

# appForum: Una aplicación para el procesamiento de foros

appForum: An application for forum processing

Alvaro Rodrigo<sup>1</sup>, José Luis Fernández-Vindel<sup>1</sup>, Jorge Pérez-Martín<sup>1</sup>, Ismael Iglesias<sup>2</sup>, Víctor Fresno<sup>1</sup>, Aitor Díaz<sup>1</sup>, Francisco Javier Sánchez<sup>2</sup> and Roberto Centeno<sup>1</sup>

<sup>1</sup>Intelligent Systems for Learning, Grupo de Innovación Docente de la UNED

<sup>2</sup>Intecca, UNED

## Abstract

Los foros siguen siendo la forma predominante de comunicación en algunas comunidades y sobre todo en cursos virtuales. Estos foros no solo recogen las dudas de sus usuarios y las consiguientes respuestas, sino que contienen una gran información relativa a su frecuencia de uso, las dinámicas que se generan entre usuarios, etc. Para facilitar su procesamiento y posterior análisis, hemos desarrollado la aplicación *appForum*. Esta aplicación permite transformar a formato tabular los foros que recibe y ofrece distintas vistas y estadísticas sobre la información contenida en dichos foros. Debe servir también a futuro como plataforma sobre la que aplicar algoritmos inteligentes y basados en tecnologías del lenguaje.

**English translation.** Forums are still the main way of communication in some communities and especially in virtual courses. These forums not only collect the doubts of their users and the consequent answers, but they also contain a great deal of information regarding their frequency of use, the dynamics generated between users, etc. To facilitate their processing and subsequent analysis, we have developed the *appForum* application. This application allows the transformation of the forums it receives into a tabular format and offers different views and statistics on the information contained in these forums. It should also serve in the future as a platform on which to apply intelligent algorithms based on language technologies.

## Keywords

Keywords, of, the, paper.

## 1. Introducción

Los foros representan un recurso cada vez más importante para la comunicación entre usuarios en distintas comunidades. Esta importancia se hace todavía más evidente en cursos online, donde los usuarios normalmente no pueden comunicarse en persona y tienen como único recurso de comunicación los foros. Por otro lado, la proliferación de

cursos en línea masivos y abiertos (MOOC), así como la enseñanza a distancia y virtual, que se ha manifestado tan importante durante la pandemia, nos han mostrado la importancia de este tipo de comunicaciones.

En los foros educativos, los distintos participantes (principalmente equipo docente y estudiantes) realizan intervenciones que pueden ser relativas o expresar dudas sobre contenidos o procedimientos académicos, respuestas a las dudas de otros usuarios, aportaciones relacionadas con la asignatura, etc. Todas estas comunicaciones se hacen a través de mensajes de texto y quedan registradas dentro de un curso virtual. Como consecuencia, los foros representan una gran fuente de información para conocer las interacciones entre los distintos usuarios, el seguimiento realizado, las dudas más comunes, etc.

Teniendo en cuenta estos factores, nuestro grupo de innovación docente, *ISL: Intelligent Systems for Learning*, ha desarrollado una aplicación centrada en el procesamiento de foros en el dominio educativo. La aplicación se encarga de procesar foros de una comunidad o asignatura, y generar distintas visualizaciones, estructuradas por campos, lo que permite un mejor análisis de los foros, y con la posibilidad de ser anonimizadas. Además, se realizan distintos procesamientos, como por ejemplo análisis

SEPLN-PD 2022. Annual Conference of the Spanish Association for Natural Language Processing 2022: Projects and Demonstrations, September 21-23, 2022, A Coruña, Spain

✉ alvarory@lsi.uned.es (A. Rodrigo); jlvindel@dia.uned.es (J. L. Fernández-Vindel); jperezmartin@dia.uned.es (J. Pérez-Martín); iiglesias@intecca.uned.es (I. Iglesias); vfresno@lsi.uned.es (V. Fresno); adiazm@scc.uned.es (A. Díaz); fjsanchez@intecca.uned.es (F. J. Sánchez); rcenteno@lsi.uned.es (R. Centeno)

🌐 <http://nlp.uned.es/~alvarory/> (A. Rodrigo); <https://www.uned.es/universidad/docentes/informatica/jorge-perez-martin.html> (J. Pérez-Martín); <http://nlp.uned.es/~vfresno/> (V. Fresno); <http://nlp.uned.es/~rcenteno> (R. Centeno)

📄 0000-0002-6331-4117 (A. Rodrigo); 0000-0002-3588-7233 (J. Pérez-Martín); 0000-0003-4270-2628 (V. Fresno); 0000-0001-9095-4665 (R. Centeno)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

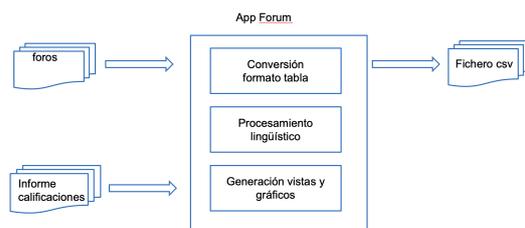


sis de sentimiento y de emociones, que nos pueden permitir realizar análisis más detallados.

El objetivo de esta aplicación es facilitar a los equipos docentes o a las coordinaciones de grado y másterla posibilidad de ejecutar diferentes tipos de analíticas sobre los mensajes en foros de una universidad. Actualmente funciona sobre los foros de la UNED, pero se pueden crear fácilmente adaptadores de los foros de cualquier plataforma. La aplicación se hará accesible a medida que se vayan desarrollando estos adaptadores a las diferentes herramientas de gestión de aprendizajes existentes. *appForum* está desarrollada en Python y utiliza Django para ofrecer sus funcionalidades a través de un interfaz web. A día de hoy la aplicación está siendo utilizada por distintos equipos docentes para analizar los datos de sus asignaturas. La evaluación de la herramienta será, por tanto, cualitativa; más adelante se agregarán todas las conclusiones extraídas por parte de los diferentes equipos docentes.

## 2. Flujo de Procesamiento

En la Figura 1 se muestra el flujo global de procesamiento de nuestra aplicación.



**Figure 1:** Flujo de procesamiento de información de la aplicación *appForum*.

**Conversión a formato tabla:** La aplicación lee los foros en texto plano y los convierte a formato tabular estándar, facilitando después las vistas y los posteriores procesamientos, con campos relativos al NOMBRE DEL FORO, NOMBRE DEL HILO, NÚMERO DE MENSAJE dentro del hilo, MENSAJE AL QUE SE RESPONDE, el AUTOR, la FECHA, el TÍTULO DEL MENSAJE y el propio TEXTO DEL MENSAJE.

**Procesamiento lingüístico:** A continuación, se analizan todos los textos de los mensajes usando la librería *spacy*<sup>1</sup>. El objetivo es disponer de información lingüística (lemas, Entidades Nombradas, qué palabras no son palabras vacías, etc) que se pueda utilizar para crear las distintas vistas. Aplicamos

análisis de sentimiento usando la librería *sentiment-spanish*<sup>2</sup> y de emoción siguiendo el modelo de afecto de Russell [1]. Creamos también representaciones de cada mensaje dentro del Modelo de Espacio Vectorial y con TF-IDF como función de pesado.

**Generación de Vistas y Gráficos:** Posteriormente, y como paso siguiente, se generan cuatro vistas complementarias en relación a cómo se agrupa la información a mostrar: “*por mensajes*”, “*por foros*”, “*por participantes*” y “*por hilos*”. Cada vista contiene información distinta (relacionada con dicha agrupación) y permite generar una serie de gráficos sobre la información mostrada en dicha vista.

Unido a lo anterior, la aplicación ofrece también las siguientes funcionalidades:

**Informe de calificaciones:** Permite cargar las calificaciones de los estudiantes de una asignatura en formato procesable y añadirla a la información de foros.

**Exportación a fichero csv:** Las distintas vistas e información de gráficos se pueden descargar en ficheros csv para su posterior procesamiento. La aplicación permite crear estos ficheros que pueden mostrar la información de foros desde distintas perspectivas, mezcladas con notas, así como con la inclusión de información lingüística.

En la siguiente sección vamos a describir las distintas vistas que ofrece la aplicación.

## 3. Vistas y Gráficos

Como hemos comentado en la sección anterior, *appForum* ofrece cuatro vistas principales en función de cómo agrupa la información contenida en los foros:

### 3.1. Vista de Mensajes

Esta es la vista por defecto, donde se muestran en formato tabla todos los mensajes publicados con su respectiva información. Esta información consta del NOMBRE DEL FORO y NOMBRE DEL HILO, el NÚMERO DE MENSAJE dentro del hilo, MENSAJE AL QUE SE RESPONDE, el AUTOR DEL MENSAJE, la FECHA y HORA DEL MENSAJE y el TÍTULO DEL MENSAJE. Además, muestra el número de caracteres de cada mensaje. Dado que la inclusión de cada mensaje en la vista podría no ser completa por su longitud, para poder ver cada mensaje hay que seleccionar la fila que lo contiene.

En la Figura 2 podemos ver un ejemplo de la vista de mensajes<sup>3</sup>. Nos permite además crear distintos gráficos sobre la información que se muestra.

<sup>1</sup><https://spacy.io/>

<sup>2</sup><https://pypi.org/project/sentiment-analysis-spanish/>

<sup>3</sup>La información del 'Autor' ha sido anonimizada.



### 3.3. Vista de Participantes

La tercera vista se centra en los usuarios que escriben mensajes en el foro, ya que no se dispone de información sobre los visitantes que no escriben. La información agrupada de esta vista permite realizar distintos análisis, como por ejemplo aquellos basados en análisis de redes sociales e interconexión de usuarios. Por ejemplo, a partir de esta vista podemos ver la estructura que representa la red de usuarios en función de los mensajes que publican y cómo se responden e interactúan (ver ejemplo en la Figura 8). Actualmente hemos incluido también medidas relativas a autoridades y hubs, pero en los foros analizados hasta la fecha no han ofrecido resultados relevantes. La Figura 7 ofrece una muestra de cómo se ofrece la información con esta vista.

Autor	Número de mensajes	Mensajes como autor	Mensajes como respondiente	Número de caracteres	Número de hilos
A	1	0	1	18	1
A	1	1	0	557	1
A	5	5	0	3441	5
A	7	4	3	3013	6
A	6	0	6	1402	3
A	2	1	1	959	1
A	2	0	2	583	1
A	48	3	45	23850	41
A	3	3	0	2943	3
A	6	2	4	4285	6

Figure 7: Ejemplo de cómo se muestra la información en la vista de participantes.

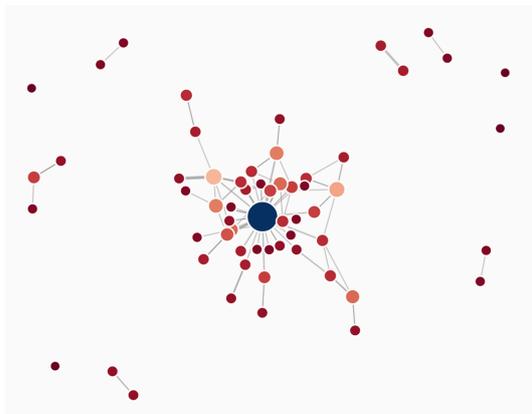


Figure 8: Representación en forma de grafo de las interacciones entre usuarios.

### 3.4. Vista de Hilos

La última vista que ofrece la aplicación agrupa los mensajes por los hilos donde se producen. La información que se ofrece es relativa al NÚMERO DE

AUTORES, MENSAJES y CARACTERES de cada hilo. Permite además seleccionar y mostrar la nube de tags de cada hilo. En la Figura 9 se ve un ejemplo de cómo se muestra la información con esta vista.

Hilo	Número de autores	Número de mensajes	Número de caracteres
Analizador Sintáctico. Etiquetas sin contenido	4	4	1390
ayudas para el examen	2	2	713
Bienvenida	1	1	2440
Bienvenida a la asignatura	2	3	1579
Bienvenida curso 2019/20	2	2	1967
Bienvenida y sesiones presenciales	2	2	3118
Certificación de la práctica	1	1	194
Cambio de fecha sesión presencial obligatoria	1	1	1220
Certificación con el enunciado	2	2	826
Consulta punto 4.8.2	2	2	518

Figure 9: Ejemplo de cómo se muestra la información en la vista de hilos.

## 4. Conclusiones y trabajos futuros

Este trabajo presenta una aplicación, llamada *appForum*, para el análisis de foros en entornos educativos. Esta aplicación permite transformar a formato tabular el contenido de uno o varios foros de una asignatura, ofreciendo distintas vistas, así como estadísticas sobre la información contenida en los mismos. *appForum* integra actualmente diferentes procesos de análisis lingüístico, como un proceso de POS tagging o de análisis de sentimiento y emoción, que se pretenden ampliar incluyendo funciones de composición semántica, algoritmos de generación automática de resúmenes extractivos, de *Community Question&Answering*, detección de toxicidad en mensajes, lenguaje ofensivo, etc. aplicando algoritmos del estado del arte.

## Agradecimientos

Este trabajo ha sido financiado con convocatorias de aplicativos de la UNED, así como por proyectos concedidos al grupo ISL: Intelligent Systems for Learning (GID2016-39) en las convocatorias PID 20/21 y 21/22.

## References

- [1] J. Russell, A circumplex model of affect, *Journal of personality and social psychology* 39 (1980) 1161–1178.