



¿De qué temas hablamos en JENUI?

Modelado de topics con Latent Dirichlet Allocation (LDA)

Raúl Marticorena Sánchez, **Carlos López Nozal**, José Miguel Ramírez-Sanz,
José Luis Garrido-Labrador

Universidad de Burgos - Área de Lenguajes y Sistemas Informáticos

Tabla de contenidos

1. Motivación
2. Trabajos relacionados
3. Método
4. Resultados y discusión
5. Conclusiones

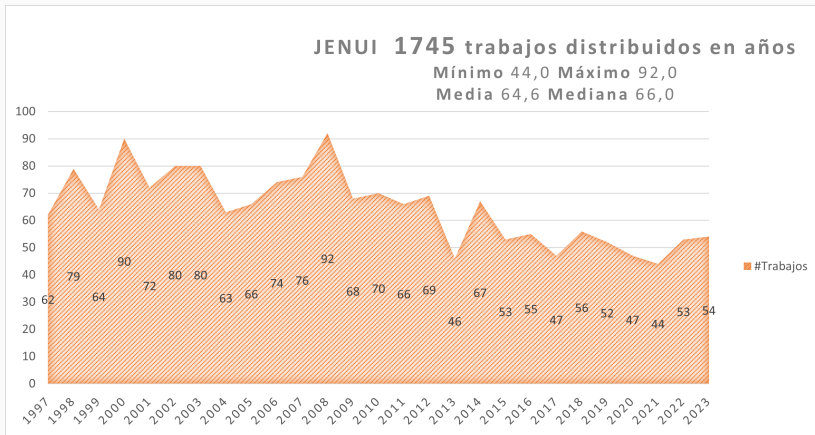
Motivación

Jornadas en Enseñanza Universitaria en Informática (JENUI)

Son un marco de referencia en la enseñanza universitaria de la informática.

- **Origen** - primera edición 1994.
- **Trabajos publicados en actas** - 1745 trabajos.
- **Contexto** - distintos ámbitos o temas de interés de la enseñanza y siempre entroncando con la enseñanza en informática.

Motivación - Contexto



Objetivo

El presente trabajo preprocesa la información sobre todos los **documentos presentados en las 27 ediciones de las JENUI** y aplica el algoritmo de **modelado de temas** conocido como *Latent Dirichlet Allocation* (LDA) [2]. Se analizan los resultados, identificando y analizando los temas, y su **evolución en el tiempo** desde una **perspectiva más objetiva y formal**.

- **PI1:** ¿Cuáles son los **temas principales** en los que podemos clasificar todas las **ponencias de JENUI**?
- **PI2:** ¿Cómo ha sido la distribución de ponencias en dichos temas a lo largo de las **27 ediciones de JENUI**?

Trabajos relacionados

Estudio de actas JENUI

Trabajos que formalizan el estudio de las actas de la JENUI

- **Miró & Alberich [6, 1]** - objetivo analizar el comportamiento de la **red de colaboraciones en JENUI**, y su similitud a las de otras jornadas y congresos de investigación y/o innovación docente.
- **Miró & López [7]** - **análisis bibliométrico**. Entre sus objetivos se establecía *entender mejor las Jenui a partir de lo recogido en sus actas*. Ponen el foco en la autoría y las citas.
- **Cernuda [3]** - explica el **proceso de digitalización completo de las actas de las JENUI**. *Describe la iniciativa llevada a cabo para preparar una edición integral de las actas de JENUI ... es un corpus digital que facilita explotar al máximo los conocimientos acumulados...*

Modelado de temas

Trabajos que formalizan la técnica de modelado de temas para clasificar artículos científicos.

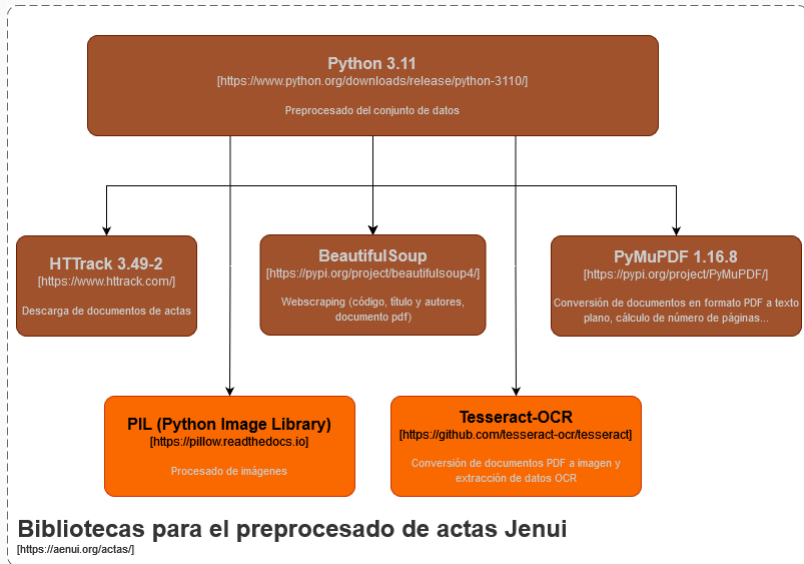
- Kim & Gil [4] - diseñan un sistema de clasificación de trabajos de investigación.
- Liao & Liu [5] - analizan 19.936 trabajos publicados para conocer una **clasificación de temas** de actualidad en la investigación sobre **innovación medioambiental**.

Método

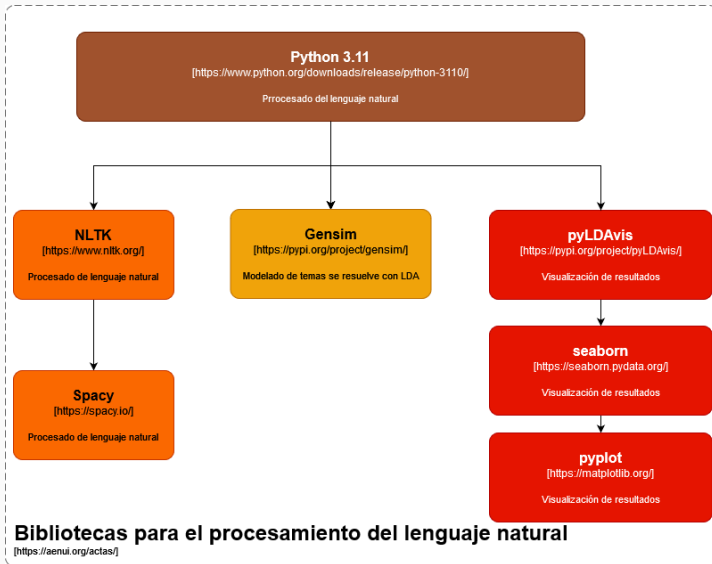
JENUI-AENUI

- Los datos utilizados proceden directamente de las **actas completas**, disponibles en formato electrónico <https://aenui.org/actas/>.
- La **información básica de edición (o año), título, autores y documento** electrónico del trabajo (en formato PDF) están disponibles en dicha web para su descarga y consulta.

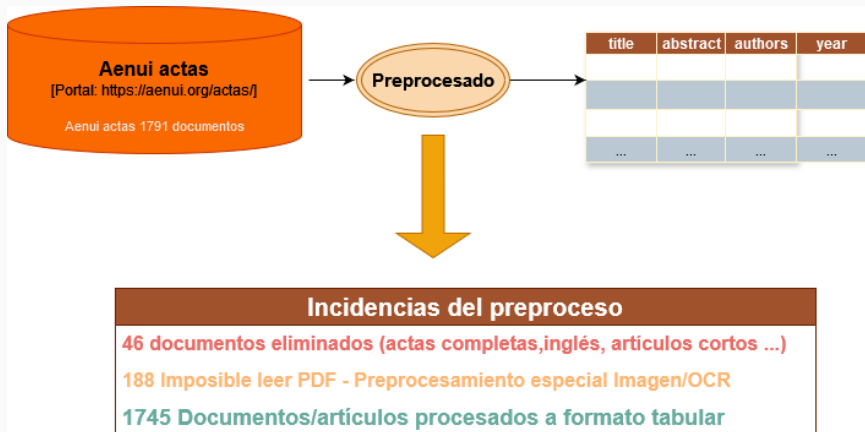
Método - Instrumentos para la preparación de los datos



Método - Instrumentos para procesamiento del lenguaje natural



Método - Procesado de datos



Procesamiento del lenguaje natural

Secuencia variable de **tareas de procesamiento** para la **extracción de palabras de los documentos**.

- Tokenización
- Eliminar caracteres extraños y palabras vacías
- Lematización
- Selección de términos gramaticales - sustantivo, verbo, adjetivo, adverbio.
- Secuencia de palabras que se utilizan juntas - bigram y trigram.
- Filtrado de token por frecuencia - poco frecuentes, en menos de 15 documentos (0,8%) y muy frecuentes en 50% documentos.

Método - Modelado de temas

Modelado de temas o topic modeling

Es una **técnica avanzada de recuperación de información** que automáticamente **encuentra los temas** generales de **en un conjunto de documentos de texto**, llamado corpus, sin la necesidad de etiquetas, datos de entrenamiento o taxonomías predefinidas.

- **Algoritmos** - LSI (Latent Semantic Indexing), **LDA (Latent Dirichlet Allocation)** [2] y HDP (Hierarchical Dirichlet Process)
- **Documento** d_i se puede modelar como una distribución multinomial θ^{d_i} sobre t temas, y cada **tema** z_k , $k = 1, \dots, t$ se modela como una distribución multinomial ϕ^k sobre el conjunto de palabras W
- **Bolsa de palabras (Bag of Words)** - Booleana, Raw TF, Escala logarítmica TF, IDF (*Inverse Document Frequency*), **TF-IDF**
- **Hiperparametrización** - Repeticiones de la ejecución del algoritmo para maximizar la coherencia (c_v) y conocer los valores óptimos de los parámetros (número de temas).

Resultados y discusión

Resultados - PI1 Temas principales de la JENUI

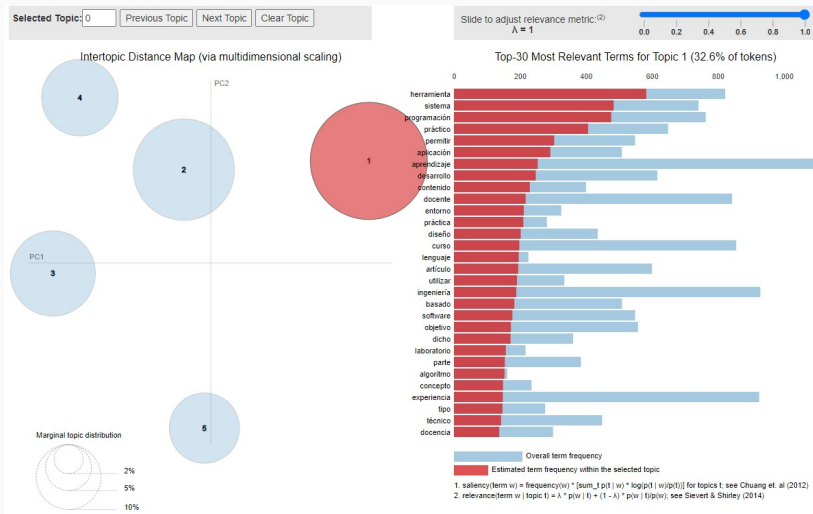


Figure 1: Visualización con pyLDAvis con 5 topics

Resultados - PI1 Temas principales de la JENUI

Tema Probabilidad/término		Etiqueta descriptiva
T1	0.020*herramienta + 0.017*sistema + 0.017*programación + 0.014*práctico + ...	herramienta_sistema_programación
T2	0.020*evaluación + 0.018*experiencia + 0.018*docente + 0.018*aprendizaje + ...	evaluación_experiencia_docente
T3	0.022*competencia + 0.015*grado + 0.015*estudiante + 0.013*programación + ...	competencia_grado_estudiante
T4	0.034*proyecto + 0.020*aprendizaje + 0.015*experiencia + 0.015*actividad + ...	proyecto_aprendizaje_experiencia
T5	0.022*estudio + 0.018*curso + 0.013*ingeniería + 0.011*titulación + ...	estudio_curso_ingeniería

Table 1: Probabilidades y términos asociados por tema.

Resultados - PI2 Distribución de ponencias en 5 temas en el tiempo

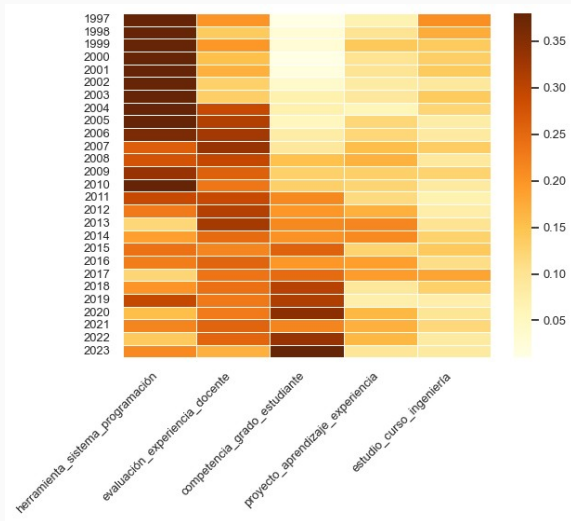


Figure 2: Mapa de calor de probabilidades medias de los documentos por años

Resultados - PI2 Distribución de ponencias en 5 temas en 6 ediciones



(a) JENUI 1997



(b) JENUI 2002



(c) JENUI 2007



(d) JENUI 2012



(e) JENUI 2017



(f) JENUI 2022

Figure 3: Gráficos de radar tomando como año de referencia el inicio de cada lustro entre 1997 a 2022. Leyenda: T1=herramienta_sistema_programación, T2=evaluación_experiencia_docente, T3=competencia_grado_estudiante, T4=proyecto_aprendizaje_experiencia, T5=estudio_curso_ingeniería

Resultados - PI2 Distribución de ponencias en 5 temas en 2023

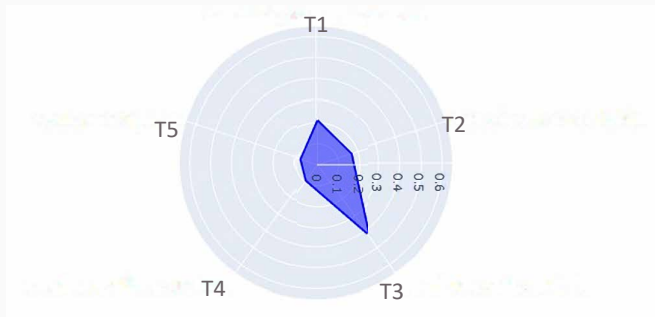


Figure 4: Radar del año 2023. Leyenda:

T1=herramienta_sistema_programación, T2=evaluación_experiencia_docente,
T3=competencia_grado_estudiante, T4=proyecto_aprendizaje_experiencia,
T5=estudio_curso_ingeniería



1997-2003

Se abordan la problemática de la **integración de nuevas tendencias en la programación e ingeniería del software**, el uso de **herramientas docentes innovadoras**.



2004-2011

El tema de principal es **programación y un cierto empuje al tema de evaluación**. Encaja con el foco puesto en la **adaptación de las titulaciones al Espacio Europeo de Educación Superior (EEES)**.



2012-2015

Aumenta el foco en el tema "**competencia_grado_estudiante**" y cuestiones relativas al "**proyecto_aprendizaje_experiencia**". La problemática de la **evaluación continua y del aprendizaje en EEES son fuente de preocupación** y estudio en estos años en particular.



2016-2020

El eje principal es la temática relativa al tema de **"competencia_grado_estudiante"** y cuestiones de **"evaluación_experiencia_docente"**.



2021-2022

Durante la pandemia se afianza los **temas de evaluación y competencias**.



2023

La aparición disruptiva de los **sistemas generativos de texto, como ChatGPT**, se observa un el fortalecimiento de los trabajos orientados a la **adquisición de competencias del estudiante y uso de herramientas**.

Conclusiones

Conclusión general

A lo largo de las **27 ediciones** ha existido una **evolución temporal en 5 temas principales** en los trabajos de las Jenui, transitando de unos temas a otros. **La evolución es explicable con cambios tecnológicos y legislativos** en la educación de la informática.

- Se complementan trabajos de estudios anteriores sobre las JENUI
- Aplicabilidad dependiente del rol
 - **Para los organizadores** - ayuda a la toma de decisiones en la elección de áreas temáticas en las llamadas a la participación.
 - **Para los investigadores** - ayuda a clasificar sus artículos en áreas temáticas consideradas.

Muchas gracias por su atención.



¿De qué temas hablamos en JENUi?

Modelado de topics con Latent Dirichlet Allocation (LDA)

Raúl Marticorena Sánchez, **Carlos López Nozal**, José Miguel Ramírez-Sanz,
José Luis Garrido-Labrador

Universidad de Burgos - Área de Lenguajes y Sistemas Informáticos



R. Alberich and J. Miró.

La colaboración en el JENUI revisited: la convergencia europea.

In Actas de las XV Jornadas de Enseñanza Universitaria de la Informática, JENUI 2009, Barcelona, jul 2009.



D. M. Blei, A. Y. Ng, and M. I. Jordan.

Latent Dirichlet Allocation.

J. Mach. Learn. Res., 3:993–1022, 2003.



A. Cernuda del Río.

La recuperación y edición integral de las actas de JENUI.

In Actas de las XXVIII Jornadas de Enseñanza Universitaria de la Informática, JENUI 2022, pages 23 – 30, La Coruña, jul 2022.



S.-W. Kim and J.-M. Gil.

Research paper classification systems based on TF-IDF and LDA schemes.

Human-Centric Computing and Information Sciences, 9(1), AUG 26 2019.



Z. Liao and X. Liu.

Trending topics and themes in environmental innovation research based on topic modeling.

Sustainable Development, 2023.



J. Miró and R. Alberich.

La colaboración en el Jenui, a quién nos parecemos y a quién no.

In *Actas de las X Jornadas de Enseñanza Universitaria de la Informática, JENUI 2004*, Alicante, jul 2004.



J. Miró and D. López.

Un análisis bibliométrico de las JENUI.

In *Actas de las XVIII Jornadas de Enseñanza Universitaria de la Informática, JENUI 2012*, Ciudad Real, jul 2012.