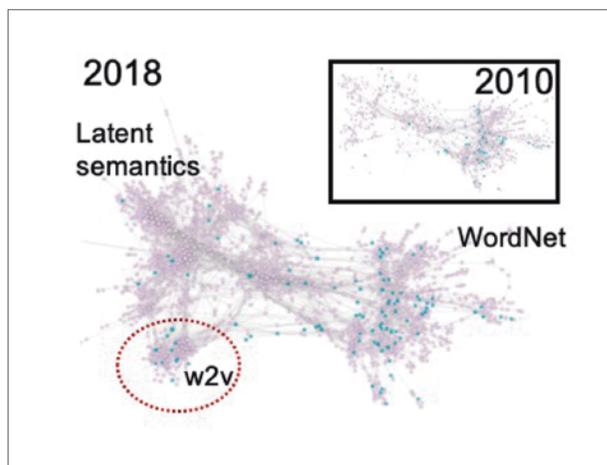


# Técnicas

- PLN incluyendo *WordEmbeddings*.
- Traducción automática español-inglés.
- Modelado de tópicos: estáticos, dinámicos y jerárquicos.
- Búsqueda textual: *LDA+Jensen-Shannon*, *BM25* y *WordEmbeddings*.
- Análisis de grafos, modularidades, distancias entre clusters y centralidad.
- Escalabilidad elástica completa: microservicios (*Rancher*, *Kubernetes*).
- Despliegue automatizado con *Ansible*.
- Componentes en *Dockers*.
- Clasificación automática de documentos con *deep learning* según taxonomías.
- Indexación optimizada para búsquedas eficientes y consultas con filtrado de metadatos.
- Agrupamiento semántico de documentos: emergencia, evolución e hibridación.

VISOR EVOLUCIÓN TEMÁTICA: grafo de identificación de áreas emergentes de I+D+i con navegación interactiva (i.e: publicaciones científicas de PLN de ACL Anthology)



## DEMOSTRADOR ONLINE CORPUS VIEWER

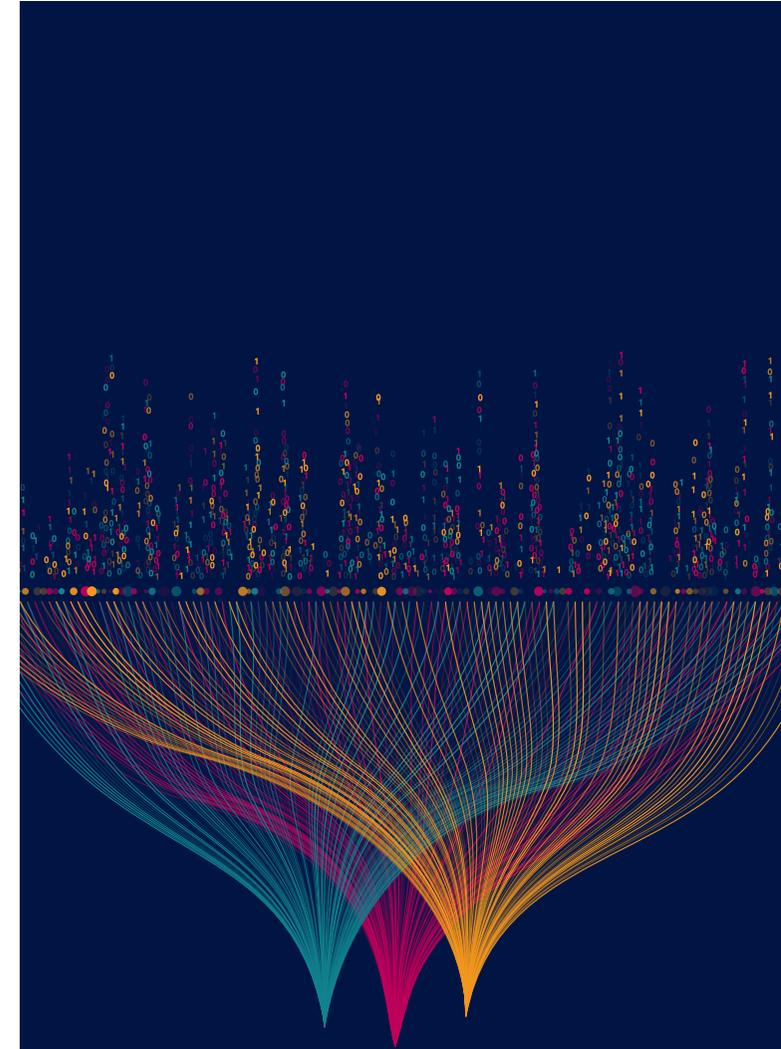
<https://cvdemo.plantl.gob.es/CorpusViewer/>



[www.PlanTL.gob.es](http://www.PlanTL.gob.es)

[plantecnologiaslenguaje@mineco.es](mailto:plantecnologiaslenguaje@mineco.es)

[Plan TL - Tecnologías Lenguaje](#)



PROYECTO FARO  
DE POLÍTICAS DE  
I+D+I

Uno de los ejes del **Plan de impulso de Tecnologías del lenguaje** es su aplicación en sectores estratégicos que actúan como proyectos faro: políticas públicas de I+D+i, Sanidad, Turismo, Justicia, etc.

En el **faro de políticas públicas de I+D+i (Inteligencia Competitiva)** colaboran las **Secretarías de Estado para el Avance Digital y de Universidades, Investigación, Desarrollo e Innovación**, así como la **Fundación Española para la Ciencia y la Tecnología**.

Como resultado de este proyecto se han desarrollado varias **herramientas de análisis de textos que ayudan en el diseño, la gestión, el seguimiento y la evaluación de las políticas de I+D+i**. La mayor parte de estas herramientas se encuentran integradas en la plataforma propia **“Corpus Viewer”** (Visor de corpus documentales).

## USUARIOS

DISEÑO DE POLÍTICAS PÚBLICAS  
(prospectiva y planificación)

GESTIÓN DE CONVOCATORIAS  
(clasificación de solicitudes, asignación de evaluadores, similitud de documentos, estimación de la innovación, evaluación y selección de solicitudes)

SEGUIMIENTO  
(seguimiento de resultados de la intervención y medición de su impacto)

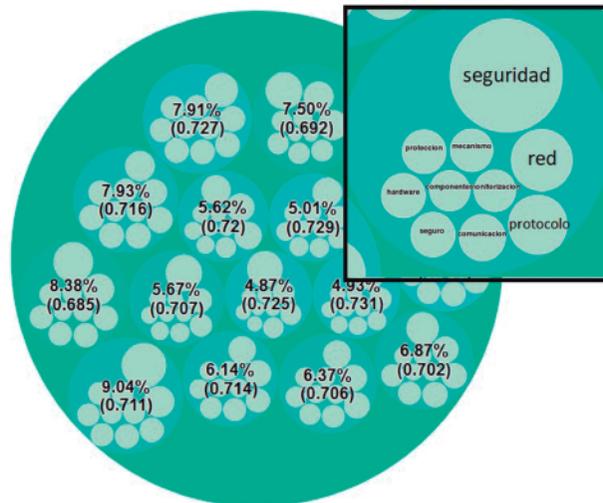
CONTROL  
(alarmas para detección de plagios y patrones de fraude)

# Plataforma Corpus Viewer

Esta plataforma se basa en **tecnologías del lenguaje** (PLN y traducción automática) y aprendizaje automático para analizar metadatos estructurados y **datos textuales no estructurados** de grandes colecciones de documentos.

Se emplea en el **diseño de políticas**, como la Estrategia de Inteligencia Artificial o la definición de misiones en Horizonte Europa, así como en la **evaluación de ayudas** otorgadas por el Ministerio de Economía y Empresa y el de Ciencia, Innovación y Universidades.

VISOR DOCUMENTAL POR TEMÁTICAS: Vista recursiva para navegar por grandes volúmenes de documentación.



## Fuentes de datos

Metadatos estructurados y datos textuales no estructurados de grandes colecciones de documentos: proyectos de I+D, publicaciones científicas y patentes.

# Utilidades

- **Identificación de áreas temáticas** a partir del texto, así como su emergencia, evolución e hibridación.
- **Perfilado** de investigadores y organizaciones, así como detección de redes de colaboración.
- **Clasificación automática** de solicitudes por área temática de gestión y evaluación.
- **Recuperación de documentos** relevantes según similitud.
- **Asociación automática** de las ayudas con sus resultados.
- **Detección de duplicados**.
- **Interfaz tipo BI**: distribución del I+D+i según áreas temáticas detectadas y metadatos disponibles, incluida área geográfica.

