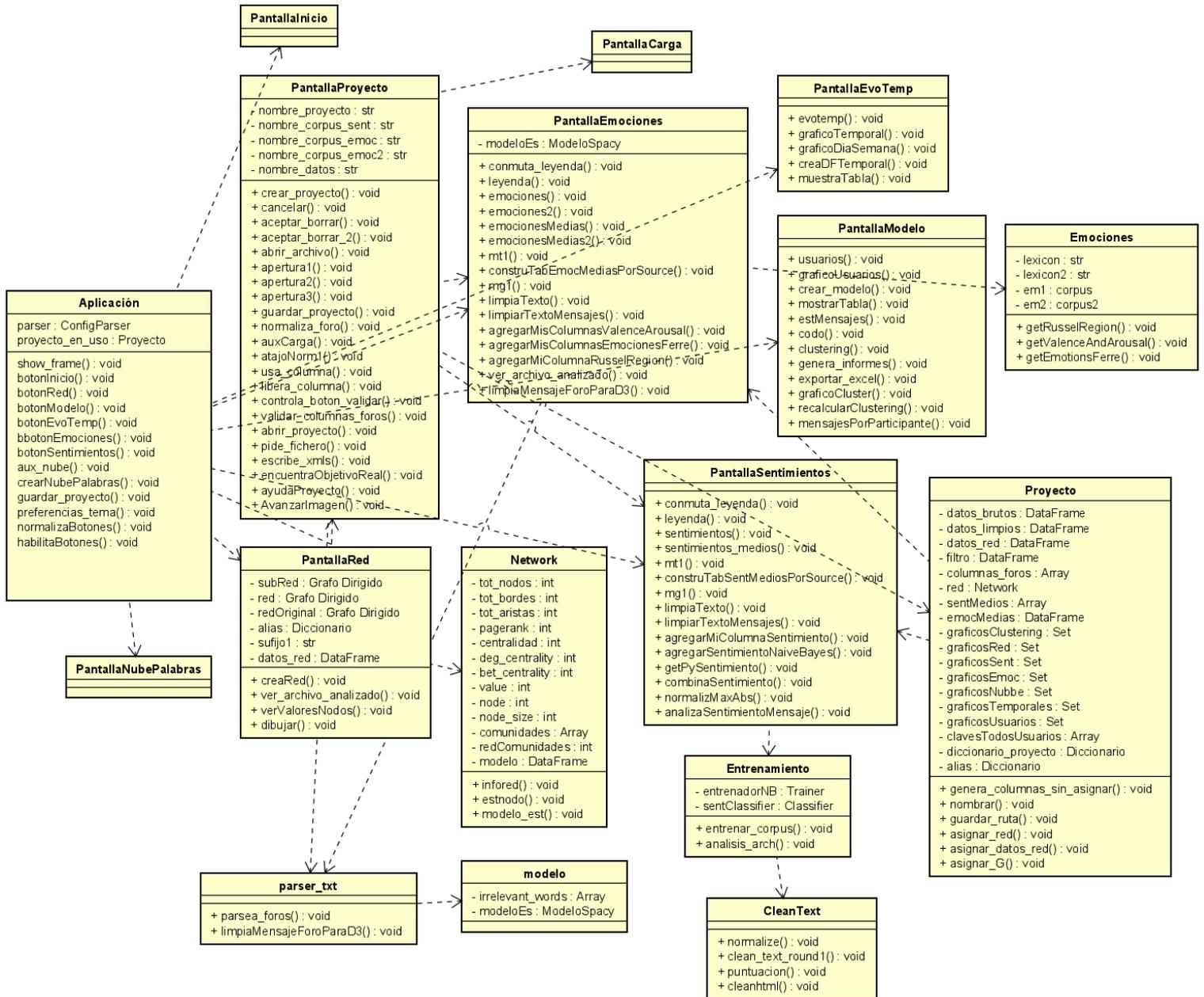


Figura 75. Modelado UML de las clases.

4.6.4. Cronograma del desarrollo y realización del TFM

Este trabajo se ha realizado a lo largo de dos cursos académicos que se pueden identificar como dos fases. En el curso 2021-2022 se abordó bajo la modalidad de tutela para simultanear el inicio del mismo con el estudio de otras asignaturas de grado y máster. En la figura 76 se muestra un cronograma de esta fase (curso). Se ha realizado un desarrollo basado en metodologías ágiles, realizando varias iteraciones principales (versiones), al estilo del Proceso Unificado²⁸ (Larman C., 2010) [36], pero dentro de las mismas se han ido realizando sub-iteraciones (sprints), por lo que se ha adoptado una aproximación a las metodologías SCRUM²⁹ sobre los diferentes módulos que componen la aplicación. Obteniendo de este modo resultados desde una fase muy temprana, pero incrementando y mejorando la funcionalidad a lo largo de las diferentes iteraciones y *sprints*.

²⁸ https://es.wikipedia.org/wiki/Proceso_unificado

²⁹ [https://es.wikipedia.org/wiki/Scrum_\(desarrollo_de_software\)](https://es.wikipedia.org/wiki/Scrum_(desarrollo_de_software))

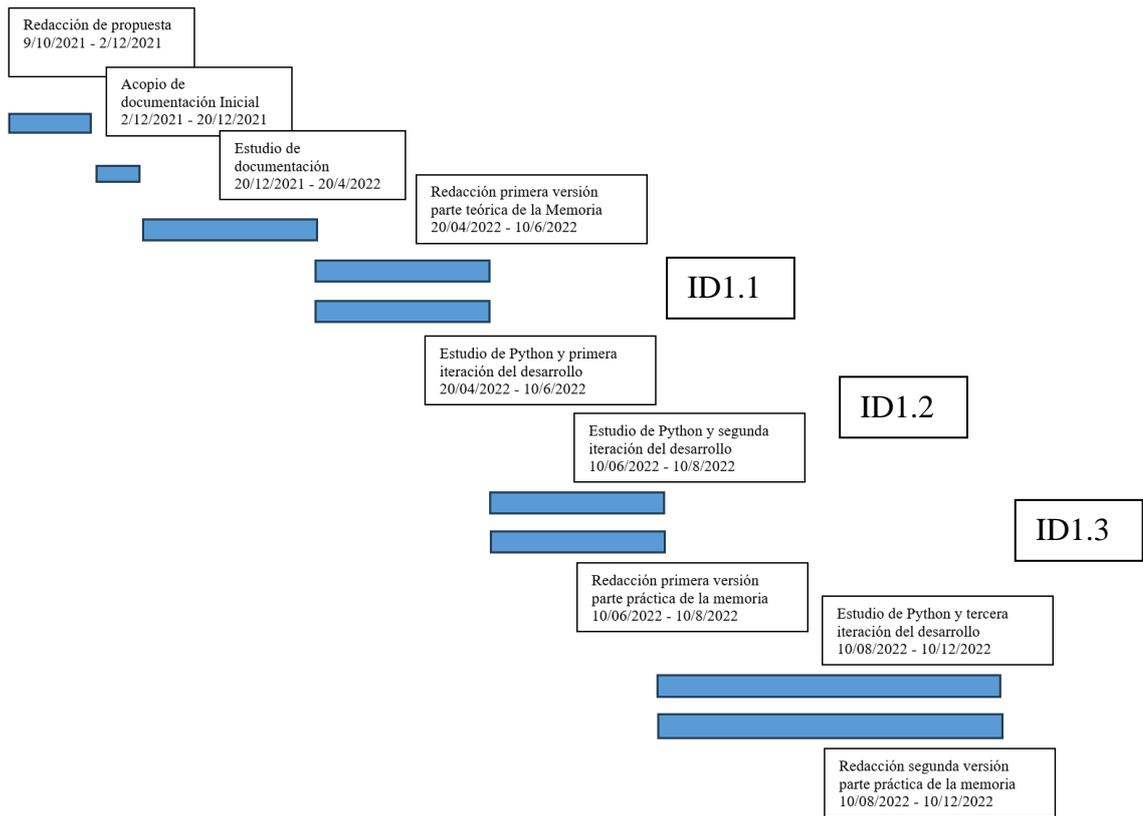


Figura 76. Primera fase del desarrollo del TFM.

Posteriormente, durante el curso 2022-2023 se realizaron nuevas iteraciones de desarrollo y sendas versiones de la memoria hasta llegar a la versión final en agosto de 2023 (figura 77).

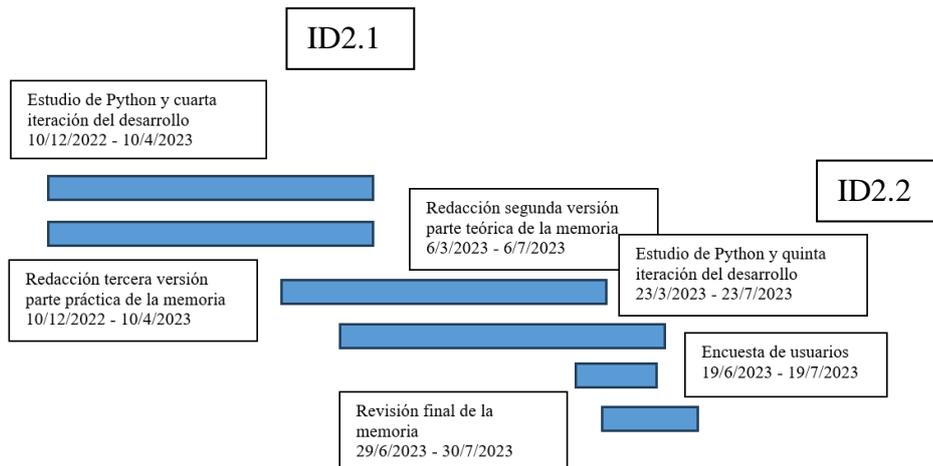


Figura 77. Segunda fase del desarrollo del TFM.

En lo que se refiere al desarrollo de la aplicación y cómo ha ido creciendo la funcionalidad a lo largo de las dos fases y las cinco iteraciones, a continuación se indica un resumen que hace referencia a dichas iteraciones identificadas en las figuras 76 y 77 como IDX.X. Cada uno de estos identificadores representa una estructura del estilo de desarrollo en cascada clásico, incluyendo: **captura de requisitos y análisis, diseño, implementación y pruebas.**

- **ID1.1:** **Análisis** de requisitos y usabilidad; **diseño** de primera versión incompleta e inmadura para probar las funcionalidades básicas; Ingeniería inversa Python para reaprovechamiento de código; **desarrollo** de la primera versión; **pruebas.** Versiones Alpha 1.X
- **ID1.2:** **Análisis** de funcionalidades útiles para Learning Analytics; **rediseño** por completo y cambio de paradigma en lo que se refiere a la entrada de datos, brindando la flexibilidad del nombre de las columnas (usabilidad); se abandona la dependencia de Google translate y se adopta *Sentiment Analysis Spanish*; **desarrollo** y corrección de errores; **pruebas.** Versiones Alpha 2.0 y 2.1.
- **ID1.3:** Revisión del **análisis**; revisión del **diseño** con mejora de la visualización y se incluye funcionalidad de informes automáticos y dos métodos de análisis de emociones; **desarrollo**, definición del sistema de distribución (entornos virtuales) y corrección de errores; **pruebas.** Versión Alpha 2.2
- **ID2.1:** Revisión del **análisis**; revisión del **diseño** con mejora de la visualización incluyendo el concepto de “alias” y rehacer función de *clustering*, mejora en la flexibilidad del formato de mensajes de los foros, ampliación de funcionalidad en módulo de red respecto al análisis de sentimientos, se implementa la posibilidad de utilizar otros corpus para análisis de sentimientos y emociones, función de exportación de datos a Excel, creación de archivos de ayuda, mejoras en informe automático PDF e inclusión del análisis de sentimientos *combinado*³⁰; **desarrollo** y corrección de errores (se cambia la librería para implementar Naive Bayes, instalando e importando *naiveBayesClassifier*³¹ en lugar de utilizar el clasificador *NaiveBayesClassifier* de *textblob.classifiers*³²); **pruebas.** Versión Alpha 3.0

³⁰ Análisis de sentimientos propio de este trabajo que consiste en la combinación de varios análisis y constituye un objeto de experimentación dentro del presente TFM.

³¹ <https://pypi.org/project/naiveBayesClassifier/>

³² <https://textblob.readthedocs.io/en/dev/classifiers.html>

- **ID2.2: Análisis** de oportunidades de mejora; mejoras de **diseño** menores en la visualización y la ayuda, también se incluye la opción de exportar a Excel los datos de la red, inclusión de Gráficos de usuarios; **desarrollo** y corrección de errores; **pruebas**. Versión Beta 3.0

Capítulo 5

Pruebas

En este capítulo se va a mostrar el proceso de prueba de la aplicación con un archivo de foros real, comparando los resultados obtenidos con los que arrojan otras soluciones.

Comparativa de análisis de Red

Cabe indicar que se ha añadido un botón en el módulo de Análisis de Red para exportar a Excel la red (Emisores y Destinatarios). Esto se ha hecho porque el formato de los datos de los foros no incluye directamente de modo general, a los destinatarios de los mensajes, sino el identificador de mensaje al que se responde. La aplicación, al preprocesar los datos, obtiene el destinatario de cada mensaje a partir del identificador de mensaje y lo incluye en una nueva columna “Destinatario”

En primer lugar, tomamos un archivo de foros (MDEME-Foro-22-11-29.csv). Se ha procesado dicho archivo con la aplicación y posteriormente se ha exportado a Excel la información.

Comparativa con *Gephi*

Gephi es una herramienta de libre uso para el análisis de redes sociales.

Para poder utilizar la información del archivo Excel obtenido en el paso explicado más arriba, se tienen que cambiar los nombres de las columnas: *userid* a *Sourcee* y *Destinatario* a *Target*. Una vez hecho esto se importa el archivo desde *Gephi* como archivo de aristas. En la figura 78 se ve la red tal como se previsualiza en *Gephi*. La red

se puede visualizar de múltiples formas realizando configuraciones que no son triviales para un usuario novel. También se aprecia el número de nodos (16) y aristas (51).

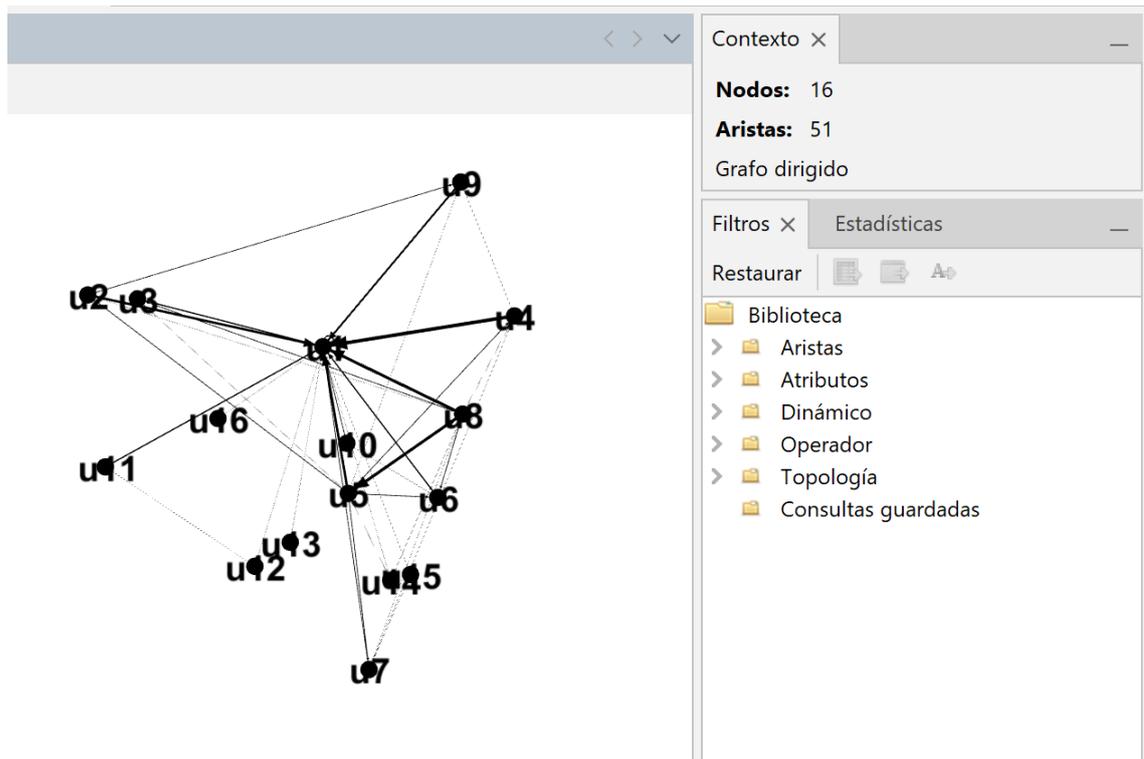


Figura 78. Previsualización de la red en Gephi.

En la figura 79 se puede ver la red representada en nuestra aplicación sin necesidad de realizar ninguna configuración.

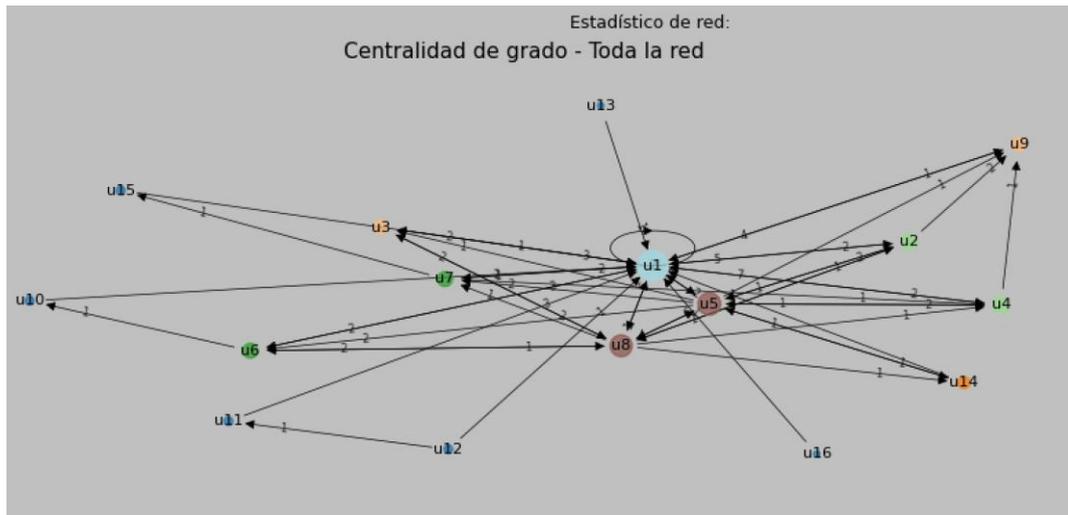


Figura 79. Representación de la red en nuestra aplicación.

En la figura 80 se aprecian los datos de los mensajes que constituyen las aristas, tal cual se ven en *Gephi*.

Origen	Destino	Tipo	Id	Label	Interval	Weight	parent	subject	alias
2946	25055	Dirigida	2096970	1.0		1.0	37266	Re: Acerc...	u4
2946	71704	Dirigida	2096971	1.0		1.0	42345	Re: Acerc...	u4
2946	79	Dirigida	2096891	7.0		7.0	32545	Re: Acerc...	u4
3709	79	Dirigida	2096915	1.0		1.0	14014	Re: Bienv...	u15
3778	79	Dirigida	2096956	1.0		1.0	14014	Re: Bienv...	u16
5055	79	Dirigida	2096896	4.0		4.0	32545	Re: Acerc...	u9
5964	23709	Dirigida	2096928	1.0		1.0	29321	Re: Bienv...	u7
5964	22946	Dirigida	2096943	1.0		1.0	50199	Re: Tarea...	u7
5964	79	Dirigida	2096957	2.0		2.0	58889	Re: Resp...	u7
9712	51194	Dirigida	2096893	1.0		1.0	16461	Re: Bienv...	u6
9712	69829	Dirigida	2096944	1.0		1.0	45813	Re: Propu...	u6
9712	79	Dirigida	2096906	3.0		3.0	58845	Re: Mode...	u6
1194	79	Dirigida	2096892	2.0		2.0	21040	Re: Charl...	u10
5100	79	Dirigida	2096894	3.0		3.0	33892	Re: Curso...	u11
8223	55100	Dirigida	2096895	1.0		1.0	17890	Re: Bienv...	u12
8223	79	Dirigida	2096907	1.0		1.0	21040	Re: Charl...	u12
0223	79	Dirigida	2096897	1.0		1.0	14014	Re: Bienv...	u13
4968	69829	Dirigida	2096936	2.0		2.0	46939	Re: Mode...	u3
4968	79	Dirigida	2096890	3.0		3.0	21040	Re: Charl...	u3
6182	69829	Dirigida	2096940	1.0		1.0	45813	Re: Propu...	u2
6182	71704	Dirigida	2096967	1.0		1.0	64694	Re: Mode...	u2
6182	25055	Dirigida	2096932	2.0		2.0	50050	Re: Mode...	u2
6182	79	Dirigida	2096904	5.0		5.0	58859	Re: Mode...	u2

Figura 80. Datos de los mensajes (aristas) en Gephi.

La tabla análoga a la anterior en nuestra aplicación se aprecia en la figura 81.

Red del proyecto Prueba ; MDEME-Foro-22-11-29.csv

Nodos: 16 Aristas: 51

Estadísticas de la red

Centralidad de Grado ?
Cercanía ?
Intermediación ?
Centralidad de Vector Propio ?
Page Rank ?
Iniciativa ?
Modularidad ?
S

Etiquetas de gráfico: Alias Sentimiento: N.Bayes (Corpus Usuario)

Filtrar por: todo Seleccionar: TODO Visualizar archivo Fecha in

	id	discussio	parent	userid	created	modified	mailed	subject	message	Alias	mess
81	59060	17861	58845	39712	16678405	16678405	1	Re: Model	hola anton	u6	2
82	59064	16841	58851	22946	16678406	16678406	1	Re: Model	teniendo a	u4	2
83	59633	19684	58889	71704	16678908	16678908	1	Re: Respue1*	aplicac	u5	2
84	60114	5193	14014	23778	16679144	16679144	1	Re: Bienve	cual es tu	u16	2
85	64265	10981	32545	25964	16682553	16682553	1	Re: Acerc	buenos di	u7	2
86	64643	14346	43000	71704	16683342	16683342	1	Re: Tema	hola	u5	2

Figura 81. Datos de los mensajes (aristas) en nuestra aplicación.

Finalmente vamos a comparar la visualización de la información relativa a los nodos y sus estadísticas. En la figura 82 la tabla de *Gephi*.

Tabla de datos x

Nodos Aristas Configuración Añadir nodo Añadir arista Buscar/Reemplazar Importar hoja de cálculo Exportar tabla Más acciones Filtro: Id

Id	Label	Interval	Grado de entrada	Grado de salida	Grado	Modularity Class	PageRank	Eigenvector Centrality
64968	u3		3	2	5	0	0.051288	0.407609
79	u1		16	9	25	0	0.253663	1.0
22946	u4		4	3	7	0	0.065982	0.487403
51194	u10		1	1	2	0	0.02407	0.079794
39712	u6		3	3	6	0	0.051288	0.407609
55100	u11		1	1	2	0	0.013359	0.002001
58223	u12		0	2	2	0	0.009375	0.0
25055	u9		4	1	5	0	0.071631	0.47686
60223	u13		0	1	1	0	0.009375	0.0
69829	u8		5	8	13	0	0.089492	0.541452
66182	u2		3	4	7	0	0.051288	0.407609
71704	u5		5	9	14	0	0.084813	0.521571
23709	u15		1	1	2	0	0.02407	0.079794
25964	u7		3	3	6	0	0.051288	0.407609
66967	u14		2	2	4	0	0.02699	0.207554
23778	u16		0	1	1	0	0.009375	0.0

Figura 82. Estadísticas de la red en Gephi.

En la figura 83 la tabla análoga en nuestra aplicación.

NODO	ALIAS	GRADO	GRADO_	GRADO_S	CENTRAL	CERCANI	INTERME	VECTOR	PAGE_R	INICIATIV	MODULA
79	u1	25	16	9	1.67	1.00	0.52	0.6	0.3	5	0
66182	u2	7	3	4	0.47	0.56	0.11	0.25	0.057	2	1
64968	u3	5	3	2	0.33	0.56	0.11	0.25	0.057	1	1
22946	u4	7	4	3	0.47	0.58	0.11	0.29	0.074	1	1
71704	u5	14	5	9	0.93	0.6	0.18	0.32	0.094	3	1
39712	u6	6	3	3	0.4	0.56	0.17	0.25	0.057	2	0
25964	u7	6	3	3	0.4	0.56	0.17	0.25	0.057	1	2
69829	u8	13	5	8	0.87	0.6	0.17	0.33	0.1	2	1
25055	u9	5	4	1	0.33	0.58	0.11	0.29	0.08	1	1
51194	u10	2	1	1	0.13	0.38	0.11	0.049	0.026	0	0
55100	u11	2	1	1	0.13	0.067	0.054	2.5e-07	0.013	0	0
58223	u12	2	0	2	0.13	0	0.054	2.5e-08	0.0094	0	0
60223	u13	1	0	1	0.067	0	0.05	2.5e-08	0.0094	0	0
66967	u14	4	2	2	0.27	0.42	0.11	0.13	0.029	1	1
23709	u15	2	1	1	0.13	0.38	0.11	0.049	0.026	0	2
23778	u16	1	0	1	0.067	0	0.05	2.5e-08	0.0094	0	0

Figura 83. Estadísticas de la red en nuestra aplicación.

Comparativa de Sentimientos

Se han realizado comparaciones del análisis de sentimientos de algunos mensajes entre la aplicación y una herramienta online (Linguakit)³³. Linguakit devuelve un literal (POSITIVO o NEGATIVO) y un valor de probabilidad para dicho valor.

Mensaje 1

Hola, yo no tengo problemas con ninguna de las fechas, así que cuando queráis. Un saludo.

Análisis linguakit.com

POSITIVE 0.999999321846056

Análisis Aplicación

Sentim. Analys. Sp.: 1.7e-06

Sentim. Analys. Sp. Norm.: 1.7e-06

Sent. N. Bayes: neg

PySentimiento: NEU

Sent. Combinado: 0.002 = REGULAR.

Mensaje 2

hola maria

me parece una propuesta a valorar pues permitiría realizar la evaluación que ha comentado antonio sobre la utilidad de las recomendaciones y en este caso de las apps mostradas

gracias por la sugerencia inma

³³ <https://linguakit.com/es/analisis-completo>

Análisis linguakit.com

POSITIVE 0.999999998132245

Análisis Aplicación

Sentim. Analys. Sp.: 3.674151831461974e-06

Sentim. Analys. Sp. Norm.: 3.6741518314895425e-06

Sent. N. Bayes: pos

PySentimiento: POS

Sent. Combinado: 0.0058 = BUENO.

Mensaje 3

yo estoy igual no puedo responder porque pone que la fecha limite es el 20 de octubre.

Análisis linguakit.com

NEGATIVE 0.999992203957731

Análisis Aplicación

Sentim. Analys. Sp.: 8.124289912789463e-05

Sentim. Analys. Sp. Norm.: 8.124289912850423e-05

Sent. N. Bayes: neg

PySentimiento: NEG

Sent. Combinado: 0.00041173408813952896 = MALO.

Como se puede apreciar, se obtiene una mayor resolución de los sentimientos con nuestra aplicación, fijándonos en el valor del Sentimiento medio combinado.

El sentimiento combinado en este proyecto se mueve entre 0.00037 y 0.0089. Además, sabemos que teóricamente, el máximo posible es 1.125.

La clasificación realizada en la aplicación para traducir el sentimiento combinado medio de los participantes a un literal se realiza mediante las particiones definidas en el siguiente fragmento de código:

```
# Sentimiento medio combinado literal
for val in medias_sent_comb:
    if val <= 0.0001 or val == 0.0:
        medias_comb_lit.append('Muy malo')
    elif val <= 0.002:
        medias_comb_lit.append('Malo')
    elif val <= 0.005:
        medias_comb_lit.append('Regular')
    elif val <= 0.006:
        medias_comb_lit.append('Bueno')
    elif val <= 100:
        medias_comb_lit.append('Muy bueno')
```

Esto se traduce en una conversión no lineal de acuerdo al siguiente gráfico de la figura, el cual es el mismo que se puede ver en la figura 84:

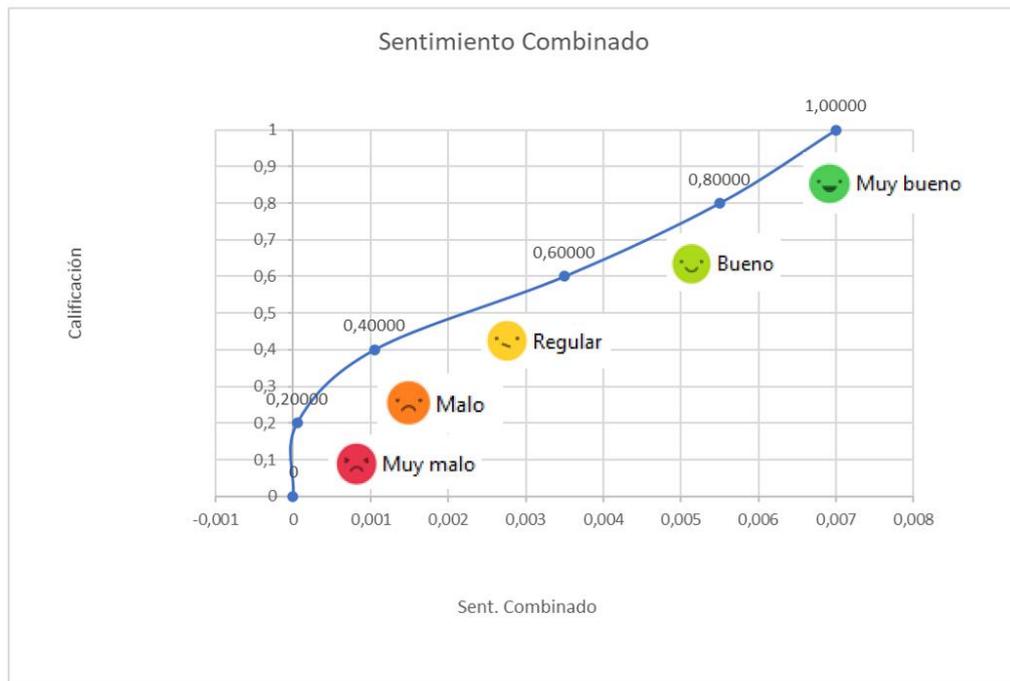


Figura 84. Correspondencia entre sentimiento medio combinado y sentimiento literal.

La comparación nos muestra que la calificación de sentimiento para *linguakit* es en ambos casos de 0.99, mientras que nuestra aplicación da a la primera un 0.002 y a la segunda un 0.0058. Según esto, el segundo mensaje analizado sería más positivo que el primero, el cual se puede considerar más bien entre neutro y positivo.

Encuesta a usuarios de prueba

En el apéndice B se recogen los resultados íntegros de una encuesta realizada a un grupo de cinco usuarios antes de la versión final. En base a estas respuestas se han realizado las últimas modificaciones del código y de los ficheros de ayuda, intentando mejorar algunos aspectos de la usabilidad y de la utilidad en la medida de lo posible.

A continuación, se reproducen los resultados más significativos de la encuesta en forma de tabla.

En la tabla 3 se muestra el resultado de la encuesta sobre la percepción inicial que tienen los usuarios de la aplicación.

Tabla 3. Resultados de la encuesta sobre experiencia de usuario inicial.

Experiencia de usuario inicial			
Es fácil saber, desde la primera pantalla, qué puede hacer la aplicación.	Totalmente de acuerdo	Parcialmente de acuerdo	En desacuerdo
Usuario 1		X	
Usuario 2		X	
Usuario 3		X	
Usuario 4	X		
Usuario 5	X		
Es fácil saber, desde la primera pantalla, cómo se utiliza la aplicación			
Usuario 1		X	
Usuario 2		X	
Usuario 3	X		
Usuario 4		X	
Usuario 5		X	
La ayuda es clara y suficiente para comenzar a utilizar la aplicación			
Usuario 1		X	
Usuario 2		X	
Usuario 3	X		
Usuario 4	X		
Usuario 5			X

En base a dichos resultados se ha mejorado sensiblemente la ayuda y la información contextual.

En la tabla 4 se muestra el resultado de la encuesta sobre la percepción de para qué sirve la aplicación.

Tabla 4. Resultado de la encuesta. Para qué sirve la aplicación.

Una vez que ha utilizado la aplicación, ¿para qué sirve la aplicación? Escriba un breve comentario.	
Usuario 1	Analizar las interacciones entre los usuarios de foros. Hace análisis automático de las emociones y sentimientos presentes en la interacciones escritas.
Usuario 2	Analizar las interacciones entre usuarios y el contenido emocional de dichas interacciones.
Usuario 3	Para tomar el pulso a los estudiantes y conocer el tono afectivo de la clase en su conjunto así como de cada estudiante en particular.
Usuario 4	Para realizar un análisis de emociones y sentimientos extraído de la interacción en el foro de los estudiantes.
Usuario 5	Para conocer información de la interacción de los alumnos en un foro.

En la tabla 5 se muestran los resultados sobre lo que se percibe que la aplicación ofrece.

Tabla 5. Resultado de la encuesta. Qué ofrece la aplicación.

Una vez que ha utilizado la aplicación, ¿qué ofrece la aplicación? Escriba un breve comentario.	
Usuario 1	Computa indicadores de interacción y lo presenta en formato tabular o gráfico en función de lo que se seleccione.
Usuario 2	El cómputo de las interacciones entre usuarios y su representación gráfica.
Usuario 3	Para que la aplicación sea útil el profesor tiene que estar formado en enfoques didácticos afectivos. De nada sirve la información que proporciona la aplicación si el profesor no sabe utilizarla en modo productivo, o incluso la utiliza en modo perjudicial para la buena marcha del curso.
Usuario 4	Análisis de Sentimientos, Gráficos de Sentimientos, Análisis de Emociones, Gráficos de Emociones, Estadísticas de los mensajes y red, Gráficos de la red, Gráficos de clustering, Gráficos temporales, Gráficos de Usuarios, Nubes de palabras del proyecto.
Usuario 5	Ver el grado de colaboración y el estado emocional reflejado en los mensajes de los foros.

A la luz de los resultados de las tablas 4 y 5, se ha mejorado la ayuda de la aplicación, intentando transmitir mejor lo que esta ofrece y para qué sirve.

En la tabla 6 se muestran los resultados de la encuesta sobre la petición de propuestas de mejora para diferentes aspectos.

Tabla 6. Resultado de la encuesta. Propuestas de mejora.

¿Qué añadiría para que la aplicación fuera más usable?	
Usuario 1	Un entorno más gráfico, basado en estándares de interacción: botones, filtros, cajas, sliders...
Usuario 2	Un entorno gráfico más amigable.
Usuario 3	Es bastante usable.
Usuario 4	Poder hacer análisis centrados en un usuario.
Usuario 5	Aclaraciones o leyendas.
¿Qué añadiría para que la aplicación fuera más intuitiva?	
Usuario 1	Ídem = Un entorno más gráfico, basado en estándares de interacción: botones, filtros, cajas, sliders...
Usuario 2	Un mejor etiquetado.
Usuario 3	Depende de los usuarios destinatarios. Para profesores de escasa formación científica está claro que tanta profusión de parámetros, datos y gráficos debe simplificarse, y quizá sustituirse o ampliarse con descripciones textuales.
Usuario 4	Aclarar algunos términos, como "atajos".
Usuario 5	Botones con lenguaje del profesor, no del desarrollador.
¿Qué añadiría para que la aplicación fuera más práctica?	
Usuario 1	Funcionalidad de filtros configurables.
Usuario 2	Poder filtrar mejor los datos.
Usuario 3	Conectar los resultados mostrados de forma explícita con el transcurrir del curso. Destacar los datos más llamativos en resúmenes.
Usuario 4	Facilitar su uso en la web sin necesidad de instalaciones.
Usuario 5	Poder mezclar info de diferentes secciones, p.e. ver la centralidad de los alumnos con determinado estado anímico.

Estos resultados constituyen el punto de partida para posibles mejoras de la aplicación.

Finalmente, en la tabla 7, se muestran los resultados de la encuesta en lo que se refiere a la usabilidad y utilidad general de la aplicación.

Tabla 7. Resultados de la encuesta. Usabilidad y utilidad general.

Valore la usabilidad de la aplicación entre 1 y 10	
Usuario 1	6
Usuario 2	5
Usuario 3	5
Usuario 4	8
Usuario 5	8
Valore la utilidad de la aplicación entre 1 y 10	
Usuario 1	9
Usuario 2	7
Usuario 3	8
Usuario 4	8
Usuario 5	8

Los resultados ponen de manifiesto que la aplicación se percibe útil, pero tiene un amplio margen de mejora, especialmente en lo que se refiere a usabilidad. También se comprueba que puede ser menos útil para usuarios que no dispongan de unas nociones mínimas acerca de la teoría de redes sociales y de la analítica de sentimientos y emociones. Incluso para profesores que cuenten con conocimientos de programación es relativamente sencillo realizar configuraciones o modificaciones del código, lo que se contrapone con la dificultad de realizar configuraciones básicas para un usuario totalmente ajeno a la programación. En este sentido, una mejora que no se ha adoptado y sería interesante es la de facilitar la configuración a través de la propia interfaz de la aplicación.

Capítulo 6

Conclusiones

El desarrollo de la aplicación denominada tentativamente *Análisis de Foros UNED* en el seno del presente Trabajo Fin de Máster, ha resultado verdaderamente enriquecedor al poder adquirir un nivel de conocimientos sobre el lenguaje de programación Python muy superior al de partida, que consistía en una base elemental del mismo. Esto constituía uno de los objetivos iniciales ([OBJ1](#)).

De igual modo, la lectura de los artículos de referencia y el estudio en profundidad de los más relevantes, ha supuesto una profundización en las técnicas de Aprendizaje Automático y de técnicas de análisis de sentimientos y emociones, así como de análisis de redes sociales. Incluso se han adquirido conocimientos que finalmente no se han aplicado, como puede ser el caso del procesamiento del lenguaje natural.

El hecho de tener que analizar un código desarrollado por otras personas, como es el caso del TFM de partida, el trabajo de [Álvaro Rodrigo et. al](#) o de algunas librerías, también ha ayudado a adquirir más conocimientos sobre Python ([OBJ1](#)), pero además ha constituido la consecución de dos de los objetivos marcados al inicio, que eran el de selección y utilización de librerías de terceros ([OBJ2](#)), desarrollado en las iteraciones [ID1.1](#) e [ID2.1](#) y el de ser capaz de reutilizar y adaptar código desarrollado por otras personas ([OBJ3](#)), que se ha desarrollado durante la iteración [ID1.1](#)

La aplicación desarrollada es relativamente fácil de distribuir mediante los mecanismos que proporcionan los entornos virtuales de Python, lo que constituía el objetivo [OBJ4](#). Se consiguió en la iteración [ID1.3](#), cuando se definió la adopción de entornos virtuales.

La representación de la red social se ha mejorado al incorporar la funcionalidad de herramientas de gráficos como zoom, desplazamiento y ajuste. También se ha incluido la posibilidad de elegir las etiquetas y se cambia el tamaño de los nodos en función del grado de los mismos. Se incluyen dos representaciones diferentes de la red atendiendo al sentimiento de los mensajes (Naive Bayes y PySentimiento). Esto cumple el objetivo [OBJ5](#) y se desarrolló en [ID1.2](#) y en [ID2.1](#).

Se ha cumplido el objetivo [OBJ6](#) al suministrar la aplicación una tabla con toda la información de los usuarios a través de la opción correspondiente dentro del módulo de Modelos e Informes ([ID1.2](#), [ID1.3](#), [ID2.1](#) y [ID2.2](#)).

Se ha creado una medida propia de los sentimientos a través de un análisis combinado basado en varios análisis diferentes (Sentimiento Combinado) y se incluyen dos tipos de análisis de sentimientos que pueden compararse, lo que supone la consecución del objetivo [OBJ7](#). Además, esta información se representa de manera gráfica, con lo que se da por alcanzado el objetivo [OBJ8](#). ([ID1.2](#), [ID1.3](#), [ID2.1](#) y [ID2.2](#)).

La interfaz de la aplicación desarrollada es más usable, puesto que disponemos de una estructura de módulos con un menú lateral. No se abren nuevas ventanas durante el uso de la aplicación salvo las de los archivos de ayuda o cuando se exportan datos o generan informes. Es decir, toda la acción se desarrolla dentro de una misma ventana gráfica. En todos los módulos se dispone de una ayuda contextual además de una ayuda básica de uso de la aplicación tanto en PDF como integrada en el caso del módulo de Proyectos. Esto supone la consecución del objetivo [OBJ9](#).

Se ha dotado de una gran flexibilidad al formato de los datos de entrada, tanto en lo que se refiere al formato de archivo como al nombre de las columnas de los datos ([OBJ10](#)).

Se ha seguido una metodología ágil siguiendo las directrices del Proceso Unificado (Larman C., 2010) [36]([OBJ11](#)).

Por último, la aplicación desarrollada se ha utilizado como herramienta para el análisis de sentimientos en la realización de otro TFM, realizado por Juan García Ruiz,

titulado *Estudio sobre la búsqueda de mentor utilizando aprendizaje automático*, bajo la dirección del Dr. Antonio Rodríguez Anaya. Esto demuestra la utilidad práctica de la aplicación desarrollada en este trabajo, al ser una manera sencilla para realizar el análisis de sentimientos de los mensajes.

Por lo tanto, se consideran alcanzados los objetivos marcados al inicio.

En la tabla 8 se muestra la consecución de los objetivos de una manera resumida.

Tabla 8. Consecución de objetivos del trabajo.

Código	Objetivo	Carácter	Consecución
OBJ1	Adquirir conocimientos sobre el lenguaje de programación Python	Aprendizaje	SI
OBJ2	Utilización de librerías	Técnico	SI
OBJ3	Comprensión y reutilización y aprovechamiento de código de terceros	Ingenieril	SI
OBJ4	Desarrollo de aplicación fácil de distribuir	Desarrollo	SI
OBJ5	Mejorar la representación de la red social	Desarrollo	SI
OBJ6	Estadísticas, medidas de centralidad y datos de los participantes mejorados y ampliados	Desarrollo	SI

Código	Objetivo	Carácter	Consecución
OBJ7	Mejora y ampliación de análisis de sentimientos y emociones	Desarrollo	SI
OBJ8	Representación gráfica de la información de forma mejorada y ampliada a nivel individual y agregado	Desarrollo	SI
OBJ9	Conseguir una aplicación con una interfaz más amigable, integrada y funcional.	Desarrollo	SI
OBJ10	Flexibilizar el formato de los datos de entrada	Desarrollo	SI

La aplicación desarrollada ha podido utilizarse por terceras personas para propósitos de investigación/aprendizaje...

Capítulo 7

Posibles líneas de mejora y evolución

A la luz de los resultados de la encuesta realizada, la aplicación puede ampliarse y mejorarse en diferentes aspectos, tanto en lo que se refiere a funcionalidad como a usabilidad.

Podría implementarse la posibilidad de elegir la información que se muestra y que se incluye en los informes, ya que algunos datos son útiles de cara a la investigación, pero no resultan prácticos para los usuarios estándar, por ejemplo, un profesor que quiere analizar el foro de una asignatura que imparte.

Puede implementarse análisis sintáctico para mejorar el análisis de los foros.

La representación de la red podría incorporar un carácter dinámico al estilo de *Gephi*, pudiendo alterar la representación arrastrando los nodos manualmente y/o eligiendo diferentes estilos de representación, por ejemplo, aristas curvas.

Se puede mejorar el análisis combinado de sentimientos, tanto incluyendo otros algoritmos, como mejorando la fórmula del mismo. En este sentido podría facilitarse la modificación de la configuración a través de la propia interfaz de la aplicación en lugar de tener que acceder directamente al archivo de configuración.

También puede mejorarse el análisis de emociones y de sentimientos utilizando otras técnicas, como pueda ser por ejemplo la integración con plataformas de IA como por ejemplo ChatGPT³⁴, mediante la API correspondiente.

³⁴ <https://es.wikipedia.org/wiki/ChatGPT>

Apéndice A

Ayuda de la aplicación

Como usar la aplicación

Para analizar foros debe crear primero un proyecto.

Si aún no ha creado ningún proyecto vaya a *Proyecto* y elija la opción *Crear Proyecto*. A continuación asigne un nombre de proyecto y pulse *Enter* o haga clic en *Aceptar*. **En la pantalla de Proyecto dispone de una ayuda** que muestra el proceso de creación de un proyecto.

Una vez haya creado un proyecto podrá guardarlo mediante el botón *Guardar Proyecto* de la parte superior. Cuando haya guardado un proyecto podrá utilizarlo en el futuro a través de la opción *Abrir Proyecto* de la misma pantalla. Cuando ya se disponga de un nombre de proyecto se puede proceder a cargar datos de Foros o corpus para los análisis de sentimientos y emociones. La aplicación incorpora corpus por defecto.

Cargar Datos de Foros

Haga clic en el botón *Cargar Datos de Foros*. Se abrirá una ventana del explorador para elegir el archivo con los datos de los foros. Esta aplicación admite tres formatos para los datos foros: csv, xlsx y txt formateado según la exportación de los foros UNED. Se dispone de gran flexibilidad tanto en el formato como en el nombre de las columnas de los datos, pero a cambio debe realizarse una tarea de preparación o normalización de los datos.

Esta normalización consiste en especificar en la aplicación cual es el significado de cada columna. La normalización se inicia al momento de elegir el archivo con los datos de los foros. A través de menús desplegables elija una columna del archivo para cada variable o atributo (remitente, objetivo, foro, hilo, id. mensaje, título del mensaje, texto del mensaje y fecha). Deberá ir fijando con el botón *Fijar* las columnas elegidas para cada atributo y una vez finalizada la asignación pulsar en *Validar y procesar*.

Nótese que dispone de la posibilidad de utilizar los botones de *Atajo de Formato* para utilizar normalizaciones predefinidas.

Cuando haya realizado la validación el archivo con los datos de los foros habrá quedado normalizado para su uso y podrá acceder a los diferentes módulos.

Análisis de Sentimientos, Análisis de Emociones, Construir una nube de palabras con las palabras más utilizadas en los foros, Análisis temporal y Modelado de usuarios y generación de informes.

Se dispone de archivos de ayuda para cada módulo específico salvo para la creación de nube de palabras, dada su simplicidad. Puede consultar los archivos de ayuda desde cada uno de dichos módulos o a través de los siguientes enlaces:

- [Ayuda del Módulo de Análisis de Red.](#)
- [Ayuda del Módulo de Análisis de Sentimientos.](#)
- [Ayuda del Módulo de Análisis de Emociones.](#)
- [Ayuda del Módulo de Análisis Temporal.](#)
- [Ayuda del Módulo de Modelado e Informes.](#)

A.1 Ayuda del módulo de Análisis de Red

Al acceder al módulo de *Análisis de la Red* se muestra en primer lugar una pantalla en la que se pueden ver los valores de los nodos. Además, se muestra en la parte superior el **número de nodos y de aristas de la red formada por los participantes de los foros.**

Valores de los Nodos

Al abrir el módulo o al pulsar el botón *Ver Valores de los Nodos* se muestra la vista de la figura 85 en la que se ven los valores de los diferentes estadísticos para cada nodo de la red formada por los participantes de los foros en forma de tabla.

La columna *NODO* se corresponde con el nodo constituido por cada usuario, por lo que aquí aparece el nombre de cada usuario (nodo de la red).

La columna *ALIAS* muestra el alias asignado automáticamente a cada usuario por la aplicación.

Las siguientes columnas muestran los valores para *GRADO*, *GRADO DE SALIDA* y *GRADO DE ENTRADA*.

- **GRADO:** Suma del número de interacciones.
- **GRADO DE SALIDA:** Suma del número de respuestas enviadas por el usuario.
- **GRADO DE ENTRADA:** Suma del número de respuestas recibidas por el usuario.

Por último, las siguientes columnas representan los estadísticos:

- **CENTRALIDAD DE GRADO:** Es la primera y más simple de las medidas de centralidad. Los valores de centralidad de grado se normalizan dividiendo por el grado máximo posible en un gráfico simple $n-1$ donde n es el número de nodos.
- **CERCANÍA:** La centralidad de cercanía de un nodo u es el recíproco de la distancia promedio del camino más corto a u sobre todos los $n-1$ nodos

alcanzables. $C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)}$, donde $d(v, u)$ es la distancia del camino más corto entre v y u , y $n-1$ es el número de nodos accesibles desde u .

- **INTERMEDIACIÓN:** La centralidad de intermediación de un nodo v es la suma de la fracción de todos los pares de caminos más cortos que pasan por v $C_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$ donde V es el conjunto de nodos, $\sigma(s, t)$ es el número de caminos más cortos entre otros dos nodos s y t y $\sigma(s, t|v)$ es el número de esos caminos que pasan a través de v . Si $s = t$ $\sigma(s, t) = 1$ y si $v \in s, t$ $\sigma(s, t|v) = 0$.
- **CENTRALIDAD DE VECTOR PROPIO:** Es una medida de centralidad basada en la centralidad de los nodos vecinos.
- **PAGE RANK:** PageRank calcula una clasificación de los nodos en el gráfico en función de la estructura de los enlaces entrantes. Originalmente fue diseñado como un algoritmo para clasificar páginas web.
- **INICIATIVA:** Representa el número de “conversaciones” (hilos) iniciados por ese nodo (usuario).
- **MODULARIDAD:** Asigna a cada nodo la pertenencia a una “comunidad” en base a las aristas formadas por los envíos y recepciones de mensajes.

Analisis de Foros UNED. Versión Alfa 1.5

Archivo Configuración Preferencias Ayuda

Red del proyecto Demo ver. Alfa 1.5

Nodos: 55 Aristas: 130

Estadísticas de la red

Centralidad de Grado ? Cercanía ? Intermediación ? Centralidad de Vector Propio ? Page Rank ? Iniciativa ? Modularidad ? Sentimientos ? Ver Valores de los Nodos

Etiquetas de gráfico: Alias Sentimiento: N.Bayes (Corpus Usuario) Ayuda

Filtrar por: todo Seleccionar: TODO Visualizar archivo Fecha inicial Fecha final Calcular fecha

	NODO	ALIAS	GRADO	GRADO	GRADO	CENTRAL	CERCANI	INTERME	VECTOR	PAGE_R	INICIATIV	MODULA
1	37ac418bf9558984fbcdae5f4400780080	u1	3	1	2	0.056	0.48	0.028	0.12	0.019	1	0
2	ecafaa42d80d83a69b2f2311930cf62d55	u2	65	50	15	1.20	0.92	0.46	0.67	0.28	3	1
3	45c2fa9a13586c13ee4eff1669425d88a7	u3	6	4	2	0.11	0.5	0.03	0.17	0.048	2	0
4	eec276fc5d650417550f3a042c1a0fe718f	u4	2	1	1	0.037	0.48	0.028	0.12	0.019	1	1
5	0aad920a63c8e397ae53ee34c0ae076a	u5	16	8	8	0.3	0.52	0.077	0.31	0.06	4	5
6	e4fc98fc073fca23a32301cceb2f74a355f	u6	6	2	4	0.11	0.49	0.077	0.15	0.024	0	3
7	a8b41b15be0686c2d49490641de31047	u7	11	5	6	0.2	0.51	0.083	0.25	0.05	2	2
8	be273ceabe66f0b0ef2f29336b16ca4117	u8	9	4	5	0.17	0.5	0.06	0.2	0.034	2	5
9	1ed75217e05b580a3e51b67db2ac5737	u9	7	3	4	0.13	0.5	0.076	0.18	0.028	2	5
10	0c95e4a48870ebba5408ae3d327db2e	u10	4	2	2	0.074	0.49	0.028	0.17	0.025	1	5
11	7c95be4c421ca601382baf8b23065dfa5	u11	3	2	1	0.056	0.49	0.028	0.14	0.028	2	4
12	3b56b4a31259250e60e63a98b4326e5b3	u12	8	3	5	0.15	0.49	0.051	0.2	0.031	3	2
13	f4b542bfe437fb565b06569f2e37b61839	u13	2	1	1	0.037	0.48	0.028	0.12	0.019	1	1
14	fca7242564f380a2789f1dbb87bbe5b6fe	u14	2	1	1	0.037	0.48	0.028	0.12	0.019	0	1
15	a7a2d72220b27a8472e1e9c8b251a31a9	u15	4	2	2	0.074	0.49	0.038	0.12	0.021	0	4
16	3ad6f72cd9b5208dfe50d2b8634fa5f613	u16	13	7	6	0.24	0.39	0.1	0.2	0.039	2	3
17	24fb77d6ef839d751679f9c2db497de78a	u17	10	6	4	0.19	0.38	0.038	0.17	0.029	1	5
18	ef4fb0f3390abd9db1f6096e684833bbc	u18	3	1	2	0.056	0.28	0.028	0.036	0.0082	0	3
19	aa8c6473812dd80c8bb037015585c519	u19	7	3	4	0.13	0.29	0.065	0.038	0.012	1	3

55 rows x 12 columns

Figura 85. Valores de los Nodos.

Visualización de la red formada por los participantes atendiendo a diferentes estadísticos

Mediante los diferentes botones mostrados en la parte superior se representa el grafo de la red indicando mediante una escala de color cada medida de centralidad para todos los nodos. Por ejemplo, pulsando el botón *Centralidad de Grado* se obtiene la representación de la figura 86. Los tamaños de los nodos representan el GRADO de los mismos.

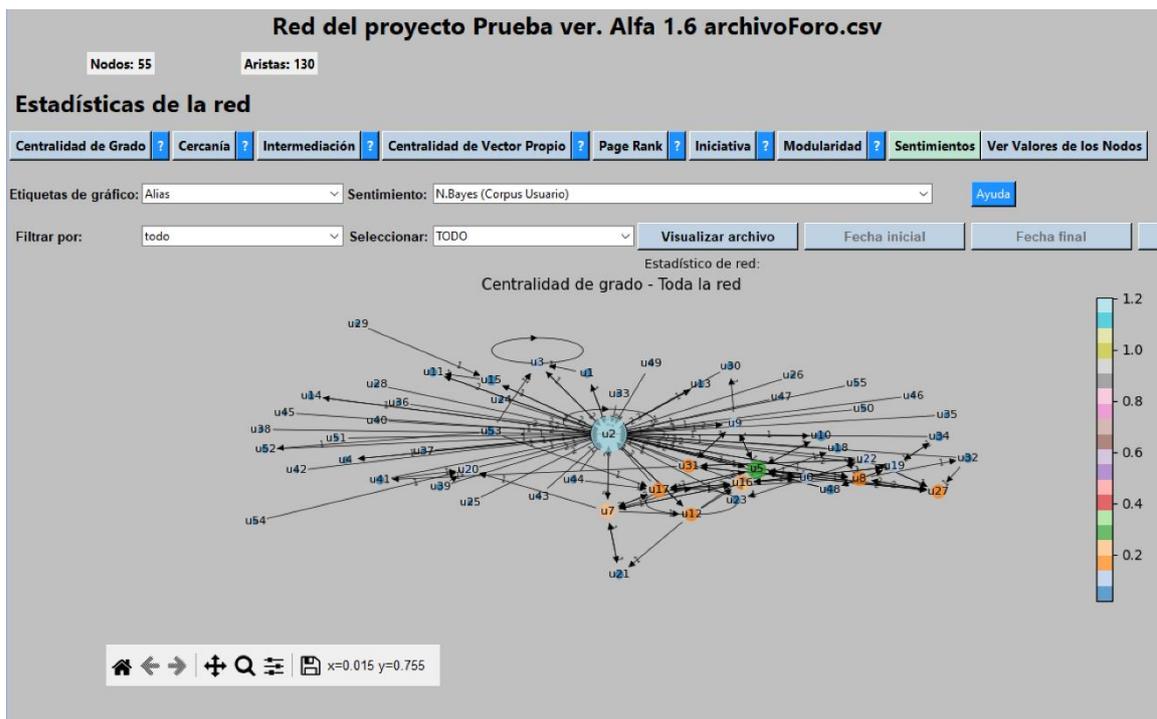


Figura 86. Representación de la centralidad de Grado.

Cada uno de estos botones se corresponden con las medidas de centralidad (estadísticos) descritos [más arriba](#).

Etiquetas de gráfico

Mediante un desplegable situado en parte superior izquierda se puede elegir si deseamos que los nodos se etiqueten con el nombre de usuario, el alias asignado automáticamente por la aplicación o bien que aparezcan sin etiqueta alguna. Ver la figura 87.



Figura 87. Selector de etiquetas para los nodos.

Red de Sentimientos

Mediante el desplegable denominado *Sentimiento* se puede elegir entre dos algoritmos para representar la “red de sentimientos”. Esto es una representación de la red en la que los colores de las aristas representan el sentimiento de los mensajes.

Los dos algoritmos elegibles son Naive Bayes mediante un corpus de entrenamiento cargado por el usuario (Se facilita un corpus por defecto) y PySentimiento³⁵, que es una librería open source.

- Naive Bayes otorga a cada mensaje (en este caso arista) un valor de positivo, representado en verde o negativo, representado en rojo. Ver figura 88.
- PySentimiento otorga a cada mensaje (en este caso arista) un valor de positivo, representado en verde, negativo, representado en rojo o neutro, representado en negro. Ver figura 89.

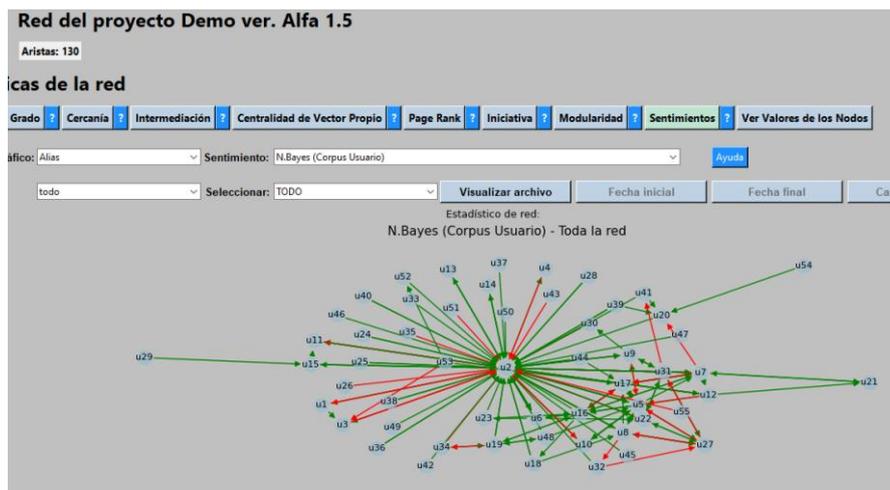


Figura 88. Red de Sentimientos N. Bayes.

³⁵ <https://github.com/pysentimiento/pysentimiento>

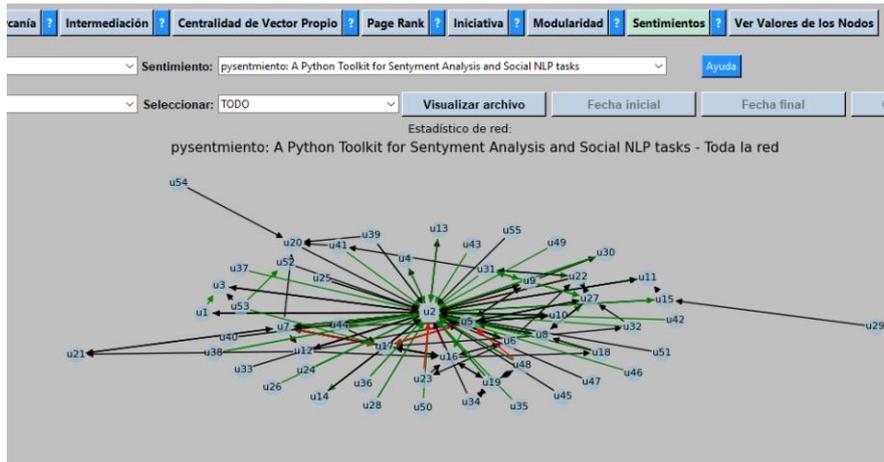


Figura 89. Red de Sentimientos PySentimiento.

Herramientas de Gráfico

Existe una caja de herramientas de gráfico en la parte inferior izquierda de los mismos. Ver figura 90.



Figura 90. Caja de Herramientas de Gráfico (grafo).

Esta caja de herramientas tiene diferentes iconos que permiten:

- Mover el gráfico 
- Hacer zoom. 
- Cambiar parámetros de visualización del gráfico 
- Guardar gráfico como un archivo de imagen por 
- Deshacer y rehacer acciones de la caja de herramientas 
- Restablecer el gráfico a la primera representación 

Visualizar Archivo

Mediante el botón *Visualizar Archivo* se muestra una tabla que contiene toda la información utilizada en el módulo. Ver figura 91.

Red del proyecto Demo ver. Alfa 1.5

Nodos: 55 Aristas: 130

Estadísticas de la red

Centralidad de Grado Cercanía Intermediación Centralidad de Vector Propio Page Rank Iniciativa Modularidad Sentimientos Ver Valores de los Nodos

Etiquetas de gráfico: Alias Sentimiento: pysentimiento: A Python Toolkit for Sentiment Analysis and Social NLP tasks Ayuda

Filtrar por: todo Seleccionar: TODO Visualizar archivo Fecha inicial Fecha final Calcular fecha

	fecha	Hora	Título mensaje	mensaje	Caractere	Hilo, idMensaje
1	4/1/2021	23:18:15	Calificaciones de las prácticas 3 y 4.	buenas tenia entendido que a quienes entr	289	Nadie
2	5/1/2021	16:48:28	Re: Calificaciones de las prácticas 3 y 4.	buenas tardes puede que no se dejara cla	196	('Calificaciones de las prácticas 3 y 4
3	5/1/2021	23:27:56	Re: Calificaciones de las prácticas 3 y 4.	despiste mio muchas gracias por la aclar	52	('Calificaciones de las prácticas 3 y 4
4	4/1/2021	11:59:45	Consulta examen	buenos dias he visto que a la hora de hac	286	Nadie
5	5/1/2021	16:43:58	Re: Consulta examen	buenas tardes avex no ofrece la opcion de	176	('Consulta examen', '407672367a1dc5a
6	12/12/2020	19:12:48	Tema 10 - no entrena con todas las muest	hola estoy siguiendo el ejemplo del libro	403	Nadie
7	12/12/2020	19:25:04	Re: Tema 10 - no entrena con todas las m	hola jorge en realidad si que esta tomand	175	('Tema 10 - no entrena con todas las m
8	12/12/2020	19:30:50	Re: Tema 10 - no entrena con todas las m	buenas jorge 1719 no son las instancias	244	('Tema 10 - no entrena con todas las m
9	12/12/2020	20:21:43	Re: Tema 10 - no entrena con todas las m	muchisimas gracias me estaba volviendo lo	149	('Tema 10 - no entrena con todas las m
10	12/12/2020	21:20:37	Re: Tema 10 - no entrena con todas las m	me paso lo mismo si asi que ejecute el n	171	('Tema 10 - no entrena con todas las m
11	13/12/2020	12:51:53	Re: Tema 10 - no entrena con todas las m	gracias de nuevo y continuando con este	288	('Tema 10 - no entrena con todas las m
12	16/12/2020	10:39:35	Re: Tema 10 - no entrena con todas las m	hola creo que a mi no me esta ocurriendo	308	('Tema 10 - no entrena con todas las m
13	17/12/2020	15:07:00	Re: Tema 10 - no entrena con todas las m	es ejecutando tal cual el notebook del te	220	('Tema 10 - no entrena con todas las m
14	0/12/2020	23:03:06	Re: Tema 10 - no entrena con todas las m	buenas noches jorge el problema que encu	607	('Tema 10 - no entrena con todas las m
15	11/12/2020	11:08:12	Re: Tema 10 - no entrena con todas las m	hola he probado como dices y sigo tenien	420	('Tema 10 - no entrena con todas las m
16	7/12/2020	21:20:22	Re: Tema 10 - no entrena con todas las m	hola solo por si sirve de ayuda dejo un	438	('Tema 10 - no entrena con todas las m
17	19/12/2020	08:24:40	Entrenamiento y evaluación de modelos	hola es un offtopic pero en el libro hab	973	Nadie
18	16/10/2020	12:19:09	Presentación de los participantes en esta	buenos dias mi nombre es antonio rodrigue	052	Nadie
19	16/10/2020	12:29:51	Re: Presentación de los participantes en	buenos dias me presento siguiendo las pre	1553	('Presentación de los participantes en e

229 rows x 26 columns

buenas tardes avex no ofrece la opcion de escribir en latex en el caso de escribir formulas la opcion de hacer una foto y subirla es lo mas acertado saludos antonio

Figura 91. Tabla de visualización de archivo.

Aquí se muestran las diferentes columnas del archivo de datos.

Filtros

Este módulo implementa el filtrado de datos mediante los valores de las diferentes columnas del archivo y de la fecha de publicación de los mensajes.

Para filtrar se debe seleccionar un elemento en el desplegable *Filtrar por* y a continuación seleccionar un valor del desplegable *Seleccionar*.

Si deseamos filtrar por fecha se habilitará primero el botón *Fecha inicial* y cuando seleccionemos del calendario que aparece una fecha se habilitará el botón *Fecha final*. Cuando seleccionemos del calendario la fecha de fin para el intervalo a filtrar se habilitará el botón *Calcular fecha*. Ver figura 92.

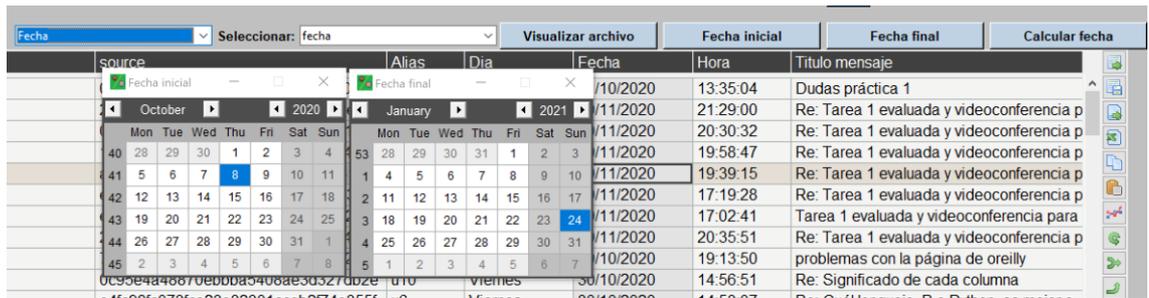


Figura 92. Filtro por fecha.

Una vez pulsado el botón *Calcular fecha* se aplicará el filtro, de manera que cualquier acción que realicemos a continuación se limitará a los mensajes que coincidan con dicho filtro. Por ejemplo, en la figura 93 se puede ver el filtro aplicado en la figura 93. Sobre la representación Page Rank de la red. El filtro se muestra en el título del gráfico.



Figura 93. Visualización de la red mostrando Page Rank con un filtro de fecha aplicado.

Los filtros aplicados también afectan a la representación de la tabla del archivo y de los valores de los nodos.

NOTA

Cuando no se aplica ningún filtro se muestran los bucles o auto-aristas (aristas que salen y entran al mismo nodo) que representan respuestas a mensajes enviados por el mismo usuario o auto-respuestas. Para omitir este tipo de aristas use el filtro y seleccione

todo. En la figura 94 se ve el grafo de cercanía de toda la red con autorrespuestas, mientras que en la figura 95 se ve el mismo grafo sin auto-respuestas.

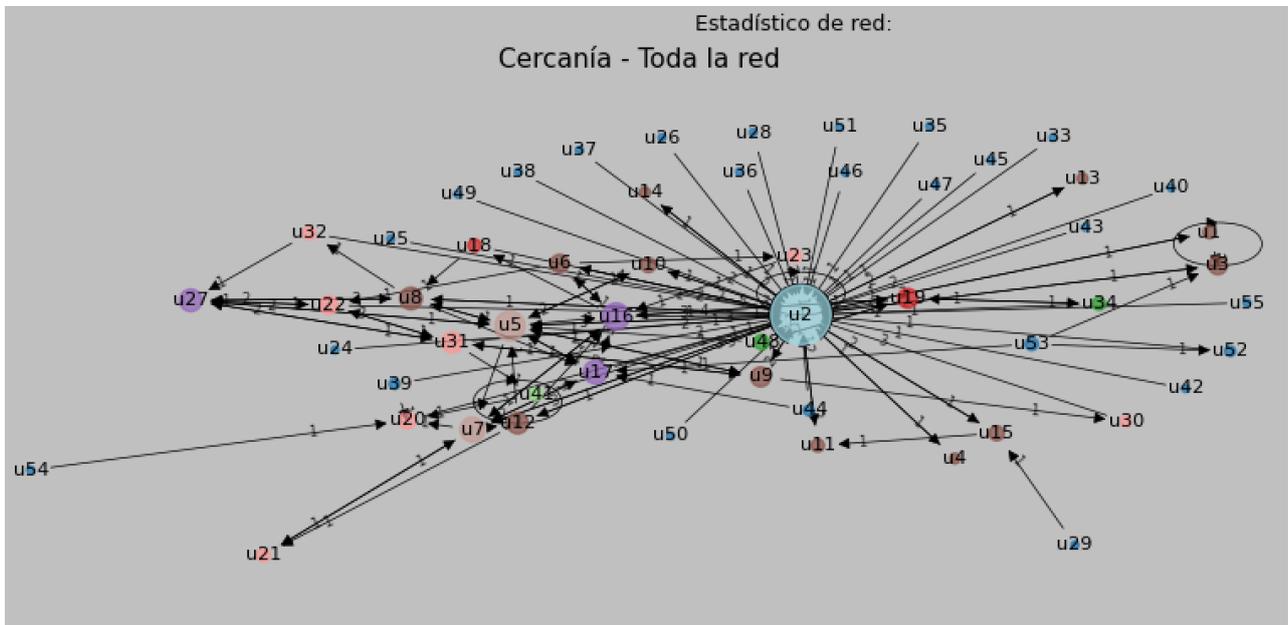


Figura 94. Grafo de Cercanía con auto-respuestas.

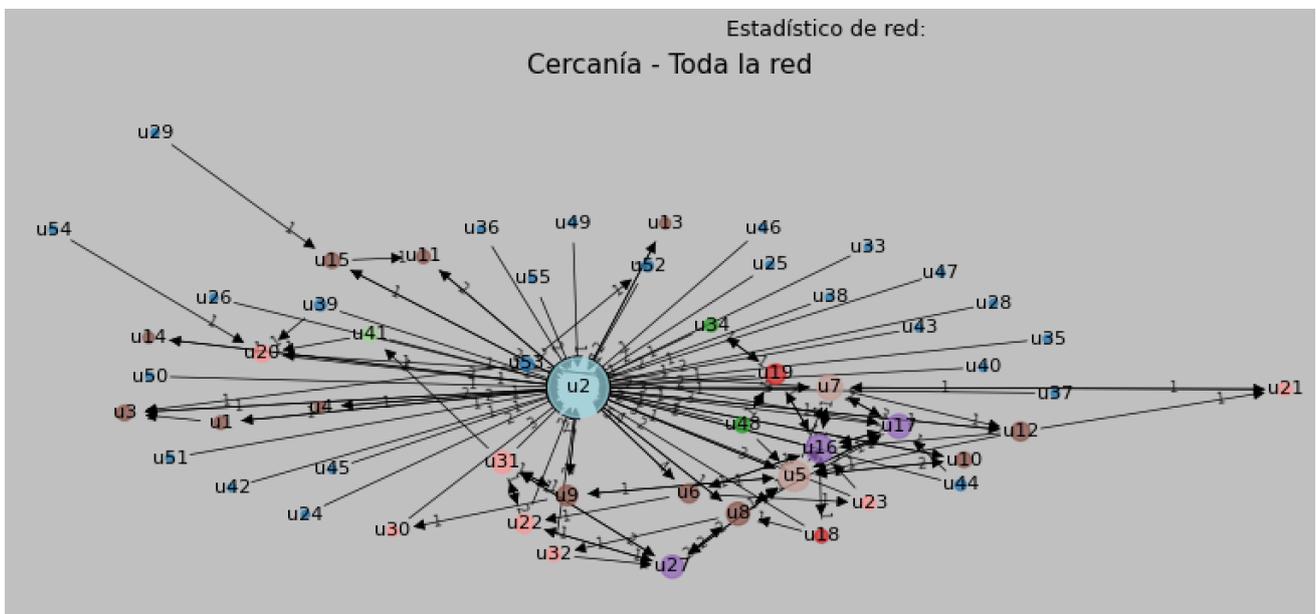


Figura 95. Grafo de Cercanía sin auto-respuestas.

Exportar Red

Desde la versión Beta 3.0 se encuentra disponible la función de exportar la red mediante un botón incluido a tal efecto en la parte superior, a la derecha del de Ayuda. Se puede ver la figura 96 a continuación, rodeado con una circunferencia de color rojo.

Red del proyecto proy 1

Nodos: 28 Aristas: 148

Estadísticas de la red

Centralidad de Grado ? Cercanía ? Intermediación ? Centralidad de Vector Propio ? Page Rank ? Iniciativa ? Modularidad ? Sentimientos Ver Valores de los Nodos

Etiquetas de gráfico: Alias Sentimiento: N.Bayes (Corpus Usuario) Ayuda **Exportar Red**

Filtrar por: todo Seleccionar: TODO Visualizar archivo Fecha inicial Fecha final Calcula fecha

	NODO	ALIAS	GRADO	GRADO	GRADO	CENTRAL	CERCANI	INTERME	VECTOR	PAGE_R	INICIATIV	MODULA
1	Prof-130	u1	0	0	0	0	0	0	4.4e-13	0.0063	1	0
2	Estu-115	u2	26	16	10	0.96	0.64	0.14	0.4	0.11	6	1
3	Estu-56	u3	16	7	9	0.59	0.5	0.073	0.25	0.046	1	2
4	Estu-39	u4	21	13	8	0.78	0.58	0.13	0.27	0.076	4	5
5	Estu-14	u5	8	5	3	0.3	0.47	0.064	0.18	0.042	2	1
6	Estu-103	u6	7	4	3	0.26	0.45	0.065	0.13	0.031	3	1
7	Estu-21	u7	16	6	10	0.59	0.49	0.08	0.2	0.041	2	2
8	Coor-36	u8	39	18	21	1.44	0.7	0.4	0.42	0.11	19	1
9	Estu-40	u9	10	5	5	0.37	0.48	0.068	0.18	0.038	2	1
10	Estu-159	u10	10	5	5	0.37	0.47	0.066	0.18	0.035	2	1

Figura 96. Botón de Exportar Red.

Pulsando este botón se crea un archivo Excel con los datos de la red, llamado *red_nombreProyecto.xlsx* en la carpeta principal del proyecto.

A.2 Ayuda del módulo de Análisis de Sentimientos

Sentimiento medio

Al abrir el módulo o al pulsar el botón *Tabla de Sentimientos* se muestra la vista de la tabla de **sentimiento medio combinado** de cada usuario de los foros. El concepto de *sentimiento Combinado* forma parte del objetivo de este TFM. Se trata de una combinación de diferentes métodos de análisis de sentimientos de código abierto y de sencilla implementación fácilmente accesibles en *la comunidad*.

Este **sentimiento combinado de cada mensaje** se calcula según la siguiente fórmula por configurable³⁶:

$$Sent\ Comb = \frac{coefNB * NBnum + coefPyS * PySnum + coefSentSpNorm * Sent.Norm^{pot}}{3} * coefTot$$

Por defecto: coefNB = 1.0; coefPyS = 1.0; coefSentSpNorm = 1.5; pot = 0.5; coefTot = 0.01

Los coeficientes anteriores se configuran en el archivo *config.ini*.

Donde

NBnum es una transformación de un literal a un valor numérico. El literal proviene de hallar el sentimiento según una implementación de *Naive Bayes*.

La correspondencia entre el literal y el valor numérico es la siguiente:

POSITIVO = 0.75

NEGATIVO = 10^{-1}

³⁶ Durante el desarrollo del trabajo se han realizado varias pruebas con diferentes fórmulas y coeficientes, comprobando los resultados por medios humanos, por lo que se puede considerar una tarea de ensayo y error supervisada. No obstante, esta cuestión queda abierta, pudiendo seguir realizando modificaciones en estas fórmulas.

PySnum es otra transformación de literal a valor numérico donde el literal proviene de hallar el sentimiento según la librería *PySentimiento*³⁷.

En este caso la correspondencia es como sigue:

POSITIVO = 1

NEGATIVO = 10^{-2}

NEUTRO = 0.5

Sent. Norm es una normalización como sigue:

$$Sent. Norm = \frac{Min(Sentimiento)}{Max(Sentimiento) - Min(Sentimiento)}$$

Donde *Sentimiento* es el sentimiento calculado por el algoritmo de la librería Python *Sentiment Analysis Spanish*³⁸.

* El *sentimiento Medio Combinado* del participante *p* es la **media aritmética** del **Sentimiento Combinado para cada usuario** según la fórmula:

$$Sent Medio Comb. (p) = \frac{1}{N} \sum_{i=1}^N Sent Comb (i)$$

Donde los ítems *i* son los mensajes emitidos por el participante *p*.

³⁷ <https://github.com/pysentimiento/pysentimiento>

³⁸ <https://pypi.org/project/sentiment-analysis-spanish/>

En la figura 97 se muestra la vista de Sentimiento Medio Combinado.

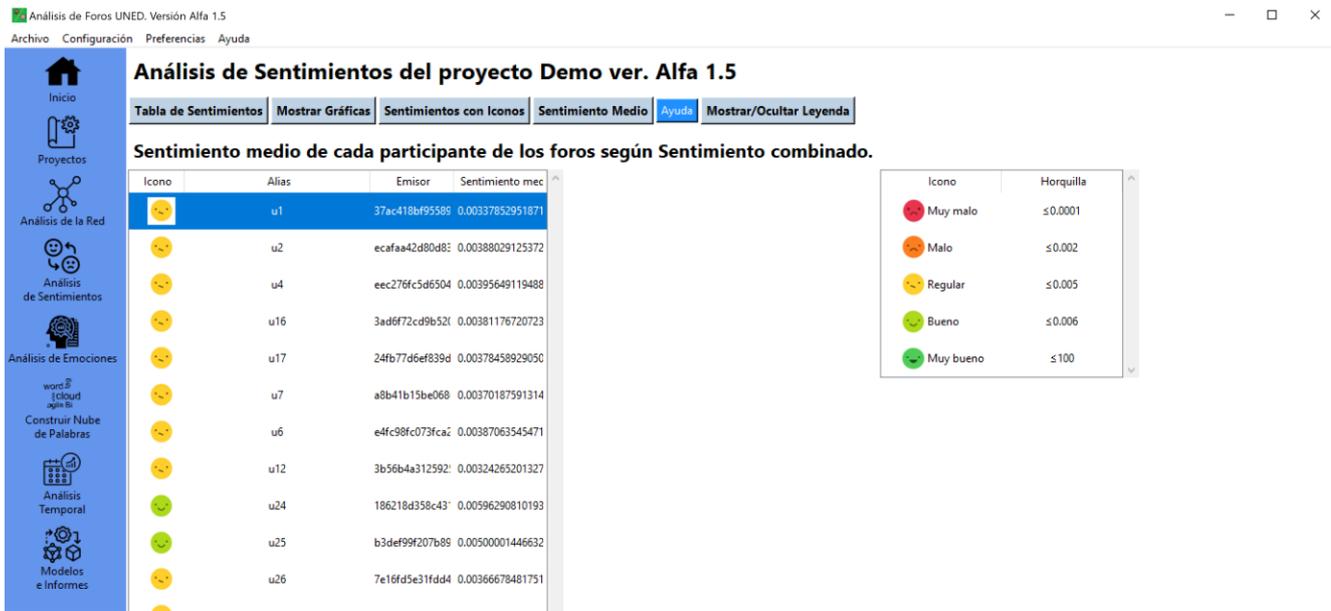


Figura 97. Vista de sentimiento medio combinado.

El valor numérico del sentimiento combinado se corresponde con un literal y un icono determinado según la siguiente gráfica (figura 98).

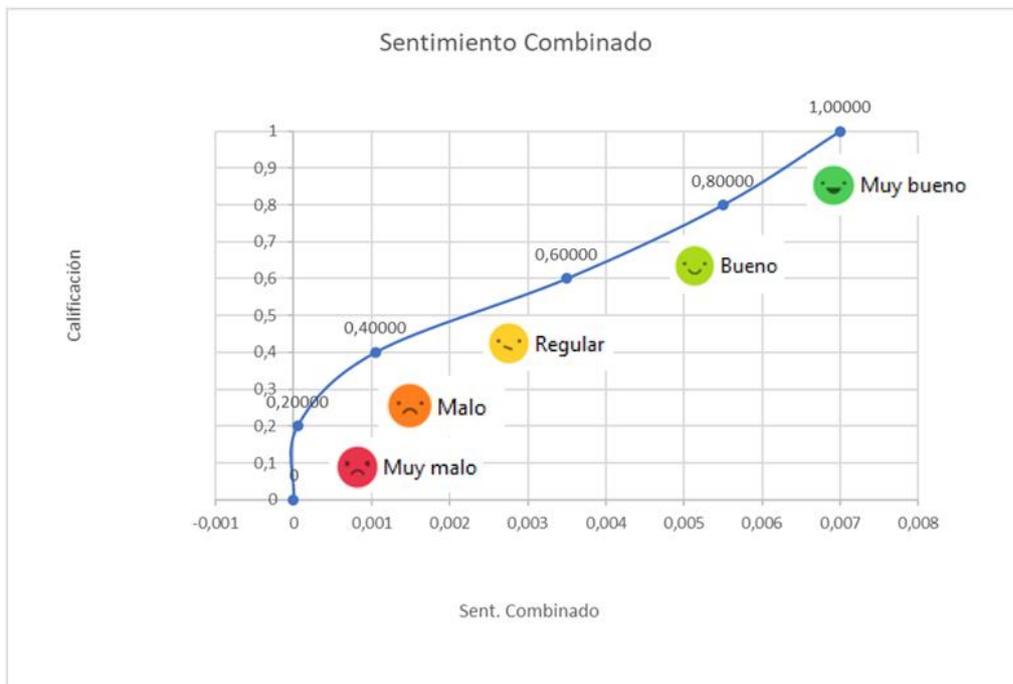


Figura 98. Correspondencia entre el valor numérico del sentimiento combinado y el literal e icono correspondiente.

La **columna *Icono*** representa de manera visual el sentimiento medio combinado de cada participante mediante un icono. Haciendo uso del botón ***Mostrar/Ocultar Leyenda*** podemos mostrar u ocultar una leyenda que informa de la correspondencia entre los iconos y los intervalos de los valores para [Sent. Comb.](#)

La **columna *Alias*** muestra el alias asignado a cada usuario de manera automática por la aplicación. Esto cumple una doble funcionalidad: por un lado se da la opción de anonimizar los datos para su representación y por otro se acorta la identificación de los usuarios para que la representación de los gráficos sea más clara.

La **columna *Emisor*** muestra el identificador de cada emisor (participante) tal cual aparece en el archivo de datos de foros utilizado para el análisis.

La **columna *Sentimiento medio*** muestra el valor numérico calculado para el sentimiento medio combinado ([Sent. Comb.](#)).

Tabla de sentimientos

Mediante el botón *Tabla de Sentimientos* se accede a la vista de la tabla que se muestra en la figura 99 con la información de sentimientos de cada mensaje de los foros.

	Emisor	Alias	Destinatario	Sentim. Analys. Sp.	Sentim. Analys. Sp. Norm.	Sent. N. B.	PySentimi	Sent. Combinad	Mensaje	Foro	Hi
30	25055	u9	79	3e-10	3e-10	pos	POS	0.0005	buenas tardes	1	11
37	79	u1	22946	3.1e-08	3.1e-08	pos	NEU	0.0042	hola pedro	1	74
38	79	u1	71704	2.2e-25	2.2e-25	pos	NEU	0.0042	buenos días adrian	1	11
39	69829	u8	79	0.012	0.012	pos	NEU	0.0047	hola	1	11
40	22946	u4	79	0.47	0.47	neg	NEU	0.0054	no desde el principio bueno	1	74
41	25055	u9	79	1.5e-70	1.5e-70	pos	NEU	0.0042	siguiendo el acercamiento e	1	10
42	66182	u2	79	1	1.00	neg	NEU	0.007	analysis of moodle analytics	1	10
43	39712	u6	79	8.5e-36	8.5e-36	pos	NEU	0.0042	buenas tardes envio mi anal	1	10
44	39712	u6	Inicia el h	7e-13	7e-13	pos	NEU	0.0042	buenos días antonio	1	14
45	25964	u7	23709	4e-16	4e-16	pos	NEU	0.0042	hola a todos para conocemo	1	51
46	71704	u5	79	1.0e-112	0	pos	NEU	0.0042	epistemologia que estamos	1	10
47	79	u1	39712	5.4e-05	5.4e-05	neg	NEU	0.002	buenos días maria	1	14
48	66967	u14	79	6.4e-16	6.4e-16	pos	NEU	0.0042	buenas tardes	1	51
49	66967	u14	Inicia el h	2.6e-11	2.6e-11	pos	NEU	0.0042	buenas tardes	1	14
50	71704	u5	Inicia el h	1.1e-26	1.1e-26	pos	NEU	0.0042	hola	1	14
51	66182	u2	25055	4.2e-07	4.2e-07	pos	NEU	0.0042	hola miguel	1	10
52	69829	u8	Inicia el h	1.8e-21	1.8e-21	pos	NEU	0.0042	hola	1	15
53	69829	u8	71704	1.3e-10	1.3e-10	pos	NEU	0.0042	buenas noches	1	14
54	69829	u8	66967	3.4e-10	3.4e-10	pos	NEU	0.0042	buenas noches	1	14

Figura 99. Tabla de Sentimientos.

La información de las diferentes columnas de la tabla de sentimientos se puede ver al seleccionar cada celda de la tabla en la parte inferior de la pantalla y es la siguiente:

- **Emisor:** Nombre del emisor del mensaje de cada fila.
- **Alias:** Alias asignado automáticamente por la aplicación para acortar los nombres de usuario y anonimizarlos.
- **Destinatario:** Destinatario del mensaje representado en cada fila.
- **Sentim. Analys. Sp.:** Valor devuelto por el algoritmo implementado en la librería [sentiment-analysis-spanish](#).
- **Sentim. Analys. Sp. Norm.:** Valor anterior normalizado [descrito más arriba](#). Se muestra con fondo verde.
- **Sent. N. Bayes:** Valor devuelto por el algoritmo Naive Bayes utilizando el corpus . Este algoritmo devuelve *neg* para mensajes negativos y *pos* para mensajes clasificados como positivos. Se muestra con fondo marrón.
- **PySentimiento:** Valor devuelto por el algoritmo implementado en la librería [PySentimiento](#). Este algoritmo devuelve tres valores posibles: *NEG*, *POS* y *NEU* para los mensajes clasificados como negativos, positivos y neutros respectivamente. Se muestra con fondo azul.
- **Sent. Combinado:** Valor de [Sent. Comb.](#) para el mensaje representado en esa fila. Se muestra con fondo amarillo.

- **Mensaje:** Se transcribe el mensaje representado en esa fila. Al seleccionar un mensaje se puede ver su contenido íntegro en la parte inferior de la pantalla.
- **Foro:** Nombre del foro donde se publicó el mensaje.
- **Hilo:** Hilo en el que se publicó el mensaje dentro del foro.

Las columnas que contienen los valores relevantes del análisis son las que se muestran con fondos de color.

- **Sentim. Analys. Sp. Norm.:** verde.
- **Sent. N. Bayes:** marrón.
- **PySentimiento:** azul.
- **Sent. Combinado:** amarillo.

Además de lo indicado anteriormente, los valores numéricos máximos de cada columna se muestran con fondo de color verde claro.

Gráficos

Al pulsar el botón *Mostrar Gráficos* se muestran los gráficos de la pantalla de la figura 100.

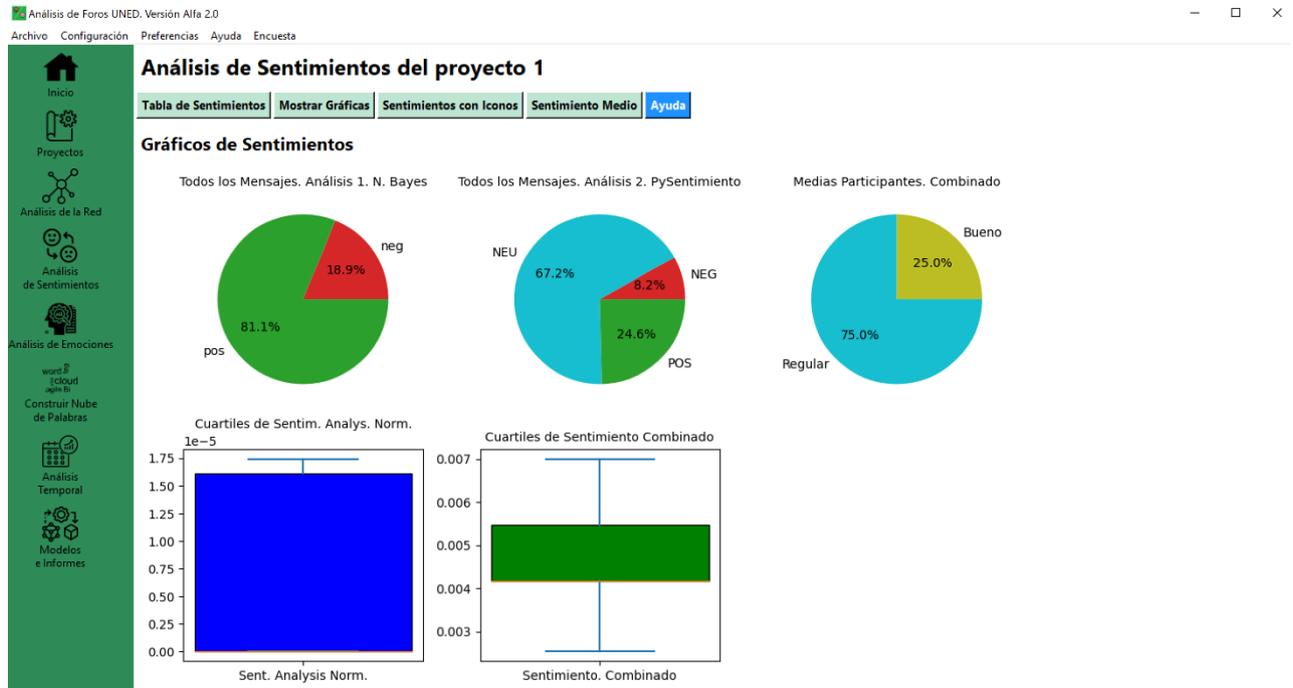


Figura 100. Gráficos de Sentimientos.

En esta pantalla se muestran tres gráficos de tarta y dos gráficos de caja.

Gráficos de tarta

- Todos los Mensajes Análisis 1. N. Bayes
 - o Se representa la proporción de mensajes positivos y negativos según este algoritmo.
- Todos los Mensajes. Análisis 2. PySentimiento
 - o Se representa la proporción de mensajes positivos, negativos y neutros según esta librería.

Gráficos de caja

- Cuartiles de Sentym. Analys. Norm.
 - o Se muestra un gráfico de caja para esta variable donde se pueden ver los cuartiles, incluida la mediana.

- Cuartiles de Sentimiento Combinado
 - o Se muestra un gráfico de caja para esta variable donde se pueden ver los cuartiles, incluida la mediana.

Análisis de Sentimientos de los mensajes mediante iconos

Al pulsar el botón *Sentimientos con Iconos* se muestra la tabla de la figura 101, la cual visualiza el sentimiento de cada mensaje.

Icono	Alias	Emisor	Destinatario	Sent. Analisis Sp	Sentimiento NB	PySentimiento	Sent. Combinado	Mensaje	Titulo	Foro	Hilo
😊	u5	71704	Inicia el hilo	1.31257192100502	neg	NEG	0.0003666666666666	hola	Tarea Tema 4 y fo	1	24137
😊	u1	79	71704	3.30547896430395	neg	NEG	0.00036666541331	en realidad buscar la buena tarde adi por las preguntas	Propuesta Model	1	14717
😊	u4	22946	79	6.99526029264355	neg	NEG	0.0003789094386	hola yo tambien	Re: Tarea 2.1	1	14001
😊	u7	25964	22946	8.12428991278946	neg	NEG	0.00041173408813	yo estoy igual no	Re: Tarea 2.1	1	14001
😊	u1	79	64950	0.00044604042104	neg	NEG	0.00047226595902	octavio no has en	Modelado Octavi	1	15233
😊	u2	95182	79	1.51499218864462	neg	NEU	0.00200001946145	hola antonio	Re: Modelado Ma	1	16582
😊	u3	64958	79	1.90632296142404	neg	NEU	0.00200002183072	hola el lunes no puede	Re: Charla telemá	1	7470
😊	u3	64958	79	5.88377856386672	neg	NEU	0.00200383529661	buenos días antoni como no voy a p	Re: Charla telemá	1	7470
😊	u4	22946	79	1.69616348910505	neg	NEU	0.00200651568010	hola yo no tengo	Re: Charla telemá	1	7470
😊	u8	69829	71704	6.05509360535465	neg	NEU	0.00201230354989	hola adrián no tengo conoc	Re: Respuesta a la	1	19064

TEXTO DEL MENSAJE:
buenos días antonio
como no voy a poder asistir a la charla telematica mi pregunta es si sera grabada para poder visualizarla luego
un saludo octavio

Figura 101. Tabla de análisis de sentimientos de cada mensaje con iconos.

La información representada en esta vista es exactamente la misma que se muestra en la [tabla de sentimientos](#) y la existencia de ambas se debe cuestiones técnicas de desarrollo, en tanto en cuanto en esta se muestra un objeto *Treeview* que permite introducir iconos de manera relativamente sencilla, cosa que no sucede con un objeto *Table*, que es el que se visualiza en otro caso. Por contrapartida resulta más sencillo capturar el contenido de una celda para representar su contenido en otro campo. Es por ello que en esta vista con iconos, sólo se pueden seleccionar filas y no celdas. Al seleccionar una fila se representa en la ventana de texto inferior el contenido íntegro del mensaje correspondiente a dicha fila.

Leyenda

Mediante el botón *Mostrar/Ocultar Leyenda* se puede mostrar u ocultar la leyenda que nos facilita interpretar los iconos. Funciona con un efecto biestable, de manera que si se pulsa cambia el estado visible u oculto de dicha leyenda. En la figura 102 se puede ver la leyenda visible en la vista de *Sentimientos con Iconos* mientras en la figura 98 se encuentra oculta.

Análisis de Sentimientos del proyecto Demo ver. Alfa 1.5

Tabla de Sentimientos | **Mostrar Gráficas** | Sentimientos con Iconos | Sentimiento Medio | Ayuda | **Mostrar/Ocultar Leyenda**

Sentimiento de cada mensaje según diferentes análisis. Icono = Sent. Combinado

Icono	Alias	Emisor	Destinatario	Sent. Analysis Sp	Sentimiento NB	PySentimiento	Sent. Combinado	Mensaje	Título	Foro	Hilo	Icono
	u1	37ac418bf95589	Inicia el hilo	4.42960700366342	neg	NEG	0.00037819446940	buenas tenia ent	Calificaciones de	Foro de consultas	Calificaciones de	
	u2	ecafaa42d80d8	37ac418bf955898	1.70316263966219	neg	NEU	0.00200022604348	buenas tardes pu	Re: Calificaciones	Foro de consultas	Calificaciones de	
	u1	37ac418bf95589	ecafaa42d80d83a	0.03888881544833	pos	NEU	0.00524679732963	despiste mio muc	Re: Calificaciones	Foro de consultas	Calificaciones de	
	u4	eec276fc5d6504	Inicia el hilo	2.64459136716653	pos	NEU	0.00416666694833	buenos dias he v	Consulta examen	Foro de consultas	Consulta examen	
	u2	ecafaa42d80d8	eec276fc5d65041	8.1509964315384	pos	NEU	0.00417161169110	buenas tardes av	Re: Consulta exar	Foro de consultas	Consulta examen	
	u16	3ad6f72cd9b52	Inicia el hilo	9.63782311071826	neg	NEU	0.00200000053771	hola estoy sigue	Tema 10 - no entr	Foro de consultas	Tema 10 - no entr	
	u17	24fb77d6ef839d	3ad6f72cd9b5208	3.17608862478775	pos	NEU	0.00417642802473	hola jorge en rea	Re: Tema 10 - no	Foro de consultas	Tema 10 - no entr	
	u7	a8b41b15be068	3ad6f72cd9b5208	2.67198946895069	pos	NEU	0.00417561992898	buenas jorge 171	Re: Tema 10 - no	Foro de consultas	Tema 10 - no entr	
	u16	3ad6f72cd9b52	a8b41b15be068c	0.01754777157660	pos	POS	0.00655889599272	muchisimas graci	Re: Tema 10 - no	Foro de consultas	Tema 10 - no entr	
	u17	24fb77d6ef839d	3ad6f72cd9b5208	6.516790826064e-	pos	NEU	0.00416806490378	me paso lo mism	Re: Tema 10 - no	Foro de consultas	Tema 10 - no entr	

TEXTO DEL MENSAJE:

buenas tenia entendido que a quienes entregaran las practicas a tiempo recibirian su correccion o como minimo la calificacion antes del examen sin embargo sigo sin ver mis practicas 3 y 4 calificadas entendi mal la dinamica o deberia tener las calificaciones saludos xabi

Figura 102. Leyenda visible en la pantalla de Sentimientos con Iconos.

A.3 Ayuda del módulo de Análisis de Emociones

Al acceder al módulo de *Análisis de Emociones* se muestra en primer lugar una pantalla en la que se pueden ver las *Emociones Promedio* de cada participante según *Regiones de Russell*.

Emociones Promedio de cada participante según regiones de Russell

Al abrir el módulo o al pulsar el botón *Medias Russell por Usuarios* se muestra la vista de la figura 103 en la que se ven las emociones promedio de cada usuario según *Regiones de Russell*³⁹.



Figura 103. Pantalla de Emociones Promedio de cada Participante según Regiones de Russell.

En esta pantalla se puede ver una tabla en la que se puede hacer *scroll* y en la que se muestra para cada usuario un icono representativo de la emoción. El identificador de usuario se muestra en la columna *Emisor*. En la columna *Alias* se representa el alias

³⁹ <https://www.affective-science.org/pubs/1999/RussellFB1999.pdf>

asignado automáticamente por la aplicación y en la columna *Emoción Promedio* aparecen las emociones promedio de cada participante de manera literal.

Estas emociones pueden ser: Feliz, Triste, Enfadado, Relajado, Sin Emoción, Miedo y Asco (Rechazo o Desprecio).

Al seleccionar una fila se muestra en la parte inferior de la pantalla el nombre de usuario completo, dado que puede ser que no quepa en el ancho por defecto de la columna, si bien el ancho de las columnas se puede redimensionar.

Para el cálculo de estas emociones se utiliza un método aritmético de bolsa de palabras y un corpus referido como corpus 1 que puede ser cambiado por el usuario. Por defecto se utiliza como corpus el archivo ubicado en la ruta:

`./recursos/ Spanish-es-NRC-VAD-Lexicon.txt`

El formato de este archivo se muestra en la figura 104.

Word	Spanish-es	Valence	Arousal	Dominance
aaaaaaah	aaaaaaah		0.479	0.606 0.291
aaaah	NO TRANSLATION	0.520	0.636	0.282
aardvark	cerdo hormiguero		0.427	0.490 0.437
aback	detrás	0.385 0.407	0.288	
abacus	ábaco	0.510 0.276	0.485	
abalone	abulón	0.500 0.480	0.412	
abandon	abandonar	0.052	0.519	0.245
abandoned	abandonado		0.046	0.481 0.130
abandonment	abandono		0.128	0.430 0.202
abashed	avergonzado	0.177	0.644	0.307
abate	disminuir	0.255	0.696	0.604
abatement	disminución		0.388	0.338 0.336
abba	Abba	0.562 0.500	0.480	
abbey	abadía	0.580 0.367	0.444	
abbot	NO TRANSLATION	0.427	0.321	0.483
abbreviate	abreviar		0.531	0.375 0.330
abbreviation	abreviatura		0.469	0.306 0.345
abdomen	abdomen	0.469 0.462	0.471	
abdominal	abdominal		0.490	0.456 0.445
abduct	secuestrar	0.173	0.720	0.615
abduction	secuestro		0.062	0.990 0.673
aberrant	aberrante		0.146	0.765 0.431
aberration	aberración		0.125	0.816 0.417
abeyance	suspensión		0.330	0.510 0.292
abhor	aborrecer	0.125	0.602	0.349
abhorrence	aborrecimiento		0.167	0.684 0.420
abhorrent	aborrecible		0.229	0.750 0.474
abide	acatar	0.635 0.354	0.705	
abiding	permanencia	0.796	0.327	0.750
ability	capacidad	0.875	0.510	0.816

Figura 104. Archivo utilizado como corpus por defecto para el cálculo de las emociones según Regiones de Russell.

Tabla de Emociones

Al pulsar el botón *Mostrar Tabla con Emociones* se muestra una tabla como la que se ve en la figura 105.

	Emisor	Alias	Destinatario	Emoción Re	Felicidad	Enojo	Tristeza	Miedo	Asco	Dominante	Mensaje	Foro	Hilo
1	37ac418bf95	u1	Inicia el hilo	Ninguna	0	0	0	0	0	N/A	buenas tenia	Foro de cons	Calificaciones
2	ecafaa42d80	u2	37ac418bf95	Ninguna	0	0	0	0	0	N/A	buenas tardes	Foro de cons	Calificaciones
3	37ac418bf95	u1	ecafaa42d80	Ninguna	0	0	0	0	0	N/A	despiste mio	Foro de cons	Calificaciones
4	eec276fc5d6	u4	Inicia el hilo	Triste	0	0	0	0	0	N/A	buenos dias	Foro de cons	Consulta ex
5	ecafaa42d80	u2	eec276fc5d6	Triste	0	0	0	0	0	N/A	buenas tardes	Foro de cons	Consulta ex
6	3ad6f72cd9b	u16	Inicia el hilo	Ninguna	0	0	0	0	0	N/A	hola estoy s	Foro de cons	Tema 10 - no
7	24fb77d6ef8	u17	3ad6f72cd9b	Ninguna	0	0	0	0	0	N/A	hola jorge e	Foro de cons	Tema 10 - no
8	a8b41b15be	u7	3ad6f72cd9b	Ninguna	0	0	0	0	0	N/A	buenas jorge	Foro de cons	Tema 10 - no
9	3ad6f72cd9b	u16	a8b41b15be	Ninguna	0	0	0	0	0	N/A	muchisimas gr	Foro de cons	Tema 10 - no
10	24fb77d6ef8	u17	3ad6f72cd9b	Ninguna	0	0	0	0	0	N/A	me paso lo mi	Foro de cons	Tema 10 - no
11	3ad6f72cd9b	u16	24fb77d6ef8	Ninguna	0	0	0	0	0	N/A	gracias de nu	Foro de cons	Tema 10 - no
12	24fb77d6ef8	u17	3ad6f72cd9b	Ninguna	0	0	0	0	0	N/A	hola creo qu	Foro de cons	Tema 10 - no
13	3ad6f72cd9b	u16	24fb77d6ef8	Ninguna	0	0	0	0	0	N/A	es ejecutando	Foro de cons	Tema 10 - no
14	e4fc98fc073	u6	3ad6f72cd9b	Ninguna	0	0	0	0	0	N/A	buenas noche	Foro de cons	Tema 10 - no
15	3ad6f72cd9b	u16	e4fc98fc073	Ninguna	0	0	0	0	0	N/A	hola he prob	Foro de cons	Tema 10 - no
16	3b56b4a312	u12	3ad6f72cd9b	Ninguna	0	0	0	0	0	N/A	hola solo po	Foro de cons	Tema 10 - no
17	3b56b4a312	u12	Inicia el hilo	Triste	0	0	0	0	0	N/A	hola es un o	Foro de cons	Entrenamient
18	ecafaa42d80	u2	Inicia el hilo	Ninguna	0	0	0	0	0	N/A	buenos dias	Foro de cons	Presentación
19	ecafaa42d80	u2	ecafaa42d80	Ninguna	0	0	0	0	0	N/A	buenos dias	Foro de cons	Presentación

Figura 105. Tabla de mensajes y emociones.

En esta tabla se representan todos los mensajes contenidos en el archivo cargado. Para cada mensaje se puede ver el Emisor, el alias asignado automáticamente a dicho emisor, el usuario destinatario del mensaje, la emoción según la región de Russell (primer análisis de emociones con fondo color mostaza), y las componentes del segundo análisis de emociones junto con su dominante (columna de fondo verde).

A continuación de las columnas mencionadas se muestran sendas columnas con el contenido del mensaje, el Foro en el que se publicó y el Hilo. Si seleccionamos una

celda de la tabla su contenido se muestra en la parte inferior de la pantalla. Ver figura 106.

147	a488	u10	ecataa4zda0	Triste	0	0	0	0	0	0	N/A	puntos noia al Foro de la Pr	l area 1 evalu
148	42d80	u2	0c95e4a488	Triste	0	0	0	0	0	0	N/A	puntos buenosForo de la Pr	Tarea 1 evalu
149	4c42	u11	Inicia el hilo	Triste	0	0	0	0	0	0	N/A	buenas tardes Foro de la Pr	Duda fecha d

229 rows x 13 columns

buenas tardes tengo una duda respecto a la fecha de entrega de la practica en el apartado tareas aparece como fecha limite 22 de noviembre pero en detalles aparece que puede ser entregada en la fecha de finales de diciembre si no recuerdo mal hay algun tipo de fecha limite o con que se entregue antes de los finales es suficiente me gustaria saberlo unicamente con otras asignaturas un saludo y gracias

Figura 106. Detalle del contenido de una celda de la tabla.

Gráficos de Emociones

Al pulsar el botón *Mostrar Gráficos Emociones* se muestra una pantalla como la que se ve en la figura 107.

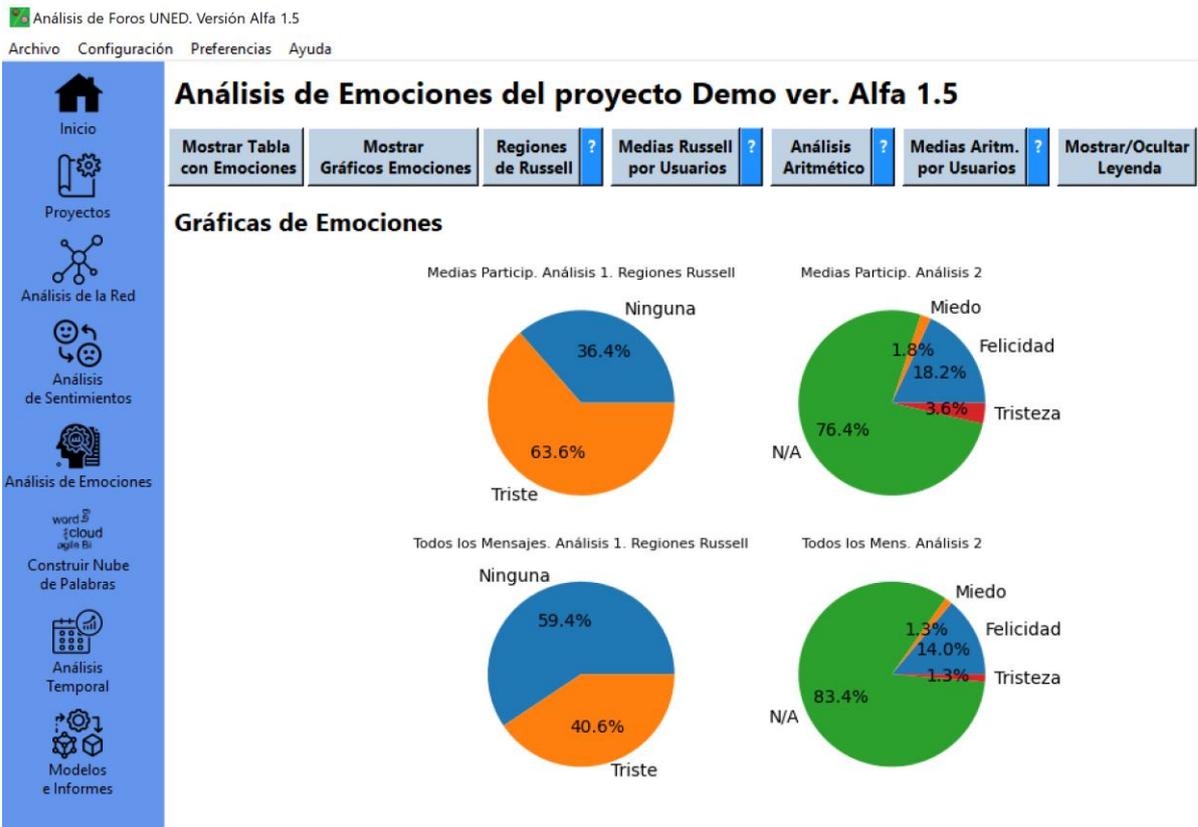


Figura 107. Gráficos de Emociones.

En esta pantalla se muestran cuatro gráficos de tarta:

- Medias de los Participantes según el análisis de Regiones de Russell.
- Medias de los Participantes según el análisis puramente aritmético.
- Proporción de cada región de Russell entre todos los mensajes de los foros.
- Proporción de cada emoción dominante según el análisis 2 entre todos los mensajes de los foros.

Regiones de Russell

Al pulsar el botón *Regiones de Russell* se muestra una pantalla como la que se ve en la figura 108. La región de Russell se halla a partir de un corpus de palabras a las que se le ha asignado un valor de excitación y un valor de valencia. Se busca cada palabra del mensaje en el corpus y se realizan los sumatorios de valencia y excitación para cada mensaje. La región de Russell se halla a partir de dichos valores de valencia y excitación.

El método utilizado en esta aplicación es muy simple y se basa en la función arcotangente. Se ha adoptado de la aplicación mencionada en el [4.1](#). Dicho método se reproduce a continuación.

```
def getRusselRegion(self, arousal, valence):
    """ Function to get the Russell Region from arousal and valence """

    russel = {
        270: C_RELAX,
        180: C_SAD,
        90: C_ANGRY,
        0: C_HAPPY
    }

    a = math.atan2(arousal - 5, valence - 5)
    mydegrees = math.degrees(a)
    deg = mydegrees if mydegrees > 0 else mydegrees + 360
    # print ("Grado:",deg)
    result = ''
    for x in russel.keys():
        if deg > x:
            result = russel[x]
            break
    if valence == 0:
        result = C_NONE
    # print("Resultado:",result)
    return result
```

Análisis de Foros UNED. Versión Alfa 1.5
 Archivo Configuración Preferencias Ayuda

Análisis de Emociones del proyecto Demo ver. Alfa 1.5

Mostrar Tabla con Emociones Mostrar Gráficos Emociones **Regiones de Russell** Medias Russell por Usuarios Análisis Aritmético Medias Aritm. por Usuarios Mostrar/Ocultar Leyenda

Emociones de los mensajes según Regiones de Russell y el Corpus 1.

Icono	Emisor	Alias	Destinatario	Emoción	Mensaje	Título	Foro	Hilo
	37ac418bf95589	u1	Inicia el hilo	Ninguna	buenas tenia er	Calificaciones d	Foro de consult	Calificaciones d
	ecafaa42d80d8e	u2	37ac418bf95589	Ninguna	buenas tardes ¡	Re: Calificacion	Foro de consult	Calificaciones d
	37ac418bf95589	u1	ecafaa42d80d8e	Ninguna	despiste mio m	Re: Calificacion	Foro de consult	Calificaciones d
	eec276fc5d6504	u4	Inicia el hilo	Triste	buenos días he	Consulta exami	Foro de consult	Consulta exami
	ecafaa42d80d8e	u2	eec276fc5d6504	Triste	buenas tardes i	Re: Consulta ex	Foro de consult	Consulta exami
	3ad6f72cd9b52c	u16	Inicia el hilo	Ninguna	hola estoy sigu	Tema 10 - no er	Foro de consult	Tema 10 - no er
	24fb77d6ef839d	u17	3ad6f72cd9b52c	Ninguna	hola jorge en r	Re: Tema 10 - n	Foro de consult	Tema 10 - no er
	a8b41b15be068	u7	3ad6f72cd9b52c	Ninguna	buenas jorge t	Re: Tema 10 - n	Foro de consult	Tema 10 - no er
	3ad6f72cd9b52c	u16	a8b41b15be068	Ninguna	muchisimas gr	Re: Tema 10 - n	Foro de consult	Tema 10 - no er
	24fb77d6ef839d	u17	3ad6f72cd9b52c	Ninguna	me paso lo misi	Re: Tema 10 - n	Foro de consult	Tema 10 - no er

TEXTO DEL MENSAJE:
 buenas tenia entendido que a quienes entregaran las practicas a tiempo recibirian su correccion o como minimo la calific
 sin ver mis practicas 3 y 4 calificadas entendi mal la dinamica o deberia tener las calificaciones saludos xabi

Figura 108. Pantalla de Regiones de Russell de cada mensaje.

En esta tabla se puede ver la emoción calculada mediante el método de Regiones de Russell para cada mensaje. Se muestra en primer lugar un icono representativo de la emoción, a continuación, aparecen el identificador del usuario emisor y el alias que la aplicación le ha asignado de manera automática para anonimizar y acortar el identificador de usuario. Seguidamente aparece el destinatario del mensaje (su identificador) y la emoción calculada. Por último se muestra el contenido del mensaje, el título, el foro en el que se ha publicado y el hilo.

Análisis Puramente Aritmético

Al pulsar el botón *Análisis Aritmético* se muestra una tabla como la de la figura 109. En esta tabla se puede ver para cada mensaje los valores devueltos para cada emoción primaria por un algoritmo puramente aritmético de bolsa de palabras.

Se dispone de un archivo de palabras etiquetadas con una puntuación. Cada palabra tiene una puntuación asociada a cada una de cinco emociones consideradas

primarias (Felicidad, Enojo, Tristeza, Miedo y Asco o Rechazo). Se busca cada palabra del mensaje en el archivo de palabras etiquetadas (corpus) y se suma el valor de cada emoción y palabra. Este método tiene el inconveniente de que si no aparece ninguna palabra relevante del mensaje en el corpus, el mensaje aparece calificado como sin emoción.

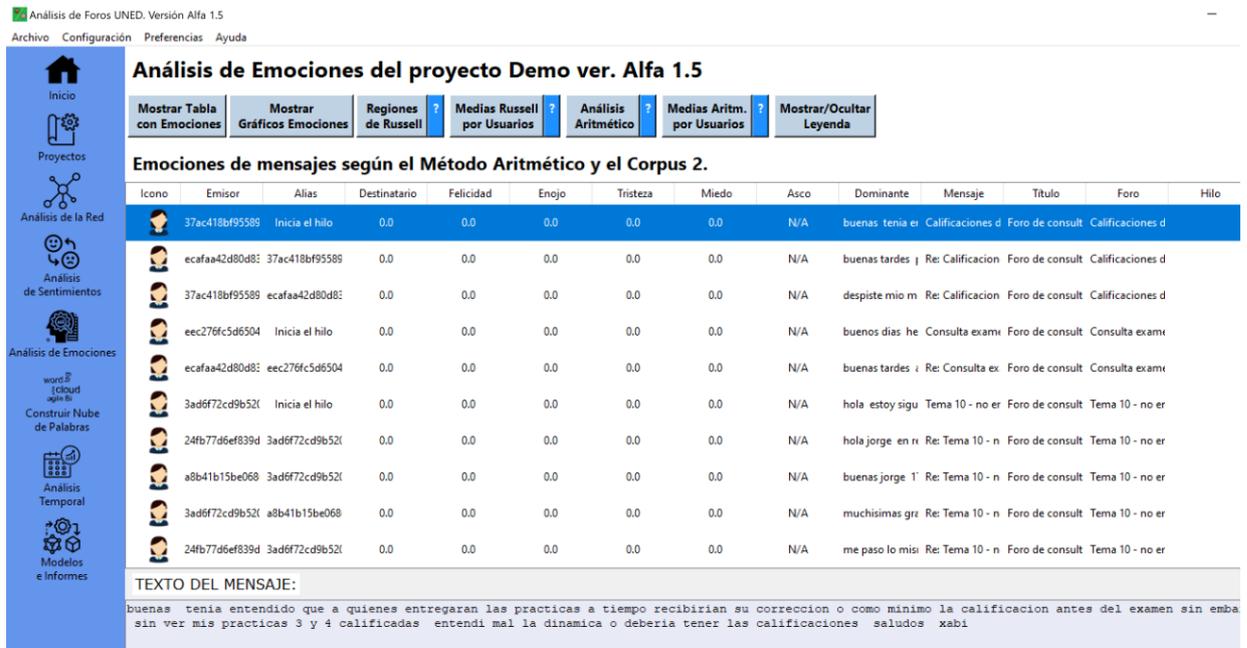


Figura 109. Tabla de análisis puramente aritmético (análisis 2 con corpus 2).

Medias Aritméticas por Usuario

Al pulsar el botón *Medias Aritm. Por Usuarios* se muestra una tabla como la que se ve en la figura 110. En esta tabla se muestra para cada usuario el valor medio de cada componente emocional primaria y se indica como dominante la de mayor valor. También se muestra un icono representativo de la emoción.

Al seleccionar una fila se muestra en la parte inferior de la pantalla la emoción dominante para ese usuario.

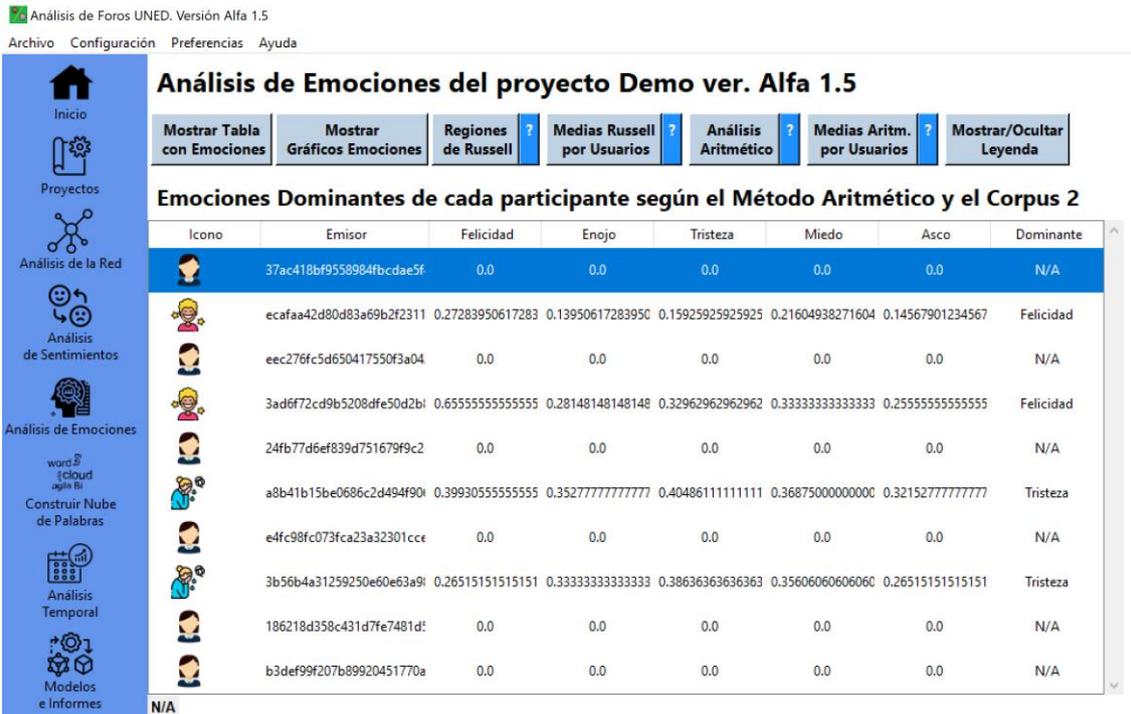


Figura 110. Medias aritméticas de emociones por usuario y emoción dominante por usuario.

Leyenda de los iconos

Mediante el botón *Mostrar/Ocultar Leyenda* se puede mostrar u ocultar la leyenda que nos facilita interpretar los iconos. Funciona con un efecto biestable, de manera que si se pulsa cambia el estado visible u oculto de dicha leyenda. En la figura 111 se puede ver la leyenda de los iconos visible en la tabla de emociones dominantes en cada usuario según el método aritmético (método 2).

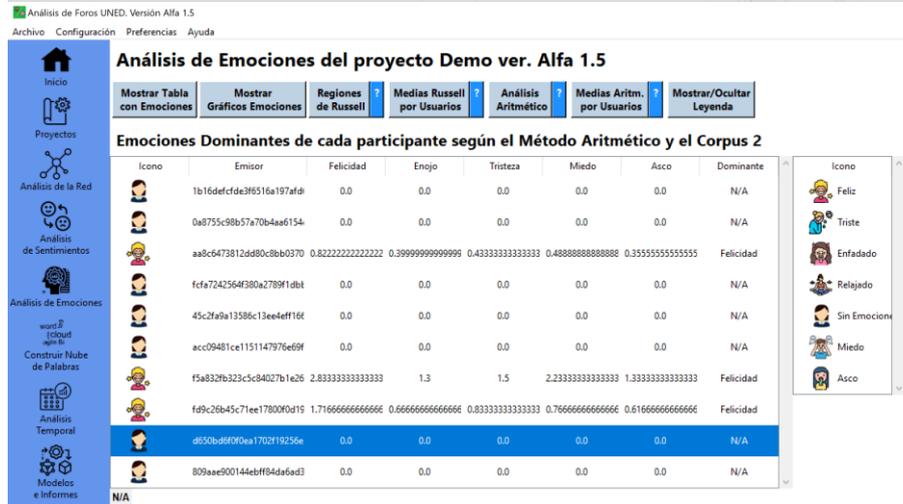


Figura 111. Leyenda de los iconos visible en la tabla de emociones dominantes en cada usuario según el método aritmético (método 2).

A.4 Ayuda del módulo de Análisis Temporal

Gráfica de Mensajes por Fecha

Al abrir el módulo o al pulsar el botón *Nº Mensajes por Fecha*



se muestra un gráfico similar al de la figura 112.

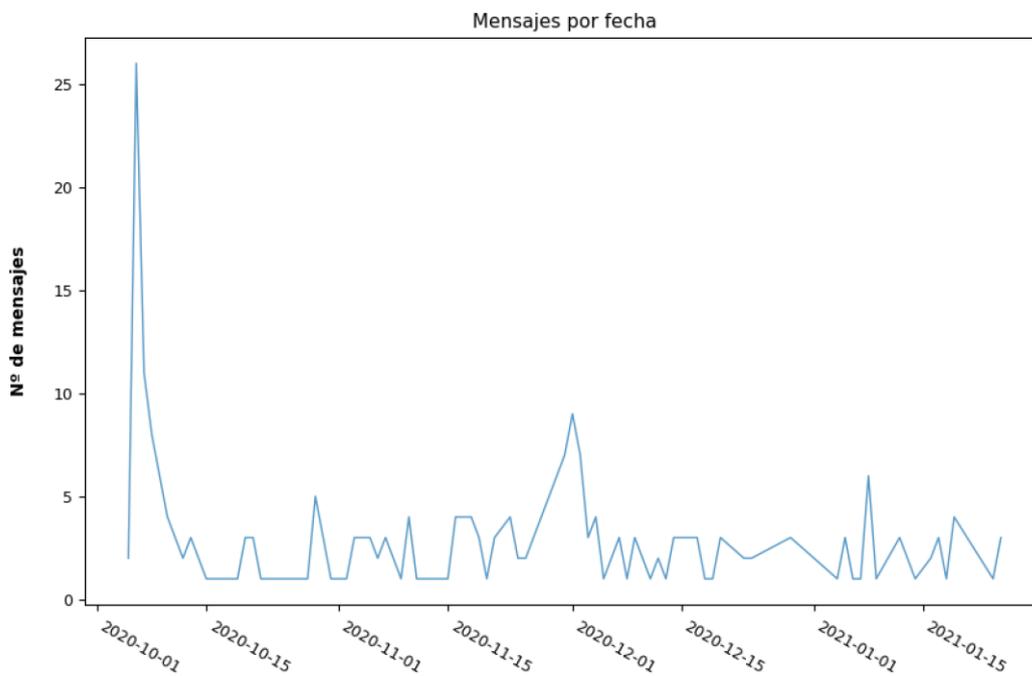


Figura 112. Gráfico de Mensajes por Fecha.

En este gráfico se representa el número de mensajes publicados a lo largo del tiempo.

Visualización de Datos

Pulsando el botón *Visualizar Datos*  se accede a la visualización de una tabla como la que se muestra en la figura 113.

Evolución Temporal del proyecto Demo ver. Alfa 1.5

Visualizar Datos N° Mensajes por Fecha N° Mensajes por día de la semana

	Fechas	numero_m	Día de la sem
1	2020-10-05	2	Lunes
2	2020-10-06	26	Martes
3	2020-10-07	11	Miércoles
4	2020-10-08	8	Jueves
5	2020-10-09	6	Viernes
6	2020-10-10	4	Sábado
7	2020-10-12	2	Lunes
8	2020-10-13	3	Martes
9	2020-10-15	1	Jueves
10	2020-10-16	1	Viernes
11	2020-10-19	1	Lunes
12	2020-10-20	3	Martes
13	2020-10-21	3	Miércoles
14	2020-10-22	1	Jueves
15	2020-10-24	1	Sábado
16	2020-10-26	1	Lunes
17	2020-10-27	1	Martes
18	2020-10-28	1	Miércoles
19	2020-10-29	5	Jueves

78 rows x 3 columns

Figura 113. Datos temporales.

En esta tabla se representa el número de mensajes enviados en cada fecha y el día de la semana de dicha fecha.

Gráfica de Mensajes por Día de la Semana

Pulsando el botón *N° Mensajes por día de la semana*  se accede a la visualización de un gráfico como el de la figura 114.

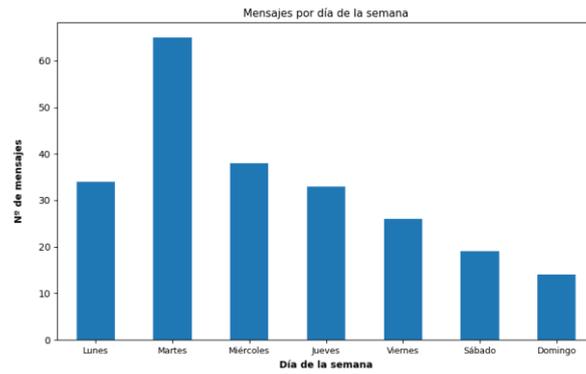


Figura 114. Gráfica de mensajes por día de la semana.

En esta gráfica se puede ver la distribución de mensajes publicados por día de la semana. En la gráfica de ejemplo de la figura 3 se aprecia que los martes se publican más mensajes que en cualquier otro día de la semana, mientras que el día de la semana en el que menos mensajes se publican es el domingo.

A.5 Ayuda del módulo de Modelos e Informes

Visualización de Datos

Al abrir el módulo o al pulsar el botón *Visualizar Datos* se muestra una tabla como la de la figura 115.



Análisis de Foros UNED. Versión Alfa 3.0
 Archivo Configuración Preferencias Ayuda Encuesta

Modelos del proyecto Prueba ; MDEME-Foro-22-11-29.csv
 Mensajes: 122 Participantes: 16

Visualizar Datos Algoritmo de Clustering Gráficos de Usuarios Generar Informe PDF Exportar a Excel Ayuda

Información de cada participante en los foros ---> Desplace la barra inferior para ver todas las columnas. --->

	Participan	Alias Usual	N.Mensaj	Hilos Inic	Resp. A O	A sí mismo	A Sus Hil	A Otros Hi	Grado	Grado Ent	Grado Sal	Centralid	Cercanía	Int
1	79	u1	20	5	13	7	10	10	25	16	9	1.67	1.00	0.0
2	66182	u2	11	2	9	2	4	7	7	3	4	0.47	0.56	0.0
3	64968	u3	6	1	5	1	1	5	5	3	2	0.33	0.56	0.0
4	22946	u4	10	1	9	1	2	8	7	4	3	0.47	0.58	0.0
5	71704	u5	23	3	20	3	5	18	14	5	9	0.93	0.6	0.0
6	39712	u6	7	2	5	2	3	4	6	3	3	0.4	0.56	0.0
7	25964	u7	5	1	4	1	1	4	6	3	3	0.4	0.56	0.0
8	69829	u8	22	2	20	2	6	16	13	5	8	0.87	0.6	0.0
9	25055	u9	5	1	4	1	1	4	5	4	1	0.33	0.58	0.0
10	51194	u10	2	0	2	0	0	2	2	1	1	0.13	0.38	0.0
11	55100	u11	3	0	3	0	0	3	2	1	1	0.13	0.067	0.0
12	58223	u12	2	0	2	0	0	2	2	0	2	0.13	0	0.0
13	60223	u13	1	0	1	0	0	1	1	0	1	0.067	0	0.0
14	66967	u14	3	1	2	1	2	1	4	2	2	0.27	0.42	0.0
15	23709	u15	1	0	1	0	0	1	2	1	1	0.13	0.38	0.0
16	23778	u16	1	0	1	0	0	1	1	0	1	0.067	0	0.0

16 rows x 32 columns

Figura 115. Tabla de datos de los participantes de los foros.

En esta tabla se pueden ver las diferentes variables que se evalúan, calculan y visualizan en el resto de módulos. Cada fila corresponde a un participante. La primera columna muestra el identificador de usuario. La segunda columna muestra el alias asignado automáticamente por la aplicación a cada usuario. A continuación aparece la siguiente información:

Columnas de la 3ª a la 8ª

Se muestra información relativa al número de mensajes publicados.

- *N.Mensajes* (Número total de mensajes)
- *Hilos iniciados* (Número de hilos iniciados)
- *Resp. A Otros Autores* (Número de mensajes enviados como respuesta a mensajes de otros autores)
- *A sí mismos + inicios* (Suma del número de mensajes publicados en respuesta a sus propios mensajes más mensajes iniciando hilos)

- *A sus Hilos* (Número de mensajes publicados en hilos iniciados por el mismo usuario)
- *A Otros Hilos* (Número de mensajes publicados en hilos no iniciados por el propio usuario)

Columnas de la 9ª a la 18ª

Se muestra información relativa a medidas de centralidad y otras estadísticas de la red. Ver la ayuda del módulo de análisis de Red.

- *Grado*
- *Grado Ent.*
- *Grado Sal.*
- *Centralidad Grado*
- *Cercanía*
- *Intermediación*
- *Cent. Vect. Propio*
- *Page Rank*
- *Iniciativa*
- *Modularidad*

Columnas de la 19ª a la 23ª

Se muestra información relativa a al análisis de sentimientos. Ver la ayuda del módulo de análisis de sentimientos.

- *Sent. Analysis*
- *Sentimiento N.B*
- *Py Sentimiento*
- *Sent. Comb*
- *Sent. Comb Lit.*

Columnas de la 24ª a la 32ª

Se muestra información relativa al análisis de emociones. Ver ayuda del módulo de análisis de sentimientos.

- *Valencia Media*
- *Excit. Media*

- *Región Russell*
- *Felicidad*
- *Enfado*
- *Tristeza*
- *Miedo*
- *Disgusto*
- *Dominante*

Agrupamiento (*Clustering*)

Pulsando el botón *Algoritmo de Clustering*  se accede a la visualización de un gráfico similar al de la figura 116. Puede rotarse la vista con el ratón.

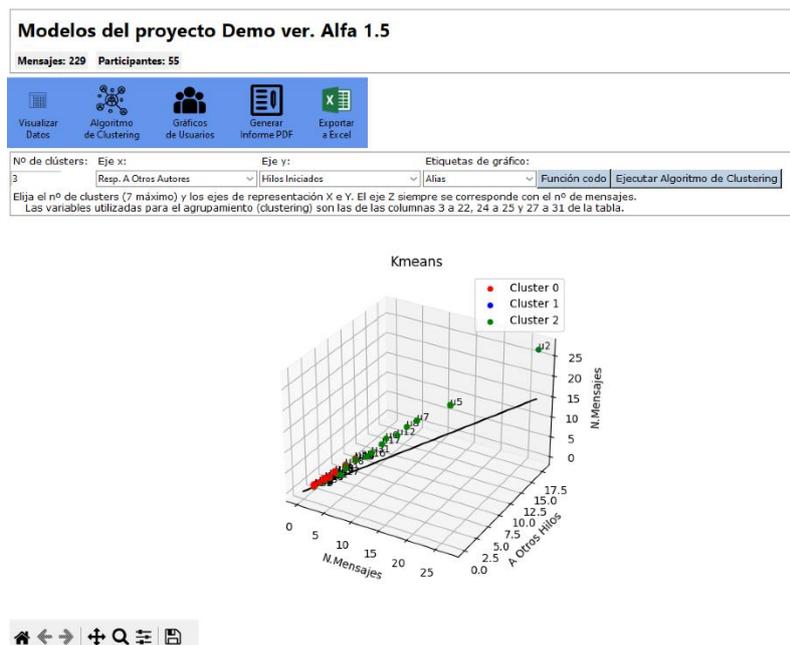


Figura 116. Gráfico de agrupamiento de usuarios (clustering).

Al acceder a la función de agrupamiento se genera un gráfico de agrupamiento inicial. En esta ejecución inicial del algoritmo de *clustering* se agrupan los usuarios como

pertenecientes a tres categorías o grupos. El algoritmo utilizado es K-medias (*Kmeans*)⁴⁰.

Se trata de una representación espacial en la que el eje X se corresponde con el número de mensajes publicados como respuestas a otros autores; el eje Y se corresponde con el número de hilos iniciados y el eje Z con el número de mensajes.

Las variables utilizadas para ejecutar el algoritmo de *clustering* son las columnas 3 a 22, 24 a 25 y 27 a 31 de la tabla.

Configuración del agrupamiento

En la parte superior, bajo los botones del módulo se encuentran unos campos de configuración. Mediante esta configuración se puede elegir:

- Número grupos (*clusters*). Pueden ser 7 como máximo.
- Eje X de la representación espacial.
- Eje Y de la representación espacial.
- Etiquetas a mostrar junto a cada punto (usuarios, alias generados automáticamente por la aplicación o ninguna etiqueta).

Una vez configurado el algoritmo del agrupamiento puede ejecutarse y dibujar el gráfico pulsando el botón *Ejecutar Algoritmo de Clustering*.

Función codo

Dentro del apartado de *clustering*, si se pulsa el botón *Función codo* se muestra un gráfico como el de la figura 117. El número de clusters idóneo según este método coincide con la zona que se asemeja a un codo. En este caso estaría entre 2 y 3.

⁴⁰ https://scikit-learn.org/stable/modules/generated/sklearn.cluster.k_means.html

Nº de clústers:	Eje x:	Eje y:	Etiquetas de gráfico:	
3	Resp. A Otros Autores	Hilos Iniciados	Alias	Función codo Ejecutar Algoritmo de Clustering

Elija el nº de clusters (7 máximo) y los ejes de representación X e Y. El eje Z siempre se corresponde con el nº de mensajes.
Las variables utilizadas para el agrupamiento (clustering) son las de las columnas 3 a 22, 24 a 25 y 27 a 31 de la tabla.

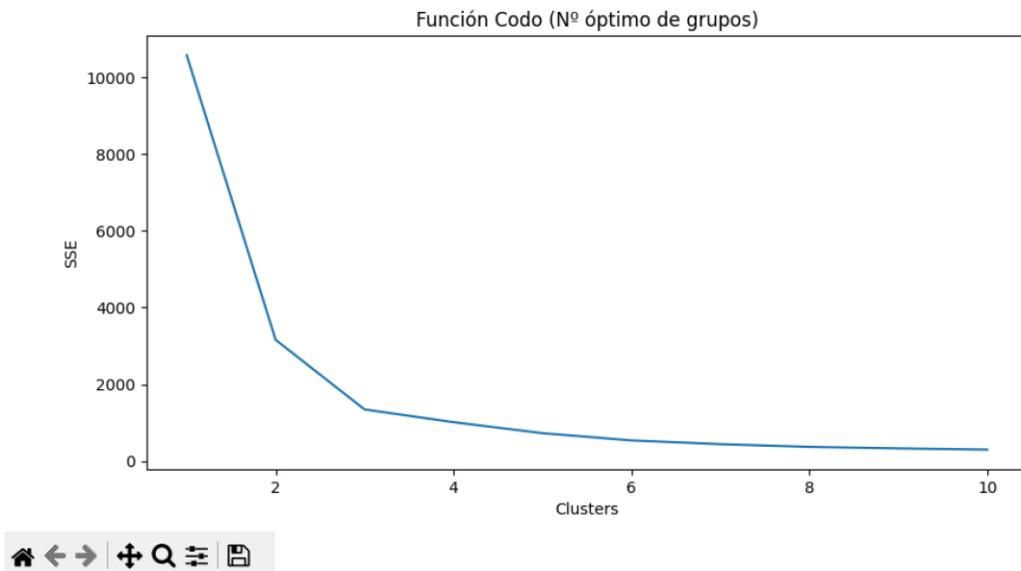


Figura 117. Función codo que facilita la elección del número de grupos (clusters).

Gráficos de Usuarios

Al pulsar el botón *Gráficos de Usuarios*  se accede a una pantalla como la que se ve en la figura 118.

Modelos del proyecto Prueba ver. Alfa 1.6 archivoForo.csv

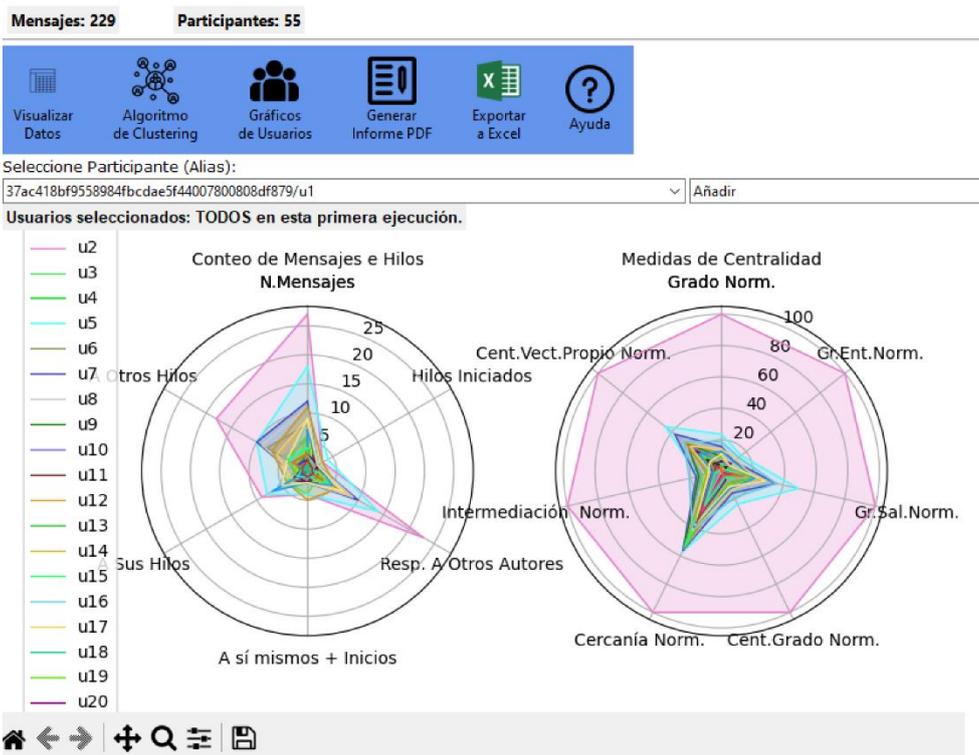


Figura 118. Selección de usuario para dibujar su gráfico.

Lo que se muestra son dos gráficos de tipo radar. En el de la izquierda se representan variables relacionadas con el número de mensajes e hilos iniciados, mientras que en el de la derecha se representan las medidas de centralidad. [Ver el documento de Ayuda del módulo de Red](#). Los valores de este segundo gráfico se encuentran normalizados mediante el siguiente código:

```
# Nueva escala de 0 a 100.
new_max = 100
new_min = 0
new_range = new_max - new_min

# Hacer transformación lineal en cada variable para cambiar
el valor
# to [0, 100].
for factor in factors:
    max_val = df[factor].max()
    min_val = df[factor].min()
    val_range = max_val - min_val
    df[factor + ' Norm'] = df[factor].apply(
        lambda x: ((x - min_val) * new_range) / val_range
    + new_min)
```

Donde *factors* es un array que contiene todas las variables y *factor* representa a cada una de esas variables (medidas de centralidad).

En la primera visualización se representan todos los usuarios. Mediante la rueda del ratón podemos hacer *scroll* para desplazarnos por la leyenda cuando no se puede ver completa por contener muchos usuarios.

Mediante el desplegable de la izquierda podemos seleccionar cualquier usuario para añadirlo o eliminarlo en las gráficas de tipo radar. Una vez se seleccione un usuario, podemos seleccionar en el desplegable de la derecha una de las opciones (*Añadir* o *Quitar*). Al seleccionar cualquiera de estas opciones se dibujarán automáticamente los dos gráficos de tipo radar. La primera vez debe realizarse la adición de cualquier usuario para borrar los gráficos iniciales de todos los usuarios y dibujar los gráficos de ese primer usuario. A partir de ese momento podemos ir añadiendo o quitando usuarios y los gráficos se irán actualizando automáticamente. Ver figura 119.

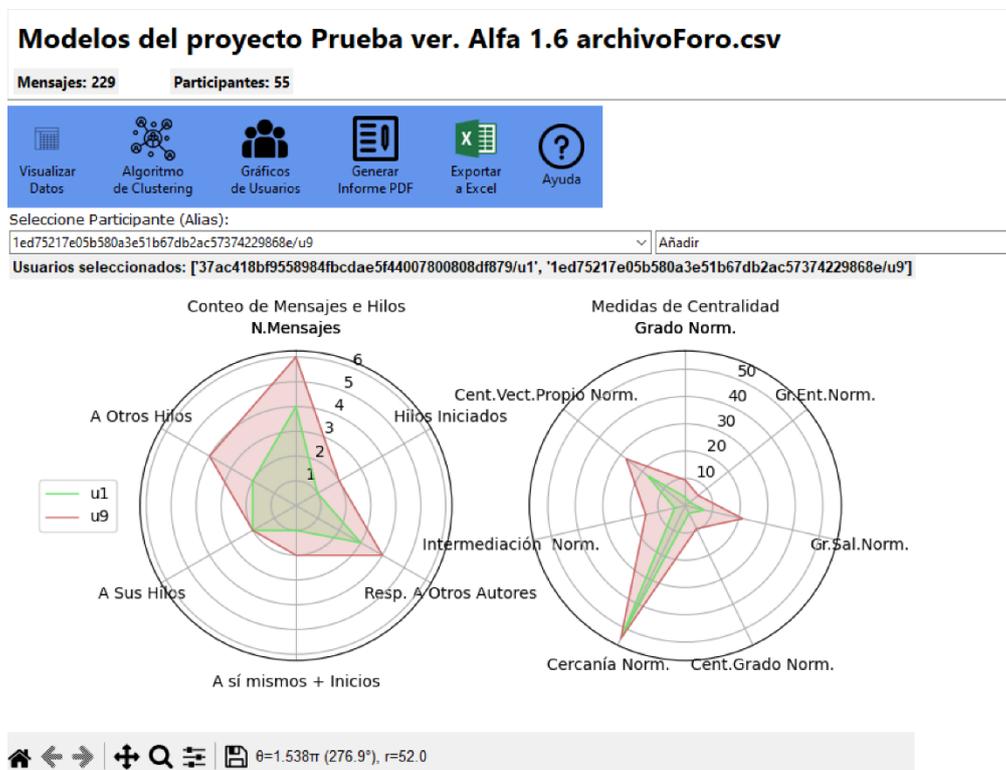


Figura 119. Gráficos de radar de dos usuarios.

En la versión 3.0 Beta de la aplicación se representan:

VARIABLES ASOCIADAS A LAS ESTADÍSTICAS DE NÚMERO DE MENSAJES

- N. Mensajes
- Hilos Iniciados
- Resp. A Otros Autores
- A sí mismos + Inicios
- A Sus Hilos
- A Otros Hilos

VARIABLES DE LAS MEDIDAS DE CENTRALIDAD (NORMALIZADAS)

- Grado
- Grado de Entrada
- Grado de Salida
- Centralidad de Grado
- Cercanía
- Intermediación
- Centralidad de Vector Propio

INFORMES EN PDF

Haciendo clic en el botón *Generar Informe PDF*  se generará un informe en PDF de manera automática y se abrirá. Si ya existe un informe para el mismo proyecto se sobrescribirá siempre que no esté abierto. El nombre del autor del informe se puede configurar en el archivo config.ini. En el informe se añaden todos los gráficos que se han generado en la sesión de trabajo. Si se desea omitir alguna imagen de gráfico debe borrarse manualmente de la carpeta correspondiente. Las carpetas que contienen las imágenes para el informe son:

`./Proyectos/Nombre_Proyecto/Gráficos_Clustering/`

`./Proyectos/Nombre_Proyecto/Gráficos_Emociones/`

`./Proyectos/Nombre_Proyecto/Gráficos_Nube/`

`./Proyectos/Nombre_Proyecto/Gráficos_Red/`

`./Proyectos/Nombre_Proyecto/Gráficos_Sentimientos/`

`./Proyectos/Nombre_Proyecto/Gráficos_Temporales/`

`./Proyectos/Nombre_Proyecto/Gráficos_Usuarios/`

El informe en PDF se guarda en la ruta:

`./Proyectos/Nombre_Proyecto/Infor. mes/PDF/`

Adicionalmente la aplicación también guarda dos pequeños informes en formato html:

`./Proyectos/Nombre_Proyecto/Infor. mes/informe_1.html` es un informe que contiene información de los usuarios, sus alias, estadísticas de mensajes y estadísticas de la red.

En la figura 120 se muestra uno de estos informes.

Participante	Alias Usuario	N Mensajes	Hilos Iniciados	(Resp. A Otros Autores)	A ni mímos + Inicio	A San Hilos	A Otros Hilos
37ac415b9f518948f8e4a784407800808879	al1	4	1	3	11	2	1
ee42ca2c085831902c21119105051709a10	al2	37	3	33	14	9	18
41c2929a1556e11e04af166942588a744386	al3	6	2	2	4	4	2
eecc76658f04175502a0421a267188826f	al4	2	1	1	11	1	1
0aa0206d5f5a397ae1ae0a676a156168d0	al5	18	4	14	10	10	10
4e839620738a2423210e02f4451f68a87	al6	6	2	4	10	2	3
ab841319a56862494941a3104499104	al7	12	6	10	12	10	10
9a719a06606c72919166411711218	al8	11	4	8	11	8	8
1ae1921705b38a241867b2a25734229686	al9	6	2	4	12	2	4
0c95e4e8370bbba448ba82178c2a687e780	al10	4	2	2	11	2	4
02939e421ca801382a8f8230624a34a494	al11	11	2	2	12	3	4
19598aa1129252a0e0a1984156a5446e0b	al12	11	3	6	15	4	4
4a5424da178b56506692a76d1839a6eab	al13	3	1	2	11	1	1
8da1245428824289188878ba36676c0	al14	2	0	2	10	0	2
2a7012420827a812a4e0b3231a30935186	al15	3	0	3	10	0	3
3a8f92a09510388a5043834648132a7445	al16	9	2	7	12	6	3
24b770eaf39d7116799c2b497de7ba2d7c8	al17	9	1	8	11	3	6
4e0805192a0a0a1809a4848478a78e70	al18	3	0	3	10	0	3
aa349479812880c80a03701558451978a30359	al19	4	1	3	11	2	4
80a07ae720422676129a4e4a885015a6d21af	al20	3	2	1	12	2	1
803a9a0114a68844a6a644444444444444	al21	2	1	1	11	2	0
7486f7c04405937154243a9a0a0e0e0e0e	al22	6	0	6	10	0	6
699c26a43c1ae18006a19284a8326a83758	al23	2	0	2	10	0	2
188118d3541478e481a211942f96a8a2b	al24	2	0	2	10	0	2
03a4992073899024170ae1a9e7a3031979	al25	1	0	1	10	0	1
7a1686a1884a3a435a2c0c0d8941a2a00b	al26	1	0	1	10	0	1
97a813a232c5c8a027a2a068189429a50f	al27	1	0	1	10	0	1
238a71aedd32a0ba3009811756a8c506e	al28	1	0	1	10	0	1
169a254295121248a6621a7012663c03327	al29	1	0	1	10	0	1
0646078a6a754d78787441870a19a754a83330	al30	3	0	3	10	0	3
73114a4a91996149f1c21050e02956d89b	al31	8	0	8	10	0	8
53116247584da1a166a1786a740247a69	al32	2	0	2	10	0	2
c0848a1971a7a8838a8a8a8a8a8a8a8a8a8	al33	1	0	1	10	0	1
44a47a4e495a66a2b7e71233a2a148149a03	al34	3	1	2	11	2	1
1896eac78a9f7515818a1a4434a49a0b	al35	1	0	1	10	0	1
9312665508a1e0a0a0a10a185ea558764	al36	2	0	2	10	0	2
6c2929a726a9a29701c1e1c154278234	al37	1	0	1	10	0	1
0c11184a4111a4e0519a9a19a19a19a19a	al38	1	0	1	10	0	1
873308a831166217a4186d6786cc31a6cb	al39	3	0	3	10	0	3
1a2891a3030a87409040a4b448a787181	al40	2	0	2	10	0	2
039a71a31a1a88781a8a1a8a1a8a1a8a1a8a	al41	1	0	1	10	0	1
408e36133c702a1a2a1a0379154a49a2d6	al42	1	0	1	10	0	1
04849781899a134020a0a07a7a813a2a2a	al43	2	0	2	10	0	2
841ae7a5f6742a7028798a7105c67e1	al44	1	0	1	10	0	1
39a67788ca017a0c684a4a4a4a4a4a4a4a	al45	1	0	1	10	0	1
2af728561a0c4181c0a0a6aa3a0728036a	al46	1	0	1	10	0	1
030a0790a0a4991e1a0916a0a4a0a0a2a2c7	al47	1	0	1	10	0	1
98a57a2c6a899101a0a1524a0a0a0a0a0a0	al48	4	0	4	10	0	4
338a0a0a078c0c02a0389540a188aa7	al49	1	0	1	10	0	1
80c39a4e0a8888a9a037e715a835106a2f	al50	1	0	1	10	0	1
1ae11a08137120a0a1c7e0a4c1819a031aa	al51	1	0	1	10	0	1
1516a4c8a18816a197a684a4a3591a977	al52	1	0	1	10	0	1
0af735898a7a70a4a815a740a558a92c080	al53	3	0	3	10	0	3
aa09a41ae1111197a499a020a9a7813a0d0	al54	1	0	1	10	0	1
0530a0a0a170219255a850a254a0a0e	al55	2	1	1	11	1	1

Participante	Alias Usuario	Grado (Grado Em)	Grado Sal	Centralidad Grado	Cercanía	Intermedios	Cent. Vect.	Propio	Page Rank	Iniciativa	Modularidad
37ac415b9f518948f8e4a784407800808879	al1	3	1	2	0.05556	0.481143	0.027946	1.188044+01	0.018771	1	2
ee42ca2c085831902c21119105051709a10	al2	65	15	15	0.203704	0.912544	0.452110	0.704671+01	0.031311	1	2
41c2929a1556e11e04af166942588a744386	al3	6	4	2	0.111111	0.495413	0.029607	1.899841+01	0.047867	2	2
eecc76658f04175502a0421a267188826f	al4	2	1	1	0.037037	0.481143	0.027946	1.188044+01	0.018771	1	1
0aa0206d5f5a397ae1ae0a676a156168d0	al5	18	8	8	0.296296	0.524723	0.077492	1.020800+01	0.059990	4	4
4e839620738a2423210e02f4451f68a87	al6	6	2	4	0.111111	0.484485	0.079112	1.548752+01	0.024254	0	3
ab841319a56862494941a3104499104	al7	11	5	6	0.203704	0.514286	0.081399	2.233511+01	0.049662	2	3
9a719a06606c72919166411711218	al8	9	4	5	0.166667	0.500000	0.059501	2.002154+01	0.024282	2	4
1ae1921705b38a241867b2a25734229686	al9	7	4	4	0.125830	0.499269	0.021163	1.989756+01	0.023092	2	4
0c95e4e8370bbba448ba82178c2a687e780	al10	4	2	2	0.074074	0.499269	0.027946	1.728170+01	0.025145	1	3
02939e421ca801382a8f8230624a34a494	al11	3	2	1	0.055556	0.499269	0.027946	1.988617+01	0.027754	2	6
19598aa1129252a0e0a1984156a5446e0b	al12	11	3	2	0.148148	0.499269	0.021163	1.989756+01	0.023092	2	4
4a5424da178b56506692a76d1839a6eab	al13	2	1	1	0.037037	0.481143	0.027946	1.188044+01	0.018771	1	1
8da1245428824289188878ba36676c0	al14	2	1	1	0.037037	0.481143	0.027946	1.188044+01	0.018771	1	1
2a7012420827a812a4e0b3231a30935186	al15	3	0	3	0.203704	0.912544	0.452110	0.704671+01	0.031311	1	2
3a8f92a09510388a5043834648132a7445	al16	13	7	6	0.240741	0.387114	0.103541	2.055472+01	0.038706	2	3
24b770eaf39d7116799c2b497de7ba2d7c8	al17	10	6	4	0.185185	0.382979	0.038488	1.704870+01	0.023980	1	3
4e0805192a0a0a1809a4848478a78e70	al18	3	1	2	0.055556	0.282723	0.028550	1.606713+02	0.008210	0	4
aa349479812880c80a03701558451978a30359	al19	5	4	4	0.125830	0.285114	0.044751	3.848313+02	0.011939	1	5
80a07ae720422676129a4e4a885015a6d21af	al20	3	2	1	0.092593	0.362416	0.037710	4.495066+02	0.013784	0	0
803a9a0114a68844a6a644444444444444	al21	2	1	1	0.055556	0.348154	0.028114	8.114927+02	0.015069	1	3
7486f7c04405937154243a9a0a0e0e0e0e	al22	6	3	3	0.111111	0.350649	0.030234	2.966583+02	0.017888	0	4
699c26a43c1ae18006a19284a8326a83758	al23	3	1	2	0.055556	0.331288	0.027946	2.444434+02	0.007882	0	3
188118d3541478e481a211942f96a8a2b	al24	1	0	1	0.018119	0.000000	0.010101	4.090607+13	0.002727	0	1
03a4992073899024170ae1a9e7a3031979	al25	1	0	1	0.018119	0.000000	0.010101	4.090607+13	0.002727	0	1
7a1686a1884a3a435a2c0c0d8941a2a00b	al26	1	0	1	0.018119	0.000000	0.010101	4.090607+13	0.002727	0	1
97a813a232c5c8a027a2a068189429a50f	al27	1	0	1	0.166667	0.367347	0.034492	1.177883+01	0.026579	2	4
238a71aedd32a0ba3009811756a8c506e	al28	1	0	1	0.018119	0.000000	0.010101	4.090607+13	0.002727	0	1
169a254295121248a6621a7012663c03327	al29	1	0	1	0.018119	0.000000	0.010101	4.090607+13	0.002727	0	1
0646078a6a754d78787441870a19a754a83330	al30	2	1	1	0.037037	0.333404	0.027946	3.264924+02	0.008763	0	4
73114a4a91996149f1c21050e02956d89b	al31	8	0	8	0.148148	0.348189	0.041942	6.408253+02	0.019189	4	4
53116247584da1a166a1786a740247a69	al32	3	1	2	0.055556	0.337500	0.027946	3.548292+02	0.008515	0	4
c0848a1971a7a8838a8a8a8a8a8a8a8a8a8	al33	1	0	1	0.018119	0.000000	0.010101	4.090607+13	0.002727	0	1
44a47a4e495a66a2b7e71233a2a148149a03	al34	3	1	2	0.055556	0.223140	0.027946	8.818244+03	0.005244	1	1

Figura 120. Visualización del archivo informe_1.html.

./Proyectos/Nombre_Proyecto/Infor. mes/informe_modelo.html

es un informe que contiene solamente los participantes e información numérica sobre los mensajes. En la figura 121 se puede ver un informe de este tipo.

Participante	N.Mensajes	Hilos Iniciados	Resp. A Otros Autores	A si mismos + Inicios	A Sus Hilos	A Otros Hilos
37ac418b9558948fcdac544007800804df879	4	1	3	1	2	2
ecafaa42480483a69b2c2311930c624550be1f0	27	3	23	4	9	18
45c2f69a13586c13ee4e1f1669425d88a7a443fc	6	2	2	4	4	2
fec2766c4650417550f3a042c1a0fe718f8626f	2	1	1	1	1	1
0aad920a63c8e397ae53ec34c0ae076a1561dbd0	18	4	14	4	8	10
e4c98e6073fa23a23301cecb274a355f8bbd7	8	0	8	0	0	8
8bb41b15be0686c249490641de31047499104	12	2	10	2	2	10
bc273ceab66f80e7f29336b16ca411731a218	11	2	9	2	2	9
1ed75217e05b580a3e51b67db2ac5734239868a	6	2	4	2	2	4
0e95e4a48870ebba5408ae34327db2e6867e780	6	1	5	1	2	4
7e95bete421ca601382ba8b33065af554a4c94	4	2	2	2	3	1
3b56b4a31259250e60e63a98a433e65b344d506b	11	3	6	5	4	7
f4b542b6e437b565b065692e237b61839e66cab	3	1	2	1	1	2
fcfa7242564f380a2789f1dbb87b8e306fe76c00	2	0	2	0	0	2
a7a27d220b27a8472e1e9c8e251a31a9935538f	3	0	3	0	0	3
3ad6f72ced9652084e50d2b863465f6132e7645	9	2	7	2	6	3
24b77d6e839d751679f9c2db497de78a3cd7e8	9	1	8	1	3	6
e4f8f0f3390abd9ab1f609e684833bb37e70	3	0	3	0	0	3
aa8c647812d480c8bb037015585c51978a30355	6	1	5	1	2	4
fc0e05ee72042b76129dc42ed8561b0e6d1d5	3	2	1	2	2	1
809aae900144ebf84da6ad36445a9dd3bb4072a	2	1	1	1	2	0
7486f5cb94460937b718542c8e940a40ec45bde	5	0	5	0	0	5
f49c26b43c71ee17800f0d1928458326a83758e	2	0	2	0	0	2
1862184358c431d7fe7481d511042f5fb6edba2b	2	0	2	0	0	2
b3de99207b89920451770ad7e957a3b5303379	2	0	2	0	0	2
7e16fd4e31f4d43ba453b2c20cced6941e2a60fb	1	0	1	0	0	1
f5a832fb232c5c84027b1e260661d394229e65e6	7	2	5	2	7	0
28be17e1cd3b2aebba3a9069415415b685ad56e	2	0	2	0	0	2
1f90e25629523248cba6621e730126fc3c05327	1	0	1	0	0	1
0646079add6754d787db7441870e19e3549c833	3	0	3	0	0	3
73114c6ea4639b6149f70c19f5c28259edd698b	9	2	7	2	5	4
53716f5d7564da1aa1f66a41786aa740d247e69	2	0	2	0	0	2
cd648ad19775a7d8b89ab4ecbe80b10fd7b34e6	1	0	1	0	0	1
4da45a54500ebd2b7e721285ac417d421a9d53	3	1	2	1	2	1
1899e64c739a967935f88684e43b43d8e99405b	1	0	1	0	0	1
9312665508f1e5c00da8e1c70a1f35ce558f764	2	0	2	0	0	2
dc29203e726ae9a0b67074c21c1c35d50786234	1	0	1	0	0	1
3e01319490457214c4668539693a1428c1f52eb	2	0	2	0	0	2
8723b9bb88511662174e16b6d7b86cc313e8cb	3	0	3	0	0	3
1d2d912a30306ac87409504b4b44f8e767d181	2	0	2	0	0	2
039d7061d552c184df6763d0a461d1bf8ba8e99	2	0	2	0	0	2
d36c5f1b33c702d3ec2f1e80379354b49a3fa6	1	0	1	0	0	1
d484978b18499a01340200df6014270e8b36ba04	1	0	1	0	0	1
841ae72c8f7a036743ef02697f5d1205a5e7c1	2	0	2	0	0	2
39b7786bc3a019e2c6d64c0dbef50d70ff480c	1	0	1	0	0	1
2d7285663dccc41821ce60ab66eae58407280b36	1	0	1	0	0	1
208b3f090ba349b7e1d0916dc8a940b6a2a2c7	1	0	1	0	0	1
9b2526c2b3e9897101cbd15924f6bfa0a06995	4	0	4	0	0	4
33b8chadae67ffcc02c64438f9b546ea18f4ae7	1	0	1	0	0	1
8b5e904cef68868aa9037c721584c85506a26ff	1	0	1	0	0	1
1cd12d8c337428f94c07cf56c5835ba8a52faa	1	0	1	0	0	1
1b16defcfd3b6516a197af6f6dae1bb5b391e977	1	0	1	0	0	1
0a8755c98b5a70b4aa6154c740e556e9f52c060	3	0	3	0	0	3
acc09481ce1151147976e69f4209f4a7b8184d0	1	0	1	0	0	1
a650bd6f0dea1702f19256ed80e26426adff9e6	2	1	1	1	1	1

Figura 121. Visualización del archivo informe_modelo.html

Exportación a Excel

Pulsando el botón *Exportar a Excel*  se crea un libro de Excel con todos los datos de la tabla de usuarios y a continuación se abre. El archivo se guarda en la ruta:

./Proyectos/Nombre_Proyecto/Infor_mes/Excel/

El aspecto de este archivo se puede ver en la figura 121.

Apéndice A. Ayuda de la aplicación

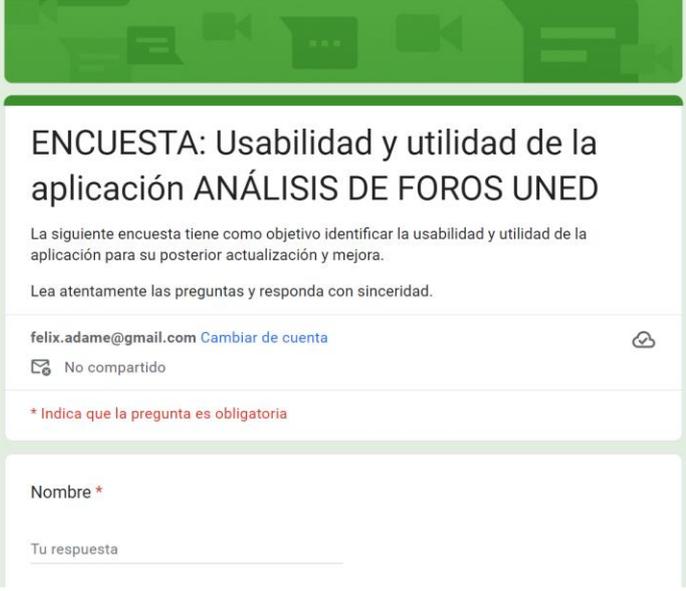
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T			
1	Participantías	Usuarios	Mensajes	Iniciadas	Otros	Achismos + Ir	A Sus Hilos	Otros Hilos	Grado	Grado Ent.	Grado Sal.	Irregularidad	Cercanía	Intermediariedad	Vect.	Prd	Page Rank	Iniciativa	Modularidad	Identificación	Analisis	Intimidad
2	0	37ac418bf1u1		4	1	3	1	2	2	3	1	2	0,055556	0,482143	0,027946	0,118808	0,018771	1	2	0,009749		
3	1	ecafa42dfu2		27	3	23	4	9	18	65	50	15	1,203704	0,915254	0,45821	0,670467	0,28313	3	1	2,02E-05	0,60	
4	2	45c2fa91u3		6	2	2	4	4	2	6	4	2	0,111111	0,495413	0,029607	0,169984	0,047867	2	2	1,2E-05		
5	3	ee276f5cu4		2	1	1	1	1	1	2	1	1	0,037037	0,482143	0,027946	0,118808	0,018771	1	1	0,000106		
6	4	0aad920a6u5		18	4	14	4	8	10	16	8	8	0,296296	0,524272	0,077492	0,305088	0,05999	4	3	0,000397	0,60	
7	5	e4fc98f07u6		8	0	8	0	0	8	6	2	4	0,111111	0,486486	0,076712	0,154876	0,024254	0	3	0,000276	0	
8	6	a8b41b15bu7		12	2	10	2	2	10	11	5	6	0,203704	0,514286	0,083395	0,253351	0,049662	2	3	0,008744	0	
9	7	be273ceab u8		11	2	9	2	2	9	9	4	5	0,166667	0,5	0,059501	0,200126	0,034282	2	4	0,006717	0,63	
10	8	1ed75217eu9		6	2	4	2	2	4	7	3	4	0,12963	0,495413	0,076184	0,184227	0,038403	2	4	0,055885	0,64	
11	9	0c95e4a48 u10		6	1	5	1	2	4	4	2	2	0,074074	0,490909	0,027946	0,172871	0,025145	1	3	3,15E-05	0,6	
12	10	7e95be4c4 u11		4	2	2	2	3	1	3	2	1	0,055556	0,490909	0,027946	0,139862	0,027734	2	6	2,52E-15	0	
13	11	3b56b4a31 u12		11	3	6	5	4	7	8	3	5	0,148148	0,490909	0,051263	0,198958	0,031092	3	3	1,43E-05	0,51	
14	12	f4b542bfe4 u13		3	1	2	1	1	2	2	1	1	0,037037	0,482143	0,027946	0,118808	0,018771	1	1	2,36E-05		
15	13	fcfa72d25e u14		2	0	2	0	0	2	2	1	1	0,037037	0,482143	0,027946	0,118808	0,018771	0	1	1,36E-06		
16	14	a7a274222 u15		3	0	3	0	0	3	4	2	2	0,074074	0,490909	0,03771	0,118808	0,021089	0	6	0,047554		
17	15	3ad6f72cdf u16		9	2	7	2	6	3	13	7	6	0,240741	0,385714	0,103541	0,203547	0,038706	2	3	0,002443	0,60	
18	16	24fb77d6e u17		9	1	8	1	3	6	10	6	4	0,185185	0,382979	0,038468	0,170487	0,02938	1	3	0,000348	0,67	
19	17	ea4ffb07331 u18		3	0	3	0	0	3	3	1	2	0,055556	0,282723	0,02835	0,036068	0,00821	0	4	5,23E-08		
20	18	aa8c60738 u19		6	1	5	1	2	4	7	3	4	0,12963	0,285714	0,064731	0,038483	0,011939	1	5	7,51E-07	0,64	
21	19	fc0e05ee7 u20		3	2	1	2	2	1	5	4	1	0,092593	0,362416	0,03771	0,046906	0,015784	2	0	6,51E-07	0,31	
22	20	809aae900 u21		2	1	1	1	2	0	3	2	1	0,055556	0,346154	0,028114	0,080149	0,015049	1	3	6,22E-08		
23	21	7486f5cb9 u22		5	0	5	0	0	5	6	3	3	0,111111	0,350649	0,030236	0,059664	0,016788	0	4	0,097718		
24	22	fd9c26b45u23		2	0	2	0	0	2	3	1	2	0,055556	0,331288	0,027946	0,027444	0,007882	0	3	0,000114		
25	23	186218d35u24		2	0	2	0	0	2	1	0	1	0,018519	0	0,010101	4,09E-13	0,002727	0	1	0,000119		
26	24	b3def992cu25		2	0	2	0	0	2	1	0	1	0,018519	0	0,010101	4,09E-13	0,002727	0	1	1,25E-11		
27	25	7e16fd5e3 u26		1	0	1	0	0	1	1	0	1	0,018519	0	0,010101	4,09E-13	0,002727	0	1	4,65E-10		
28	26	f5a832fb3u27		7	2	5	2	7	0	9	5	4	0,166667	0,367347	0,035432	0,117738	0,026579	2	4	0,017664	0,47	

Figura 122. Archivo de Excel generado mediante la exportación.

Apéndice B

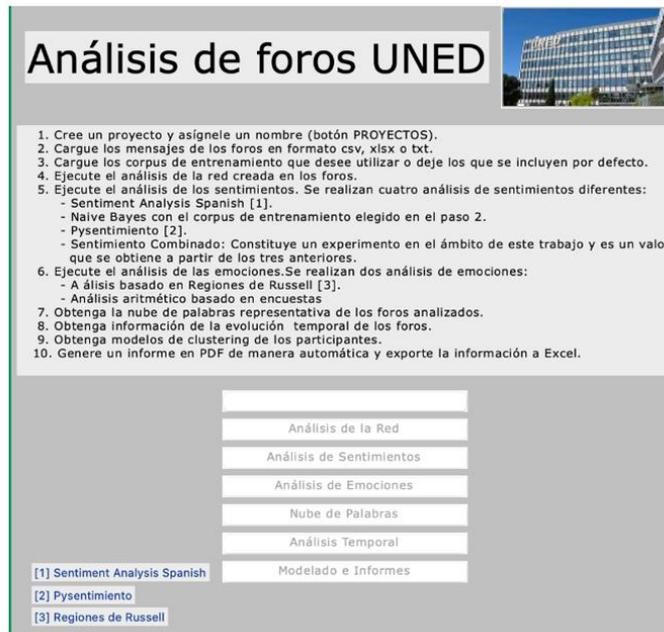
Encuesta a usuarios probadores

A fin de mejorar la aplicación y evaluar la experiencia de uso de la misma, se ha realizado una encuesta entre cinco usuarios voluntarios. A continuación, se reproduce el contenido de la encuesta realizada a través de Formularios de Google



The image shows a screenshot of a Google Form. At the top, there is a green header bar with icons for various Google services. Below the header, the form title is "ENCUESTA: Usabilidad y utilidad de la aplicación ANÁLISIS DE FOROS UNED". The text below the title states: "La siguiente encuesta tiene como objetivo identificar la usabilidad y utilidad de la aplicación para su posterior actualización y mejora." and "Lea atentamente las preguntas y responda con sinceridad." Below this, the user's email "felix.adame@gmail.com" is displayed with a "Cambiar de cuenta" link and a "No compartido" status. A red asterisk indicates that the following question is mandatory. The question is "Nombre *" and the response field is empty, showing "Tu respuesta" as a placeholder.

1- Experiencia de uso inicial



1.1. ¿Está de acuerdo con la siguiente expresión?:

Es fácil saber, desde la primera pantalla, qué puede hacer la aplicación.

- Totalmente de acuerdo
- Parcialmente de acuerdo
- En desacuerdo

RESPUESTAS

- Parcialmente de acuerdo
- Parcialmente de acuerdo
- Parcialmente de acuerdo
- Totalmente de acuerdo
- Totalmente de acuerdo

1.2. ¿Está de acuerdo con la siguiente expresión?:

Es fácil saber, desde la primera pantalla, cómo se utiliza la aplicación

Totalmente de acuerdo
Parcialmente de acuerdo
En desacuerdo

RESPUESTAS

- Parcialmente de acuerdo
- Parcialmente de acuerdo
- Totalmente de acuerdo
- Parcialmente de acuerdo
- Parcialmente de acuerdo

1.3. ¿Está de acuerdo con la siguiente expresión?:

La ayuda es clara y suficiente para comenzar a utilizar la aplicación

Totalmente de acuerdo
Parcialmente de acuerdo
En desacuerdo

RESPUESTAS

- Parcialmente de acuerdo
- Parcialmente de acuerdo
- Totalmente de acuerdo
- Totalmente de acuerdo
- En desacuerdo

2- Análisis de la Red

Estadísticas de la red

Centralidad de Grado Cercanía Intermediación Centralidad de Vector Propio Page Rank Iniciativa Modularidad

Etiquetas de gráfico: Alias Sentimiento: N.Bayes (Corpus Usuario)

Filtrar por: todo Seleccionar: TODO

	NODO	ALIAS	GRADO	GRADO_EN	GRADO_SAL	CENTRALID	CERCANIA	INTERMEDI	VEC
1	79	u1	25	16	9	1.67	1.00	0.52	0.6
2	66182	u2	7	3	4	0.47	0.56	0.11	0.25
3	64968	u3	5	3	2	0.33	0.56	0.11	0.25
4	22946	u4	7	4	3	0.47	0.58	0.11	0.29
5	71704	u5	14	5	9	0.93	0.6	0.18	0.32
6	39712	u6	6	3	3	0.4	0.56	0.17	0.25
7	25964	u7	6	3	3	0.4	0.56	0.17	0.25
8	69829	u8	13	5	8	0.87	0.6	0.17	0.33
9	25055	u9	5	4	1	0.33	0.58	0.11	0.29
10	51194	u10	2	1	1	0.13	0.38	0.11	0.04
11	55100	u11	2	1	1	0.13	0.067	0.054	2.5e-
12	58223	u12	2	0	2	0.13	0	0.054	2.5e-
13	60223	u13	1	0	1	0.067	0	0.05	2.5e-

2.1. Valore la usabilidad de este módulo. ¿Le resulta fácil navegar por este módulo y de entender qué se muestra? *

1 2 3 4 5 6 7 8 9 10

RESPUESTAS

7, 5, 4, 8, 8

2.2. Valore la utilidad de este módulo. ¿La información que se muestra le ha resultado útil? *

1 2 3 4 5 6 7 8 9 10

RESPUESTAS

9, 5, 7, 9, 8

2.5. ¿Qué te ha resultado menos práctico o útil?

RESPUESTAS

- NO RESPONDE.
- Etiquetar las cabeceras del archivo de Excel.
- En caso de un gran número de participantes es difícil sacar conclusiones a partir de los gráficos de red. No son fáciles de interpretar.
- una aclaración sobre la organización global de la tabla.
- Se necesita tener conocimiento de los indicadores para poder sacarle partido.

2.6. ¿Qué te ha resultado más práctico o útil?

RESPUESTAS

- La salida de datos en formato tabla y los gráficos.
- El valor de cercanía.
- Tanto las estadísticas como los gráficos son útiles, pero quizá lo serían más si se simplificasen pensando en profesores sin formación estadística que no saben interpretar la mayoría de los indicadores proporcionados.
- la visión conjunta de todo lo sucedido.
- La variabilidad de medidas que ofrece y el poder filtrar sobre ellas.

2.7. ¿Qué añadiría o quitaría en este módulo?

RESPUESTAS

- Más ayudas para entender lo que hace cada función y cómo se puede configurar.
- Mejorar la representación gráfica
- Proporcionaría la opción de ver solo una versión simplificada, así como una descripción textual. Sería interesante facilitar el análisis de cómo se propagan los sentimientos en la red y qué miembros de la red y mensaje (por ejemplo, de los profesores) tienen un mayor papel en ello.

- aclaraciones "flotantes" que expliquen el contenido de cada columna al situarse sobre ella.
- Que se puedan definir nuevas medidas para mostrar ad-hoc.

3- Análisis de Sentimientos

Análisis de Sentimientos del proyecto Toni-MDEME

Tabla de Sentimientos Mostrar Gráficas Sentimientos con Iconos Sentimiento Medio

Tabla de mensajes -----> Desplazar la barra inferior para ver toda la información ----->

	Emisor	Alias	Destinatario	Sentím. Analy	Sentím. Analy	Sent. N. Bay	PySentiment	Se
1	79	u1	Inicia el hilo	2.3e-29	2.3e-29	pos	NEU	0.0
2	64968	u3	79	6.5e-26	6.5e-26	pos	NEU	0.0
3	22946	u4	79	3.2e-15	3.2e-15	pos	POS	0.0
4	51194	u10	79	1.3e-24	1.3e-24	pos	POS	0.0
5	39712	u6	51194	4.1e-15	4.1e-15	pos	POS	0.0
6	55100	u11	79	1.5e-24	1.5e-24	pos	POS	0.0
7	58223	u12	55100	2.5e-14	2.5e-14	pos	NEU	0.0
8	25055	u9	79	4.4e-14	4.4e-14	pos	POS	0.0
9	60223	u13	79	1.9e-17	1.9e-17	pos	NEU	0.0
10	79	u1	Inicia el hilo	9e-08	9e-08	pos	NEU	0.0
11	55100	u11	79	7.6e-07	7.6e-07	neg	POS	0.0
12	25055	u9	79	0.0021	0.0021	pos	POS	0.0
13	60223	u13	79	4.2e-14	4.2e-14	pos	POS	0.0

3.1. Valore la usabilidad de este módulo. ¿Le resulta fácil navegar por este módulo y de entender qué se muestra? *

1 2 3 4 5 6 7 8 9 10

RESPUESTAS

8, 5, 9, 8, 8.

3.2. Valore la utilidad de este módulo. ¿La información que se muestra le ha resultado útil? *

1 2 3 4 5 6 7 8 9 10

RESPUESTAS

6, 5, 8, 9, 8.

3.3. ¿Qué te ha resultado menos práctico o útil?

RESPUESTAS

- La definición de qué hace este módulo y cómo se calculan las emociones discretas.
- La definición de sentimiento.
- Confunde el hecho de que se proporcionen los resultados de distintos métodos.
Entiendo que es interesante para utilizar la aplicación como herramienta de investigación, pero como herramienta docente debería incluir solo el o los métodos más fiables.
- Aclaraciones sobre la naturaleza del análisis de sentimiento: bases y cómo se realiza.
- La información numérica.

3.4. ¿Qué te ha resultado más práctico o útil?

RESPUESTAS

- La parte gráfica.
- La representación gráfica.
- Los gráficos se entienden muy bien.
- La posibilidad de ver una visión global del análisis de sentimientos.
- El destacar con colores la información de interés.

3.5. ¿Qué añadiría o quitaría en este módulo?

RESPUESTAS

- Más información sobre los corpus y cómo introducir otros.
- Algún corpus validado experimentalmente.
- Como he dicho, la diversidad de métodos.
- Aclaraciones sobre la naturaleza del análisis de sentimiento: bases y cómo se realiza.
- Mostrar la información agrupando usuarios según su emoción.

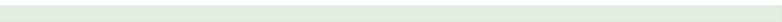
4- Análisis de Emociones

Análisis de Emociones del proyecto Toni-MDEME

Mostrar Tabla con Emociones Mostrar Gráficos Emociones Regiones de Russell Medias Russell por Usuarios Análisis Aritmético Medias Aritm. por Usuarios Mostrar Ley

Tabla de mensajes -----> Desplazar la barra inferior para ver toda la información

	Emisor	Alias	Destinatario	Emoción Reg.	Felicidad	Enojo	Tristeza
1	79	u1	Inicia el hilo	Feliz	32.23	16.83	19.43
2	64968	u3	79	Feliz	10.67	6.05	6.29
3	22946	u4	79	Feliz	2.87	2.37	2.43
4	51194	u10	79	Feliz	1.33	1.10	1.17
5	39712	u6	51194	Feliz	3.70	2.77	2.47
6	55100	u11	79	Feliz	23.90	12.40	13.80
7	58223	u12	55100	Feliz	12.40	6.70	7.93
8	25055	u9	79	Feliz	17.33	7.47	7.90
9	60223	u13	79	Feliz	5.07	2.13	2.37
10	79	u1	Inicia el hilo	Feliz	4.63	1.43	1.30
11	55100	u11	79	Ninguna	0	0	0
12	25055	u9	79	Ninguna	0	0	0
13	69829	u8	79	Feliz	9.57	3.63	4.53



4.1. Valore la usabilidad de este módulo. ¿Le resulta fácil navegar por este módulo y de entender qué se muestra? *

1 2 3 4 5 6 7 8 9 10

RESPUESTAS

8, 5, 8, 7, 9

4.2. Valore la utilidad de este módulo. ¿La información que se muestra le ha resultado útil? *

1 2 3 4 5 6 7 8 9 10

RESPUESTAS

5, 5, 9, 7, 9

4.3. ¿Qué te ha resultado menos práctico o útil?

RESPUESTAS

- La definición de qué hace este módulo.

- La definición de emociones.
- De nuevo, los resultados contradictorios por parte de distintos métodos.
- El análisis emociones es especialmente sutil y hubiese deseado unas aclaraciones sobre cómo se ha obtenido.
- La info numérica.

4.4. ¿Qué te ha resultado más práctico o útil?

RESPUESTAS

- La parte gráfica y poder sacar los valores.
- La representación gráfica.
- Las distintas opciones de gráficos.
- ver el estado global del análisis de emociones.
- Poder comparar con diferentes procesamientos.

4.5. ¿Qué añadiría o quitaría en este módulo?

RESPUESTAS

- Más información sobre los corpus y cómo introducir otros por el usuario. Poder usar un filtros para cruzar variables configurándolas por el usuario. Poder crear un modelo de usuario que se pueda consultar para alimentar otras aplicaciones (p.e. un LMS).
- Limitar el etiquetado a emociones básicas, difíciles de entender en contextos de aprendizaje.
- De nuevo, la diversidad de métodos a aplicar.
- unas aclaraciones sobre cómo se ha obtenido.
- Poder configurar otros procesamientos.

5- Nube de Palabras

Nube de palabras del proyecto Toni-MDEME



5.1. Valore la usabilidad de este módulo. ¿Le resulta fácil navegar por este módulo y de entender qué se muestra? *

1 2 3 4 5 6 7 8 9 10

RESPUESTAS

8, 5, 9, 5, 8

5.2. Valore la utilidad de este módulo. ¿La información que se muestra le ha resultado útil? *

1 2 3 4 5 6 7 8 9 10

RESPUESTAS

2, 5, 5, 5, 6

5.3 ¿Qué añadiría o quitaría en este módulo?

RESPUESTAS

- No aporta nada a nivel gráfico. Es más interesante hacer análisis de frecuencias en función de hilos, foros y usuarios para hacer Distant Reading.
- El análisis de la frecuencia de las palabras, segmentado por foros y por hilos.
- Entiendo que este módulo hay que depurarlo de modo que no salgan palabras de uso común que no revelan nada de interés.
- la posibilidad de acceder a los datos en una tabla desde la cual se pueda hacer análisis del asunto.
- NO RESPONDE.

6- Análisis Temporal

Evolución Temporal del proyecto Toni-MDEME



6.1. Valore la usabilidad de este módulo. ¿Le resulta fácil navegar por este módulo y de entender qué se muestra? *

1 2 3 4 5 6 7 8 9 10

RESPUESTAS

7, 5, 9, 9, 10

6.2. Valore la utilidad de este módulo. ¿La información que se muestra le ha resultado útil? *

1	2	3	4	5	6	7	8	9	10
<input type="radio"/>									

RESPUESTAS

6, 5, 9, 9, 10

6.3. ¿Qué te ha resultado menos práctico o útil?

RESPUESTAS

- La forma de visualización de datos.
- La forma en la que se visualizan los datos.
- Todo es útil.
- contrastar el análisis con la media acumulado.
- NO RESPONDE.

6.4. ¿Qué te ha resultado más práctico o útil?

RESPUESTAS

- No es el módulo más práctico pero sirve para ver en global la evolución de los datos de interacción a lo largo del tiempo. Por ejemplo saber en qué momentos el debate ha estado más encendido.
- Poder visualizar a lo largo del tiempo los momentos en los que ha habido más interacciones.
- Todo es útil.
- ver la evolución global de la actividad.
- NO RESPONDE.

6.5. ¿Qué añadiría o quitaría en este módulo?

RESPUESTAS

- Usar filtros para cruzar datos y poder cambiar las escala temporal. Tal vez hacer zoom en un punto en el tiempo y ver de qué mensaje parte un incremento en la interacción.
- Poder filtrar los datos en más categorías temporales.
- Para ayudar a los profesores a interpretar estos gráficos sería interesante que se integrase en ellos información sobre los diferentes hitos del curso, fechas señaladas en el calendario académico, fechas de entregas, eventos varios...
- medias del acumulado y enlaces directos desde los niveles superiores a otras medidas como la centralidad.
- Mas opciones para filtrar y comparar.

7- Modelo e Informe

Modelos del proyecto Toni-MDEME

Mensajes: 122 Participantes: 16

Información de cada participante en los foros ---> Desplace la barra inferior para ver todas las columnas. --->

	Participante	Alias Usuario	N.Mensajes	Hilos Iniciado	Resp. A Otro	A sí mismos +	A Sus Hilos	A Otros H
1	79	u1	20	5	13	7	10	10
2	66182	u2	11	2	9	2	4	7
3	64968	u3	6	1	5	1	1	5
4	22946	u4	10	1	9	1	2	8
5	71704	u5	23	3	20	3	5	18
6	39712	u6	7	2	5	2	3	4
7	25964	u7	5	1	4	1	1	4
8	69829	u8	22	2	20	2	6	16
9	25055	u9	5	1	4	1	1	4
10	51194	u10	2	0	2	0	0	2

7.1. Valore la usabilidad de este módulo. ¿Le resulta fácil navegar por este módulo y de entender qué se muestra? *

1 2 3 4 5 6 7 8 9 10

RESPUESTAS

6, 5, 9, 9, 8.

7.2. Valore la utilidad de este módulo. ¿La información que se muestra le ha resultado útil? *

1	2	3	4	5	6	7	8	9	10
<input type="radio"/>									

RESPUESTAS

9, 7, 8, 9, 8.

7.3. ¿Qué te ha resultado menos práctico o útil?

RESPUESTAS

- No saber los usuarios asignados a un cluster.
- Desconocer qué usuarios han sido asignados a un cluster.
- Todos los datos proporcionados son útiles.
- El informe es completo pero debería haber mayor control sobre lo que se incluye.
- NO REESPONDE.

7.4. ¿Qué te ha resultado más práctico o útil?

RESPUESTAS

- La exportación de datos.
- La exportación de los datos.
- Todos los datos son útiles.
- El ver de forma global todo lo analizado.
- El informe.

7.5. ¿Qué añadiría o quitaría en este módulo?

RESPUESTAS

- Información para configurar los informes, con distintos niveles de detalle.
- La posibilidad de poder configurar los elementos a incluir en el informe.
- Igualmente, un modo de poner en conexión con el calendario académico y posibles eventos destacados.
- Mayor flexibilidad para incluir unos u otros elementos en el informe.
- Poder utilizar otro algoritmos para ver las agrupaciones.

8- General

8.1. Una vez que ha utilizado la aplicación, ¿para qué sirve la aplicación? Escriba un breve comentario.

RESPUESTAS

- Analizar las interacciones entre los usuarios de foros. Hace análisis automático de las emociones y sentimientos presentes en la interacciones escritas.
- Analizar las interacciones entre usuarios y el contenido emocional de dichas interacciones.
- Para tomar el pulso a los estudiantes y conocer el tono afectivo de la clase en su conjunto así como de cada estudiante en particular.
- Para realizar un análisis de emociones y sentimientos extraído de la interacción en el foro de los estudiantes.
- Para conocer informacion de la interaccion de los alumnos en un foro.

8.2. Una vez que ha utilizado la aplicación, ¿qué ofrece la aplicación? Escriba un breve comentario.

RESPUESTAS

- Computa indicadores de interacción y lo presenta en formato tabular o gráfico en función de lo que se seleccione.
- El cómputo de las interacciones entre usuarios y su representación gráfica.
- Para que la aplicación sea útil el profesor tiene que estar formado en enfoques didácticos afectivos. De nada sirve la información que proporciona la aplicación si el profesor no sabe utilizarla en modo productivo, o incluso la utiliza en modo perjudicial para la buena marcha del curso.
- Análisis de Sentimientos, Gráficos de Sentimientos, Análisis de Emociones, Gráficos de Emociones, Estadísticas de los mensajes y red, Gráficos de la red, Gráficos de clustering, Gráficos temporales, Gráficos de Usuarios, Nubes de palabras del proyecto.
- Ver el grado de colaboración y el estado emocional reflejado en los mensajes de los foros.

¿Qué **añadiría** para que la aplicación fuera ...

8.3. **Más usable?** Escriba un breve comentario.

RESPUESTAS

- Un entorno más gráfico, basado en estándares de interacción: botones, filtros, cajas, sliders...
- Un entorno gráfico más amigable.
- Es bastante usable.
- poder hacer análisis centrados en un usuario.
- Aclaraciones o leyendas.

8.4. **Más intuitiva?** Escriba un breve comentario.

RESPUESTAS

- Ídem = Un entorno más gráfico, basado en estándares de interacción: botones, filtros, cajas, sliders...
- Un mejor etiquetado.
- Depende de los usuarios destinatarios. Para profesores de escasa formación científica está claro que tanta profusión de parámetros, datos y gráficos debe simplificarse, y quizá sustituirse o ampliarse con descripciones textuales.
- aclarar algunos términos, como "atajos".
- Botones con lenguaje del profesor, no del desarrollador.

8.5. **Más práctica?** Escriba un breve comentario.

RESPUESTAS

- Funcionalidad de filtros configurables.
- Poder filtrar mejor los datos.
- Conectar los resultados mostrados de forma explícita con el transcurrir del curso. Destacar los datos más llamativos en resúmenes.
- facilitar su uso en la web sin necesidad de instalaciones.
- Poder mezclar info de diferentes secciones, p.e. ver la centralidad de los alumnos con determinado estado anímico.

¿Qué **quitaría** para que la aplicación fuera ...

8.6. **Más usable?** Escriba un breve comentario.

RESPUESTAS

- Mejoraría la estética y las ayudas.
- Poder disponer de más ayudas.
- Descargar un poco de opciones. En particular, como ya he señalado, no ofrecer distintos métodos. Mejor hacer dos versiones de la herramienta: para investigación y para profesores comunes escasamente formados en técnicas estadísticas y minería de datos.
- nomenclaturas no explicadas.
- Las tablas, mostrarlas solo si se piden.

8.7. **Más intuitiva?** Escriba un breve comentario.

RESPUESTAS

- Simplificaría la carga y configuración de datos al crear el proyecto.
- Simplificar la forma de cargar los datos.
- Proporcionar menos indicadores.
- una visión inicial resumida mediante una presentación interactiva.
- NO RESPONDE.

8.8. **Más práctica?** Escriba un breve comentario.

RESPUESTAS

- Simplificaría la carga y configuración de datos al crear el proyecto.
- Ídem = Simplificar la forma de cargar los datos.
- Conectar con eventos y calendario académico.
- focalizar el análisis en individuos, sentimientos y emociones, y relacionar éstos con el grado de colaboración del individuo.
- Poder configurar para que muestre solo las secciones que necesitas.

8.9. Valore la **usabilidad** de la aplicación entre 1 y 10. *

1	2	3	4	5	6	7	8	9	10
<input type="radio"/>									

RESPUESTAS

6, 5, 5, 8, 8.

8.10. Valore la **utilidad** de la aplicación entre 1 y 10. *

1	2	3	4	5	6	7	8	9	10
<input type="radio"/>									

RESPUESTAS

9, 7, 8, 8, 8

Apéndice C

Informe PDF Generado Automáticamente por la Aplicación

A continuación se reproduce un informe de ejemplo generado automáticamente por la aplicación desarrollada.

Informe del proyecto: Prueba ; MDEME-Foro-22-11-29.csv



Fecha: 24/7/2023

Autor: Félix Adame Toledano

Este informe contiene un resumen de la sesión del proyecto Prueba ; MDEME-Foro-22-11-29.csv realizado con la aplicación Learning Analytics UNED.

A fin de poder mostrar la información de manera clara y dado que los nombres de usuarios pueden ser considerablemente largos, la aplicación ha generado un alias para cada usuario que se puede ver en la tabla 1 en la página siguiente:

Apartados del informe (opcionales)

Correspondencia entre los alias usados en el informe y los usuarios (tabla 1)

Análisis de Sentimientos

Gráficos de Sentimientos

Análisis de Emociones

Gráficos de Emociones

Estadísticas de los mensajes

Estadísticas de la red

Gráficos de la red

Gráficos de clustering

Gráficos temporales

Gráficos de Usuarios

Nubes de palabras del proyecto

Informe de Learning Analytics de Foros UNED. V f1.2

1. Correspondencia entre los alias usados en el informe y los usuarios (tabla 1)

Alias Usuario	Usuario
u1	79
u2	66182
u3	64968
u4	22946
u5	71704
u6	39712
u7	25964
u8	69829
u9	25055
u10	51194
u11	55100
u12	58223
u13	60223
u14	66967
u15	23709
u16	23778

Informe de Learning Analytics de Foros UNED. V f1.2

2. Análisis de Sentimientos

* El sentimiento combinado se corresponde con la fórmula:

$$\text{Sent Combinado} = ((\text{coefNB} * \text{Nbnum} + \text{coefPyS} * \text{PySnum} + \text{coefSentSpNorm} * \text{Sent.Norm}^{\wedge} \text{pot}) / 3) * \text{coefTot}$$

- NBnum es una transformación de literal a valor numérico del análisis

N.Bayes: POSITIVO = 0.75; NEGATIVO = 1e-1.

- PySnum es una transformación de literal a valor numérico del análisis

PySentimiento: POSITIVO = 1; NEGATIVO = 1e-2; NEUTRO = 0.5.

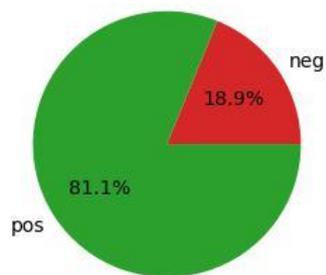
* El sentimiento normalizado es: $[\text{Sentimiento} - \text{Min}(\text{Sentimiento})] / [\text{Max}(\text{Sentimiento}) - \text{Min}(\text{Sentimiento})]$

donde 'Sentimiento' es el sentimiento calculado por el algoritmo de la librería 'Sentiment Analysis Spanish'

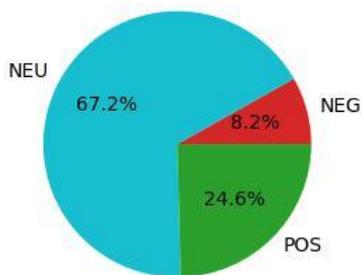
Alias Usuario	Sentiment Analysis Spanish Medio	Sent. Medio Combinado
u1	0.04	Regular
u2	0.2	Bueno
u3	0.0	Regular
u4	0.05	Regular
u5	0.06	Regular
u6	0.0	Regular
u7	0.0	Regular
u8	0.03	Regular
u9	0.0	Bueno
u10	0.0	Regular
u11	0.0	Bueno
u12	0.04	Regular
u13	0.0	Regular
u14	0.01	Regular
u15	0.0	Bueno
u16	0.0	Regular

3. Gráficos de Sentimientos

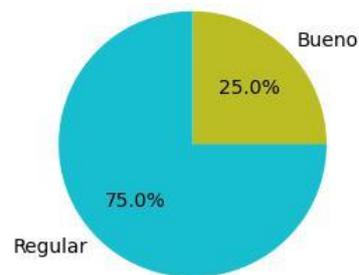
Todos los Mensajes. Análisis 1. N. Bayes



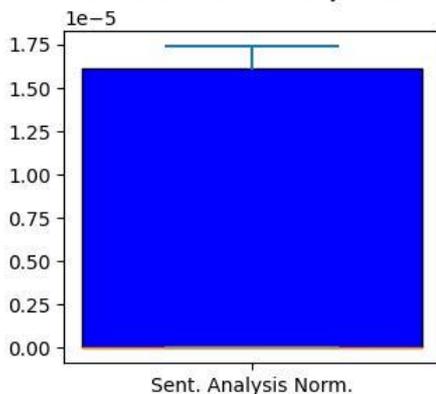
Todos los Mensajes. Análisis 2. PySentimiento



Medias Participantes. Combinado



Cuartiles de Sentim. Analys. Norm.



Cuartiles de Sentimiento Combinado

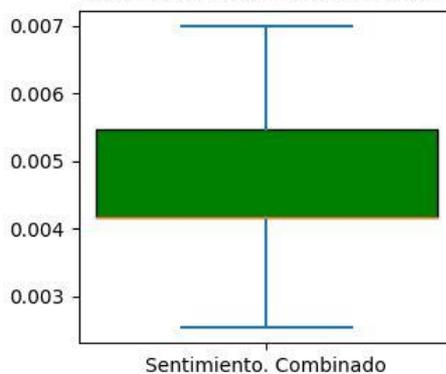


Gráfico de sentimientos N°1. Archivo: graf_sent_Prueba ; MDEME-Foro-22-11-29.csv.png.

ACLARACIONES SOBRE EL APARTADO DE GRÁFICOS SENTIMIENTOS

A continuación se hacen algunas consideraciones sobre los gráficos de sentimientos.

- Se representan cuatro gráficos (tres gráficos de tarta y un gráfico de cajas). A continuación se describe cada uno de ellos.

- 1. Medias Participantes. Combinado: Este gráfico representa la proporción de cada sentimiento en el conjunto de todos los participantes. Para llegar a él se calculan para cada participante tres sentimientos medios con tres métodos diferentes (Sentiment Analysis Spanish, Naives Bayes y PySentimiento). De estos, sólo el primero se corresponde con un valor numérico, mientras que los otros dos devuelven literales; positivo o negativo (N. Bayes) y positivo, negativo o neutro (PySentimiento). Por ello se realiza una transformación de los literales a valores numéricos. En base a numerosas pruebas realizadas, la transformación realizada es la siguiente:

- Para N.Bayes: POSITIVO = 0.75; NEGATIVO = 1e-1.
- Para PySentimiento: POSITIVO = 1; NEGATIVO = 1e-2; NEUTRO = 0.5.

- Además se hace una normalización del valor de Sentiment Analysis Spanish de la siguiente manera:

$$\frac{\text{Sentimiento} - \text{Min}(\text{Sentimiento})}{\text{Max}(\text{Sentimiento}) - \text{Min}(\text{Sentimiento})}$$

- Tras lo anterior se aplica la fórmula:

$$\frac{(\text{NBnum} + \text{PySnum} + (1.5 * \text{SQRT}(\text{Sentiment Analysis Spanish Normalizado}))/3) * 0.01}{}$$

Lo que arroja un valor que es el que denominamos sentimiento combinado.

- Para finalizar, se vuelve a realizar una transformación del valor numérico en literal en base a los siguientes tramos:

- Muy Malo: sentimiento ≤ 0.0001 .
- Malo: sentimiento ≤ 0.002 .
- Regular: sentimiento ≤ 0.005 .
- Bueno: sentimiento ≤ 0.006 .
- Muy Bueno: sentimiento ≤ 100 .

* A este último valor se le ha llamado sentimiento combinado literal.

- 2. Cuartiles de los valores numéricos: Este gráfico de cajas muestra los cuartiles de los valores globales de todos los mensajes (no por participantes), así como los máximos y mínimos (bigotes) ignorando valores atípicamente altos o bajos, para Sentiment Analysis Spanish y Sentimiento Combinado.

- 3. Todos los mensajes. N. Bayes: Gráfico de tarta que representa la proporción de mensajes positivos y negativos según N. Bayes.

- 4. Todos los mensajes. PySentimiento: Gráfico de tarta que representa la proporción de mensajes positivos, negativos y neutros según PySentimiento.

4. Análisis de Emociones

Región de Russell				Método Aritmético Alternativo					
Alias Usuario	Val. Media	Excit. Media	Región Russell	Felicidad	Enfado.	Tristeza	Miedo	Disgusto	Dominante
u1	8.48	5.89	Feliz	7.15	4.07	4.41	5.31	3.56	Felicidad
u2	4.87	2.94	Triste	3.61	1.67	1.72	2.15	1.54	Felicidad
u3	11.93	8.1	Feliz	8.1	3.59	3.97	4.77	3.43	Felicidad
u4	8.28	5.79	Feliz	4.77	2.76	2.84	3.45	2.51	Felicidad
u5	16.51	12.02	Feliz	10.23	5.86	6.32	7.62	5.23	Felicidad
u6	10.7	7.62	Feliz	7.57	4.36	4.84	6.0	3.92	Felicidad
u7	14.85	10.21	Feliz	11.67	5.26	5.84	6.33	4.82	Felicidad
u8	11.65	8.07	Feliz	7.9	3.99	4.34	4.94	3.68	Felicidad
u9	22.14	16.17	Feliz	14.99	7.79	8.31	9.84	7.41	Felicidad
u10	16.42	11.3	Feliz	1.28	1.72	2.13	2.25	1.35	Miedo
u11	13.38	9.34	Feliz	9.7	4.98	5.51	6.04	5.09	Felicidad
u12	6.85	4.43	Relajado	8.52	3.85	4.52	4.65	3.77	Felicidad
u13	12.76	8.04	Feliz	5.07	2.13	2.37	2.57	2.47	Felicidad
u14	7.18	4.85	Relajado	7.12	3.98	4.12	4.93	3.74	Felicidad
u15	22.43	18.39	Feliz	27.83	16.67	19.93	23.37	14.47	Felicidad
u16	14.49	9.61	Feliz	5.53	2.73	3.33	3.37	2.37	Felicidad

ACLARACIONES SOBRE EL APARTADO DE EMOCIONES

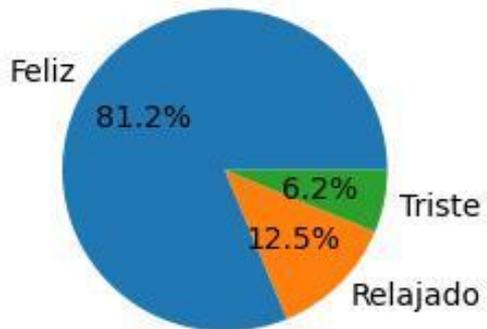
A continuación se hacen algunas consideraciones sobre este apartado.

Se utilizan dos métodos para el descubrimiento de emociones en esta aplicación: Regiones de Russell y un método aritmético basado en emociones evocadas directamente por las palabras. A continuación se describen someramente ambos métodos.

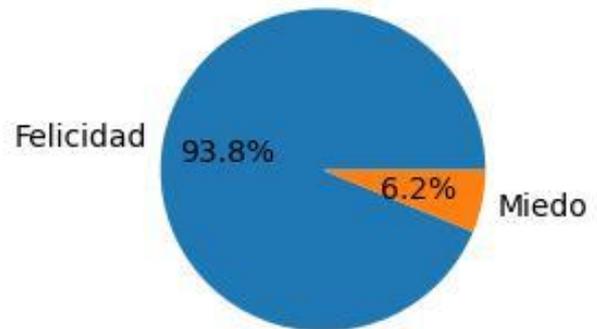
- **Regiones de Russell:** Se dispone de un corpus de palabras que por defecto se encuentra en el archivo 'Spanish-es-NRC-VAD-Lexicon.txt', en la carpeta 'recursos'. Cada palabra tiene asociado un valor de Valencia, un valor de Excitación y otro de Dominancia. Para el cálculo de la región de Russell se hace uso de la valencia y la excitación, las cuales definen una región teórica sobre un modelo circular representativo de las emociones. Las emociones consideradas son: alegre, triste, relajado, enfadado y sin emoción.
- **Método alternativo aritmético:** Se emplea un corpus en el que cada palabra tiene asociado un valor para cada una de seis emociones consideradas básicas o primarias: felicidad, enfado, tristeza, miedo y disgusto. El corpus por defecto es el archivo '13428_2016_768_MOESM1_ESM.xls', en la carpeta 'recursos'. Consta de 2266 palabras. Se ha diseñado un algoritmo para utilizar este corpus realizado en un estudio publicado por investigadores españoles (Moved by words: Affective ratings for a set of 2,266 Spanish words in five discrete emotion categories. Pilar Ferré et al.)
- La tabla está dividida en dos partes, una para cada método. En la parte izquierda se muestran los valores medios para cada participante de valencia, excitación y la correspondiente región media de Russell. En la parte derecha se muestran los valores medios de cada emoción básica y a la derecha la de mayor valor (dominante).

5. Gráficos de Emociones

Medias Particip. Análisis 1. Regiones Russell



Medias Particip. Análisis 2



Todos los Mensajes. Análisis 1. Regiones Russell



Todos los Mens. Análisis 2

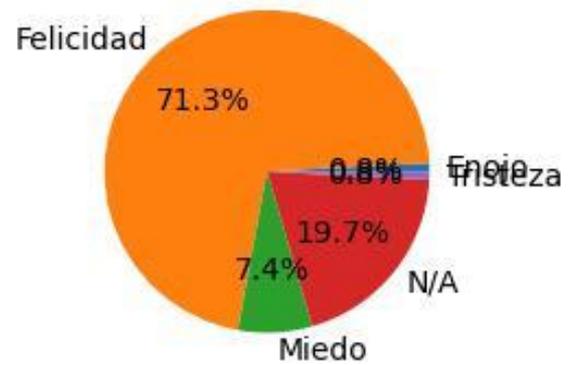


Gráfico de sentimientos N°1. Archivo: graf_emoc_Prueba ; MDEME-Foro-22-11-29.csv.png.

Informe de Learning Analytics de Foros UNED. V f1.2

6. Estadísticas de los mensajes

Alias Usuario	Nº Mensajes	Hilos Iniciados	Resp. A Otros	A sí mismos + Inicios	A Sus Hilos	A Otros Hilos
u1	20	5	13	7	10	10
u2	11	2	9	2	4	7
u3	6	1	5	1	1	5
u4	10	1	9	1	2	8
u5	23	3	20	3	5	18
u6	7	2	5	2	3	4
u7	5	1	4	1	1	4
u8	22	2	20	2	6	16
u9	5	1	4	1	1	4
u10	2	0	2	0	0	2
u11	3	0	3	0	0	3
u12	2	0	2	0	0	2
u13	1	0	1	0	0	1
u14	3	1	2	1	2	1
u15	1	0	1	0	0	1
u16	1	0	1	0	0	1

ACLARACIONES SOBRE LA TABLA DEL APARTADO 3 (Estadísticas de los mensajes)

A continuación se explica el significado de cada columna:

Alias Usuario:	Es el alias asignado automáticamente para acortar / anonimizar los usuarios.
Nº Mensajes:	Número total de mensajes escritos por cada usuario en los foros.
Hilos Iniciados:	Número total de hilos (conversaciones) iniciados por cada usuario.
Resp. A Otros:	Número de mensajes respondiendo a otros usuarios, ya sea en sus hilos o en otros hilos.
A sí mismo + Inicios:	Número de mensajes escritos que no son respuesta a otros usuarios.
A Sus Hilos:	Mensajes escritos en los hilos iniciados por el propio usuario.
A Otros Hilos:	Mensajes escritos en los hilos no iniciados por el propio usuario.
Nº Mensajes:	Número total de mensajes escritos por cada usuario en los foros.

Informe de Learning Analytics de Foros UNED. V f1.2

7. Estadísticas de la red

Alias Usuario	Grado	Grado Ent.	Grado Sal.	Cent. Grado	Cercanía	Intermed.	Cent. V. Prop.	Page Rank	Iniciativa	Modularidad
u1	25	16	9	1.67	1.0	0.52	0.6	0.3	5	0
u2	7	3	4	0.47	0.56	0.11	0.25	0.06	2	1
u3	5	3	2	0.33	0.56	0.11	0.25	0.06	1	1
u4	7	4	3	0.47	0.58	0.11	0.29	0.07	1	1
u5	14	5	9	0.93	0.6	0.18	0.32	0.09	3	1
u6	6	3	3	0.4	0.56	0.17	0.25	0.06	2	0
u7	6	3	3	0.4	0.56	0.17	0.25	0.06	1	2
u8	13	5	8	0.87	0.6	0.17	0.33	0.1	2	1
u9	5	4	1	0.33	0.58	0.11	0.29	0.08	1	1
u10	2	1	1	0.13	0.38	0.11	0.05	0.03	0	0
u11	2	1	1	0.13	0.07	0.05	0.0	0.01	0	0
u12	2	0	2	0.13	0.0	0.05	0.0	0.01	0	0
u13	1	0	1	0.07	0.0	0.05	0.0	0.01	0	0
u14	4	2	2	0.27	0.42	0.11	0.13	0.03	1	1
u15	2	1	1	0.13	0.38	0.11	0.05	0.03	0	2
u16	1	0	1	0.07	0.0	0.05	0.0	0.01	0	0

8. Gráficos de la red

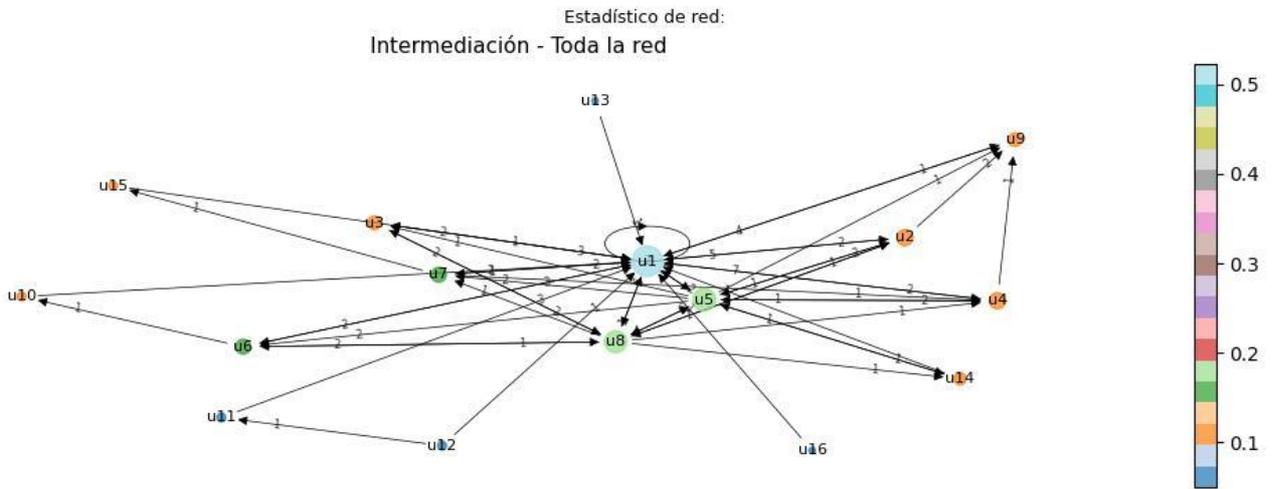


Gráfico de red N°1. Archivo: grafico_red_Intermediación - Toda la red_Alias.png.

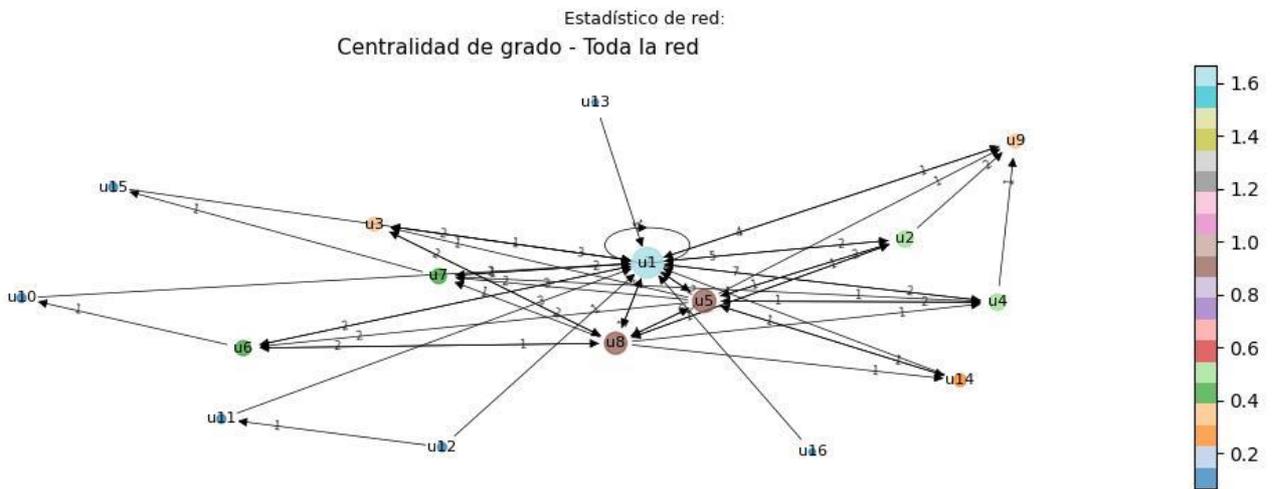


Gráfico de red N°2. Archivo: grafico_red_Centralidad de grado - Toda la red_Alias.png.

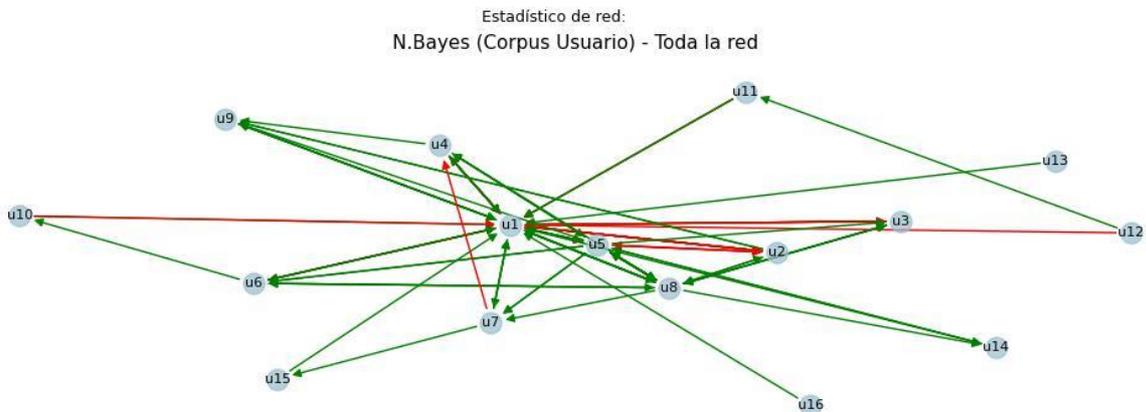


Gráfico de red N°3. Archivo: grafico_red_N.Bayes (Corpus Usuario) - Toda la red_Alias.png.

Informe de Learning Analytics de Foros UNED. V f1.2

Estadístico de red:
pysentimiento: A Python Toolkit for Sentymnet Analysis and Social NLP tasks - Toda la red

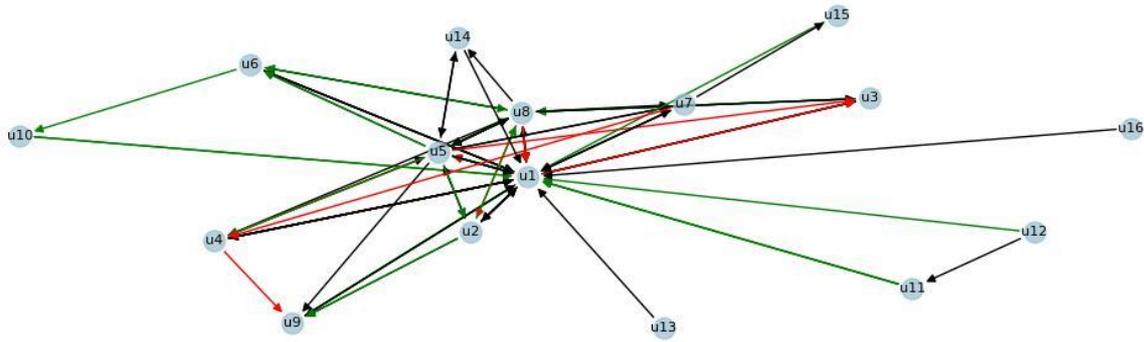


Gráfico de red N°4. Archivo: grafico_red_pysentimiento; A Python Toolkit for Sentymnet Analysis and Social NLP tasks - Toda la red_Alias.png.

ACLARACIONES SOBRE EL APARTADO DE GRÁFICOS DE RED

A continuación se hacen algunas consideraciones sobre los gráficos de red.

- Cada gráfico va acompañado de un título descriptivo, tal como muestra en la aplicación. Además al pie de cada gráfico aparece el nombre del archivo de imagen, el cual indica también el estadístico representado, el filtro aplicado o si se trata de toda la red y si se utilizan los alias o los nombres de usuarios (Ver la tabla 1 para conocer la correspondencia entre alias y nombres de usuarios).
- El tamaño de los nodos representa el grado (número de mensajes), mientras que el color representa el estadístico correspondiente de acuerdo a la barra de color situada a la derecha. En el caso de la modularidad se representan los grupos por colores y no procede la barra de color.
- En el caso de los gráficos de sentimientos de este apartado pueden ser de dos tipos: PySentimiento o N. Bayes. N. Bayes discrimina mensajes positivos de negativos, mientras que PySentimiento diferencia también mensajes de carácter neutro. El color de las aristas representa el sentimiento del mensaje: ROJO: Negativo; VERDE: Positivo y NEGRO: Neutro (Sólo para PySentimiento).
- Durante la sesión de trabajo se puede seleccionar el algoritmo que se desee para representar la 'red de sentimientos' mediante el desplegable situado en la parte superior antes pulsar el botón de 'Sentimientos'.
- Durante la sesión de trabajo se puede seleccionar también el tipo de etiqueta de nodo antes de dibujar cualquier gráfico de red (alias o usuarios).
- En los casos en que no se aprecie correctamente el gráfico es muy recomendable la utilización de la barra de herramientas para poder hacer zoom.

9. Gráficos de clustering

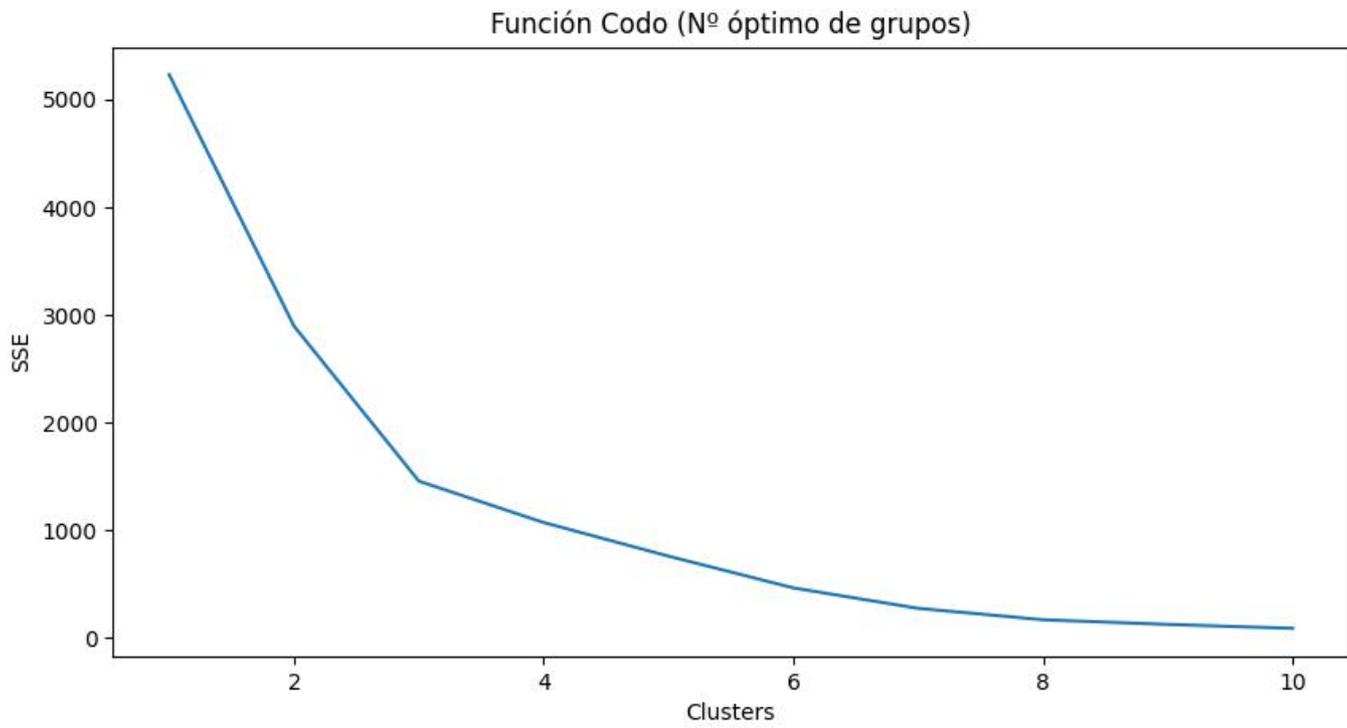


Gráfico de clustering N°1. Archivo: funcion_codo.png.

Kmeans

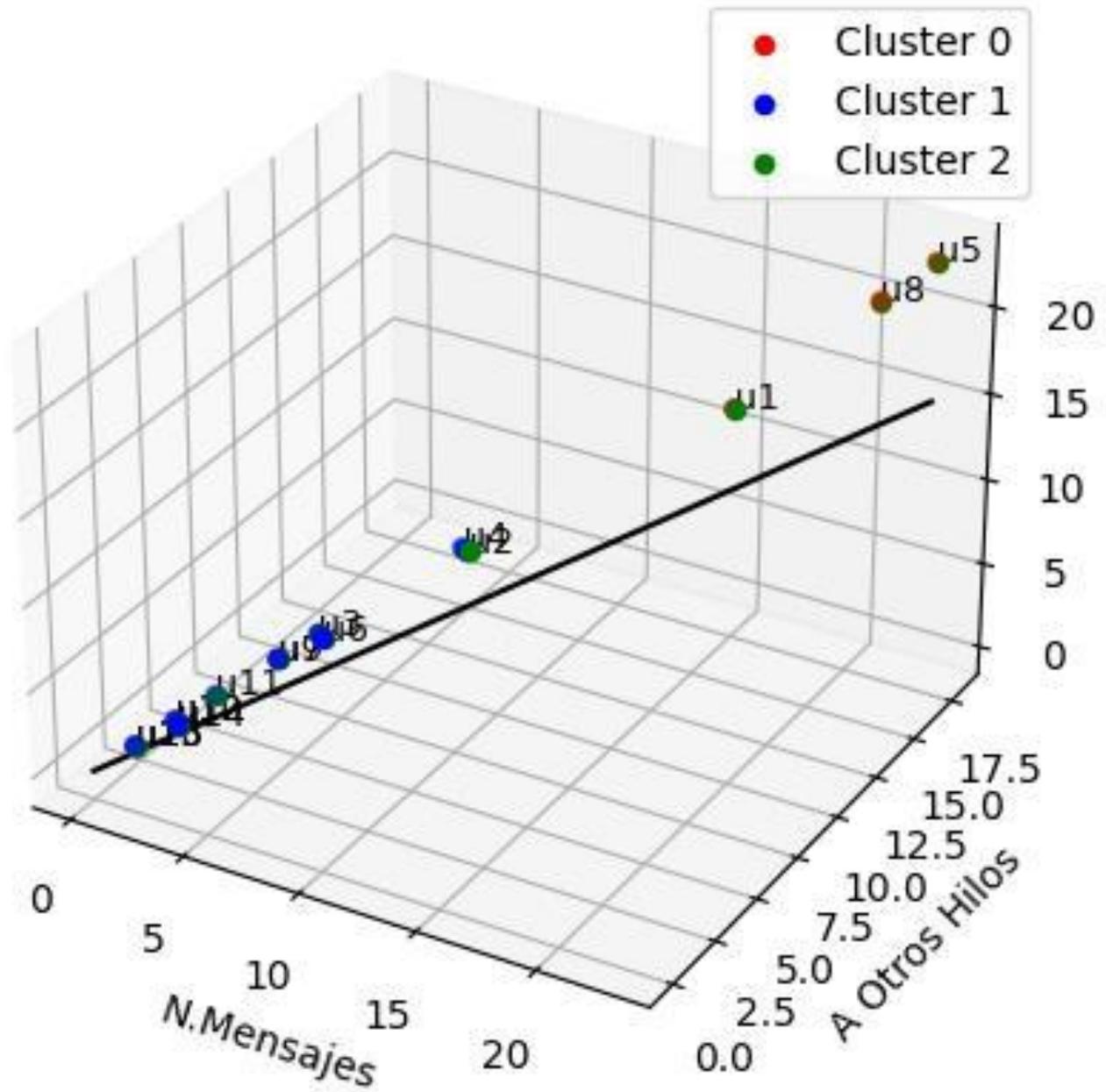


Gráfico de clustering N°2. Archivo: grafico_clusters_N.Mensajes_A Otros Hilos_3_etiquetas_Alias.png.

ACLARACIONES SOBRE EL APARTADO DE GRÁFICOS CLUSTERING

A continuación se hacen algunas consideraciones sobre los gráficos de clustering.

- El algoritmo utilizado aparece en el título del gráfico (en esta versión sólo está disponible Kmeans. Al pie de cada gráfico aparece el nombre del archivo de imagen, el cual indica las variables utilizadas para la representación del clustering, el número de clusters y el tipo de etiquetas de nodo.

Por ejemplo, dado el pie de gráfico 'Gráfico de clustering N°2. Archivo: grafico_clusters_N.Mensajes_A Otros Hilos_3_etiquetas_Alias.png', se está indicando que el gráfico generado en segundo lugar, utiliza el número de mensajes y los mensajes a hilos no iniciados por uno mismo como variables para los ejes X e Y respectivamente (el eje Z siempre es el número de mensajes), que se utilizan tres clusters y que se han seleccionado como etiquetas los alias. (Ver la tabla 1 para conocer la correspondencia entre alias y nombres de usuarios).

Un tipo de gráfico especial que se puede mostrar en esta sección es el correspondiente a la función codo. La función codo intenta determinar el número óptimo de grupos o clusters.

10. Gráficos Temporales

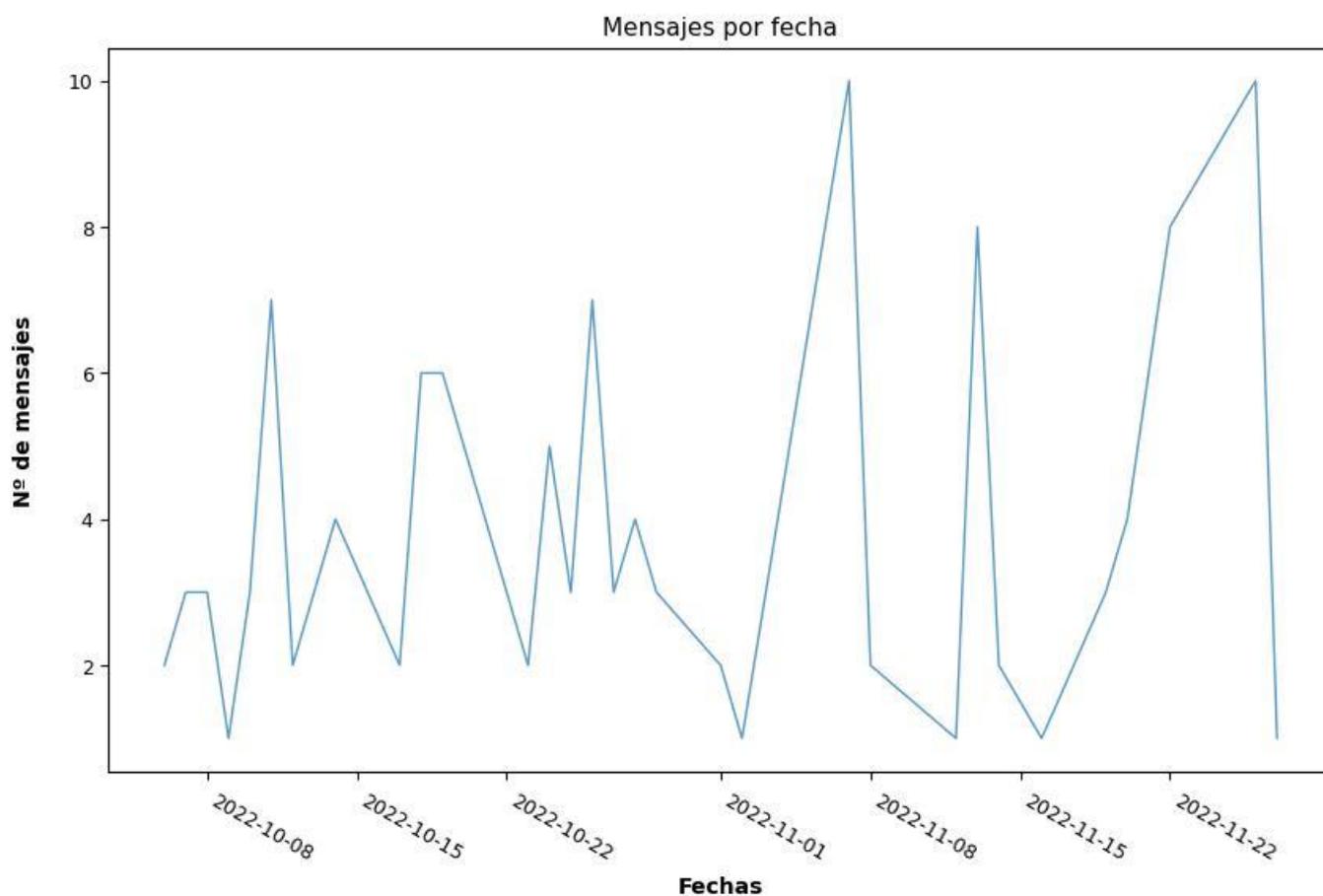


Gráfico temporal Nº1. Archivo: Mensajes por fecha.png.

Informe de Learning Analytics de Foros UNED. V f1.2

Mensajes por día de la semana

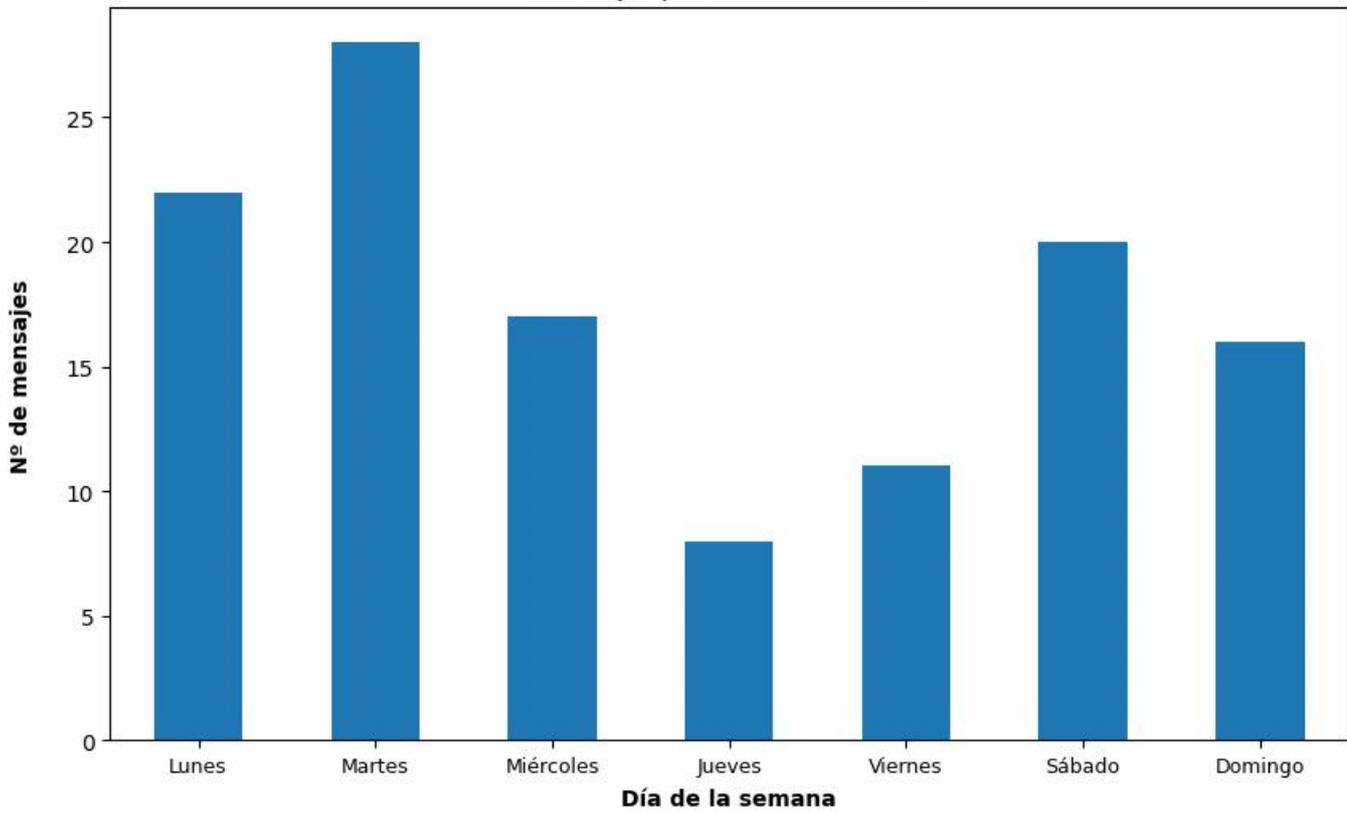


Gráfico temporal N°2. Archivo: Mensajes por día de semana.png.

ACLARACIONES SOBRE EL APARTADO DE GRÁFICOS TEMPORALES

A continuación se hacen algunas consideraciones sobre los gráficos temporales.

- Se pueden mostrar dos tipos de gráficos temporales:
- Gráfico de mensajes por fecha.
- Gráfico de mensajes por día de la semana.

En ambos casos se representan todos los mensajes del archivo.

Referencias bibliográficas

- [1] Lori Lockyer and Shane Dawson. 2011. Learning designs and learning analytics. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK '11). Association for Computing Machinery, New York, NY, USA, 153–156. <https://doi.org/10.1145/2090116.2090140>
- [2] Hernández-de-Menéndez, M., Morales-Menendez, R., Escobar, C.A. et al. Learning analytics: state of the art. *Int J Interact Des Manuf* 16, 1209–1230 (2022). <https://doi.org/10.1007/s12008-022-00930-0>
- [3] Klamma, R. (2013). Community Learning Analytics – Challenges and Opportunities. In: Wang, J.F., Lau, R. (eds) *Advances in Web-Based Learning – ICWL 2013*. ICWL 2013. Lecture Notes in Computer Science, vol 8167. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-41175-5_29
- [4] Song, E., Petrushyna, Z., Cao, Y., Klamma, R.: Learning Analytics at Large: The Lifelong Learning Network of 160,000 European Teachers. In: Kloos, C.D., Gillet, D., Crespo García, R.M., Wild, F., Wolpers, M. (eds.) *EC-TEL 2011*. LNCS, vol. 6964, pp. 398–411. Springer, Heidelberg (2011)
- [5] Baker, R.S.J.D., Yacef, K.: The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining* 1, 3–17 (2009).
- [6] Shum, S.B., Ferguson, R.: Social Learning Analytics. *Educational Technology & Society* 15, 3–26 (2012).
- [7] Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., Ludgate, H.: *NMC Horizon Report: 2013 Higher Education Edition* (2013).
- [8] Wenger, E.: *Community of Practice: Learning, Meaning, and Identity*. Cambridge University Press, Cambridge (1998)

- [9] Samuels, Antony and John Mcgonical. "News Sentiment Analysis." ArXiv abs/2007.02238 (2020): n. <https://arxiv.org/abs/2007.02238>
- [10] Fabio Clarizia, Francesco Colace, Massimo De Santo, Marco Lombardi, Francesco Pascale, and Antonio Pietrosanto. 2018. E-learning and sentiment analysis: a case study. In Proceedings of the 6th International Conference on Information and Education Technology (ICIET '18). Association for Computing Machinery, New York, NY, USA, 111–118. <https://doi.org/10.1145/3178158.3178181>
- [11] R. Colbaugh and K. Glass, "Estimating sentiment orientation in social media for intelligence monitoring and analysis," 2010 IEEE International Conference on Intelligence and Security Informatics, Vancouver, BC, Canada, 2010, pp. 135-137, doi: 10.1109/ISI.2010.5484760.
- [12] José Alberto Benítez Andrades. 2011. Estado del arte en Probabilistic Latent Semantic Analysis aplicado a problemas de acceso a la información en la Web.
- [13] M.D. Devika, C. Sunitha, Amal Ganesh, Sentiment Analysis: A Comparative Study on Different Approaches, Procedia Computer Science, Volume 87, 2016, Pages 44-49, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2016.05.124> (<https://www.sciencedirect.com/science/article/pii/S187705091630463X>).
- [14] Enrique J. Carmona Suárez. 2013. Tutorial sobre Máquinas de Vectores Soporte (SVM). <http://www.ia.uned.es/personal/ejcarmona>
- [15] Neha S. Joshi, Suhasini A. Itkat, "A Survey on Feature Level Sentiment Analysis" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5422-5425. <https://www.ijcsit.com/docs/Volume%205/vol5issue04/ijcsit20140504135.pdf>
- [16] P. Šaloun, M. Hruzík and I. Zelinka, "Sentiment analysis - e-bussines and e-learning common issue," 2013 IEEE 11th International Conference on Emerging eLearning Technologies and Applications (ICETA), Stara Lesna, Slovakia, 2013, pp. 339-343, doi: 10.1109/ICETA.2013.6674455. <https://ieeexplore.ieee.org/abstract/document/6674455>

- [17] Pablo Gamallo, Marcos Garcia, “Citius: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets” Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 171–175, Dublin, Ireland, August 23-24 2014.
- [18] A. Tamilselvi, M. ParveenTaj, “Sentiment Analysis of Micro blogs using Opinion Mining Classification Algorithm” International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Volume 2 Issue 10, October 2013.
- [19] Tan, Pang-Ning., Sanjay. Chawla, Chin Kuan. Ho, and James. Bailey. Advances in Knowledge Discovery and Data Mining, Part I 16th Pacific-Asia Conference, PAKDD 2012, Kuala Lumpur, Malaysia, May 29-June1, 2012, Proceedings, Part I. 1st Ed. 2012. ed. 2012. Lecture Notes in Artificial Intelligence ; 7301. Web.
https://buscador.biblioteca.uned.es/permalink/f/ffr813/34UNED_ALMA51103053860004215
- [20] Ankitha Srivastava, Dr.M.P. Singh, “Supervised SA of product reviews using Weighted k-NN Algorithm,” 2014 11th International Conference on Information Technology.
- [21] Rehman, A.U., Malik, A.K., Raza, B. et al. A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis. *Multimed Tools Appl* 78, 26597–26613 (2019). <https://doi.org/10.1007/s11042-019-07788-7>
- [22] Kawathekar, Swati. (2012). Movie Review analysis using Rule-Based & Support Vector Machines methods. *IOSR Journal of Engineering*. 02. 389-391. 10.9790/3021-0203389391.
https://www.researchgate.net/publication/271251630_Movie_Review_analysis_using_Rule-Based_Support_Vector_Machines_methods
- [23] Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225.
<https://doi.org/10.1609/icwsm.v8i1.14550>
- [24] Klekovkina, Maria and Eugeny V. Kotelnikov. “Метод автоматической классификации текстов по тональности, основанный на словаре эмоциональной

- лексики (The Automatic Sentiment Text Classification Method Based on Emotional Vocabulary).” Russian Conference on Digital Libraries (2012). <https://ceur-ws.org/Vol-934/paper15.pdf>
- [25] Man Lan, Chew Lim Tan, Jian Su, & Yue Lu. (2009). Supervised and Traditional Term Weighting Methods for Automatic Text Categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4), 721–735. doi:10.1109/tpami.2008.110. <https://doi.org/10.1109/TPAMI.2008.110>
- [26] Russell, James A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. doi:10.1037/h0077714
- [27] Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology*, 44(4), 229–237. doi:10.1037/h0055778
- [28] Schlosberg, H. (1954). Three dimensions of emotion. *Psychological review*, 61 2, 81-8.
- [29] Izard, C.E. (1977). *Human emotions*. New York: Plenum
- [30] Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124–129. <https://doi.org/10.1037/h0030377>
- [31] Posner J, Russell JA, Peterson BS. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev Psychopathol.* 2005 Summer;17(3):715-34. doi: 10.1017/S0954579405050340. PMID: 16262989; PMCID: PMC2367156.
- [32] Borrajo Millán D. et al. *Aprendizaje Automático*. SANZ Y TORRES, 2006. <http://www.editorialsanzytorres.com/libros/aprendizaje-automatico/9788496094734/>
- [33] Rosenblatt, F., *Principles of Neurodynamics*, Spartan books, New York, 1962

- [34] Ferré, P., Guasch, M., Martínez-García, N. et al. Moved by words: Affective ratings for a set of 2,266 Spanish words in five discrete emotion categories. *Behav Res* 49, 1082–1094 (2017). <https://doi.org/10.3758/s13428-016-0768-3>
- [35] Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441–453. <https://doi.org/10.1177/016555150202800601>
- [36] Larman, Craig. *UML Y Patrones : Una Introducción Al Análisis Y Diseño Orientado a Objetos Y Al Proceso Unificado*. 2ª ed. Madrid: Pearson; Prentice Hall, 2010. Print.

Índice alfabético

A

- activación (*activation*) · 47
- agrupamiento · 31, 55, 62, 103, 106, 107, 178, 179, *Véase* clustering
- Agrupamiento · 106, 178
- algebraico · 37, 49
- alias · 58, 80, 81, 85, 93, 95, 104, 107, 122, 145, 147, 157, 163, 165, 169, 176, 179, 184, 208, 209, 218, 222, 225
- Alias** · 89, 93, 157, 158, 163, 209, 210, 214, 217, 218, 219, 220, 221, 224, 225
- Álvaro Rodrigo · 59, 61, 138
- Análisis · 1, 22, 25, 26, 28, 30, 36, 43, 49, 52, 54, 55, 56, 79, 84, 85, 87, 88, 91, 92, 97, 98, 100, 101, 102, 122, 123, 124, 129, 130, 138, 144, 145, 154, 160, 161, 163, 169, 173, 203, 208, 210, 214
- Anonimización · 58
- aplicación · 2, 3, 25, 26, 27, 28, 41, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 63, 67, 68, 69, 70, 72, 73, 75, 77, 78, 80, 81, 85, 87, 89, 92, 93, 97, 99, 104, 107, 110, 111, 115, 116, 119, 120, 122, 124, 125, 126, 127, 128, 129, 130, 131, 132, 138, 139, 140, 141, 142, 143, 144, 145, 147, 157, 158, 164, 168, 169, 176, 179, 183, 184, 188, 189, 190, 194, 202, 203, 204, 207, 208, 215, 222
- Aprendizaje · 22, 26, 28, 31, 37, 49, 138, 140, 233
- Aprendizaje Automático · 31, 37, 49, 138, 233
- aprendizaje supervisado** · 31, 42
- aspecto** · 36, 54, 57, 58, 68, 74, 77, 186
- attention-rejection* (atención-rechazo) · 44, 45
- Ayuda · 57, 64, 67, 71, 108, 143, 144, 145, 154, 163, 173, 175, 181

B

- BI · 21, 25, *Véase* Business Intelligence
- Business Intelligence** · 21, 25

C

- Casos de Uso · 116
- Centralidad · 80, 81, 82, 105, 110, 147, 177, 183, 220
- Cercanía · 55, 105, 110, 177, 183, 219
- CIS · 21, 29, *Véase* Community Information System

CLA · 21, 29, *Véase* Community Learning Analytics

Clases · 117

Clasificador · 41

clustering · 31, 55, 56, 57, 62, 65, 103, 106, 107, 122, 178, 179, 203, 208, 223, 224, 225

Clustering · 66, 106, 107, 178, 179, 183

CNN · 21, 40, 232, *Véase* Convolutional Neural Networks

combinado · 85, 87, 122, 130, 131, 132, 139, 142, 154, 156, 157, 210, 213

Communities of Practice · 29

Community Information System · 21, 29

Computer-Supported Behavioral Analytics · 21, 28

Computer-Supported Predictive Analytics · 21, 28

Computer-Supported Visualization Analytics · 21, 29

Conclusiones · 138

Configuración · 67, 69, 107, 179

conocimiento del dominio · 29

Convolutional Neural Networks · 40

CoP · 21, 29, *Véase* Communities of Practice

Correspondencia entre el Sentimiento Combinado y su traducción literal (Particiones) · 88

Cronograma · 120

CSBA · 21, 28, *Véase* Computer-Supported Behavioral Analytics

CSLA · 21, 28, *Véase* Computer-Supported Learning Analytics

CSPA · 21, 28, *Véase* Computer Supported Predictive Analytics

csv · 64, 65, 66, 73, 124, 143, 208, 212, 216, 229

CSVA · 21, 29, *Véase* Computer-Supported Visualization Analytics

D

Data Mining · 21, 25, 230, 232

dataset · 31, 42

desarrollo · 24, 36, 51, 60, 76, 92, 111, 113, 120, 121, 122, 123, 138, 154, 161

Desarrollo · 26, 27, 140, 141

Desprecio · 93, 164

Destinatario · 89, 124, 158

Diagrama · 116, 117, 118

diseño · 22, 53, 61, 116, 122, 123

Distribución · 60, 111

DM · 21, 25, *Véase* Data Mining

documento · 36, 37, 42, 43, 65, 99, 108, 181

Dominante · 105, 178, 214

E

EDM · 21, 24, *Véase* Educational Data Mining

Educational Data Mining · 21, 24, 230

Emisor · 89, 93, 95, 157, 158, 163, 165

emociones · 2, 3, 24, 26, 28, 30, 31, 32, 43, 44, 45, 46, 47, 48, 49, 55, 61, 63, 76, 92, 93, 94, 95, 98, 105, 122, 138, 141, 142, 143, 163, 164, 165, 169, 171, 172, 177, 194, 196, 202, 205, 215

Emociones · 30, 43, 44, 66, 92, 93, 94, 95, 96, 99, 144, 163, 165, 166, 167, 184, 203, 208, 214, 216

Enfoque · 37, 42

enojo · 44, 45

entornos virtuales · 59, 122, 138

Entropía · 41

Estructura · 63

F

fase · 76, 120, 121

fechas · 60, 84, 100, 129, 200

felicidad · 43, 44, 45, 98, 215

ficheros · 60, 62, 63, 111

Foro · 89, 95, 124, 159, 165, 208, 212, 216, 229

Frank Rosenblatt · 34

freeze · 60, 111

Función codo · 107, 179, 180

G

Gephi · 59, 124, 125, 126, 127, 142

Grado · 55, 80, 81, 82, 97, 105, 110, 147, 168, 177, 183, 219

Gráficos · 62, 90, 91, 95, 96, 108, 110, 123, 160, 166, 167, 180, 182, 183, 184, 203, 208, 212, 216, 220, 223, 226

I

informes · 58, 59, 62, 103, 122, 139, 142, 144, 184, 202

Iniciativa · 105, 177, 219

Instalación · 60, 111, 115

Inteligencia de Negocio · 24, *Véase* Business Intelligence

interfaz · 2, 26, 27, 51, 53, 61, 67, 68, 69, 71, 77, 139, 141

Interfaz · 51, 54, 67, 78

iteraciones · 31, 32, 120, 121, 122, 138

K

k-means · 31

K-medias · 32, 106, 179, Véase K-means

K-Nearest Neighbors · 33

K-NN · 21, 33, 40, Véase K-Nearest Neighbors

L

LA · 21, 22, 23, 24, 29, 218, Véase Learning Analytics

LDA · 21, 43

Learning Analytics · 1, 3, 5, 21, 22, 23, 24, 28, 29, 122, 208, 209, 210, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230

Linguakit · 59, 129

Long Short Term Memory · 40

LSTM · 21, 40, 232, Véase Long Sort Term Memory

M

Máquinas de Vectores Soporte · 32, 231, Véase SVM y Support Vector Machines

Minería de Datos · 25, 28, Véase Data Mining

Modelado · 119, 144

Modularidad · 84, 105, 177, 219

MOOC · 21, 23

N

N. Bayes · 82, 89, 90, 91, 129, 130, 148, 158, 159, 160, 213, 222, Véase Naïve Bayes

Naïve Bayes · 60, 83, 89, 122, 139, 148, 154, 158, Véase Naïve Bayes

Naïve Bayes · 32, 37, Véase Naïve Bayes

Natural Language Processing · 21, 30, 35, Véase *Procesamiento del Lenguaje Natural*

n-grama · 41

NLP · 21, 30, 35, 221, Véase Natural Language Processing y *Procesamiento del Lenguaje Natural*

Nodos · 80, 84, 145, 146

Normalización · 73, 74, 75

O

oración · 36, 38, 41, 42

Óvalo de Posner de las emociones · 47

P

Page Rank · 105, 151, 177, 219
PDF · 59, 66, 122, 139, 183, 184, 207
Persistencia · 58
peso · 35, 42, 43
pip · 60, 111, 115
pleasantness-unpleasantness (agrado-desagrado) · 44, 45
pLSA · 21, 43
PMI · 41, Véase Point wise Mutual information
Point wise Mutual information · 41
Posner · 47, 233
Procesamiento del Lenguaje Natural · 30, 35
Proceso Unificado · 120, 139
proyecto · 63, 65, 70, 71, 72, 73, 76, 78, 79, 80, 111, 112, 113, 117, 130, 143, 183, 203, 205, 208, 229
pruebas · 31, 62, 76, 86, 122, 123, 154, 213
Pruebas · 124
PyCharm · 59, 111, 112, 113, 114
PySentimiento · 62, 82, 83, 89, 90, 91, 129, 130, 139, 148, 149, 155, 158, 159, 160, 210, 213, 222
Python · 2, 4, 25, 26, 59, 111, 114, 122, 138, 140, 155, 221

R

Rechazo · 93, 164, 170
red · 2, 26, 28, 34, 49, 51, 54, 56, 57, 60, 63, 64, 65, 80, 81, 82, 84, 85, 105, 122, 123, 124, 125, 126, 127, 128, 139, 140, 142, 145, 147, 148, 151, 152, 177, 184, 192, 203, 208, 219, 220, 221, 222
Red · 23, 66, 79, 82, 83, 85, 105, 108, 124, 144, 145, 148, 149, 177, 181, 184
Redes Neuronales · 33
Redes Neuronales Convolucionales · 40, Véase CNN, Convolutional Neural Networks
Redes Sociales · 49
reglas · 36, 42
requirements · 60, 63, 111, 113, 114
Russell · 3, 4, 44, 46, 47, 61, 92, 93, 94, 95, 96, 97, 105, 163, 164, 165, 167, 168, 169, 178, 214, 215, 233

S

Scholberg · 44, 45, 46
SCRUM · 120
Sent. Norm · 86, 154

sentimientos · 2, 3, 24, 26, 28, 30, 31, 32, 36, 37, 38, 40, 42, 43, 48, 49, 51, 52, 53, 55, 56, 57, 59, 60, 61, 63, 76, 82, 89, 91, 92, 105, 115, 122, 129, 130, 138, 139, 141, 142, 143, 148, 154, 158, 161, 177, 192, 194, 202, 205, 212, 213, 216, 222

Sentimientos · 30, 36, 49, 66, 82, 83, 85, 87, 88, 89, 90, 91, 129, 144, 148, 149, 154, 158, 160, 161, 162, 184, 203, 208, 210, 212, 222

SNA · 21, 49, Véase Social Network Analysis

Social Network Analysis · 3, 5, 21, 24, 49

sorpresa · 44, 45

sprints · 120

Support Vector Machines · 32, 232

SVM · 21, 32, 231, Véase Support Vector Machines y Máquinas de Vectores Soporte

T

Tabla · 87, 88, 89, 91, 94, 95, 103, 104, 150, 154, 158, 161, 165, 170, 176

tensión (*arousal*) · 47

TFM · 21, 49, 59, 103, 111, 120, 121, 122, 138, 154

token · 38, 39

tristeza · 44, 98, 215

txt · 60, 63, 64, 65, 66, 73, 79, 93, 111, 113, 114, 143, 164, 215

U

UML · 116, 119

UNED · 2, 3, 4, 5, 61, 63, 68, 138, 143, 208, 209, 210, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229

Usuarios · 66, 98, 108, 163, 170, 180, 184, 203, 208

V

valencia (*valence*) · 47

virtualenv · 59, 111

X

xls · 66, 73, 215

Z

zoom · 81, 82, 139, 149, 200, 222