



PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE

Análisis de políticas públicas

- Definición del **Universo** objeto de estudio (ej. IA). Comparativa con respecto al total. Cuantificar geográficamente, temporalmente, otros metadatos.
- Integración **múltiples corpus** en un mismo espacio (I+D: al menos, Publicaciones + Patentes + Ayudas)
- **Subáreas** (agrupaciones semánticas, tópicos) en el Universo. Diferenciación técnicas, funcionalidades, dominios de aplicación
- **Análisis temporal**. Emergencia/desaparición de áreas temáticas (tópicos). Relaciones entre áreas temáticas consolidadas (TRLs). Fenómenos de Lead-Lag entre distintos corpus.

- Principales **agentes** de cada subárea y sus cifras (GEO + TEMP)
- Caracterización de Agentes (investigadores, startups, promotores públicos I+D, etc.) a partir de su producción/financiación. **Perfilado** y agrupación de perfiles
- **Redes de colaboración** entre agentes. Descomposición de proyectos en perfiles colaboradores.

3

Dimensiones: geográfica, temporal, edad, sexo, etc.

Medidas: importe proyectos, impacto publicaciones, etc.

- Proyectos/Patentes/Publicaciones **similares**
- Análisis del **estado del arte**
- Selección de **evaluadores**
- Perfilado de **agentes investigadores/innovadores**

- **Corpus** completos, ej. Concesiones Plan Nacional I+D
Nuevos corpus: empleo, contratación pública
- Conjunto mínimo común de **metadatos** (ej. NIH Reporter).
Enlace entre distintos corpus, desambigüación de autores.
- Compatibilidad con **bases información** bien establecidas;
PatStat EPO, Cordis, NSF/NIH)
- Más **texto**. Ej. Entregables, memorias técnicas, informes finales.
- **Traducción automática a inglés** adaptada a dominio (NMT
basado en corpus paralelos de especialidad. ej. Abstracts
proyectos, patentes, etc.)

Ej. selección universo IA:

WIPO

https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf

https://www.wipo.int/export/sites/www/tech_trends/en/artificial_intelligence/docs/techtrends_ai_methodology.pdf

OCDE

Patentes, ayudas.

ELSEVIER

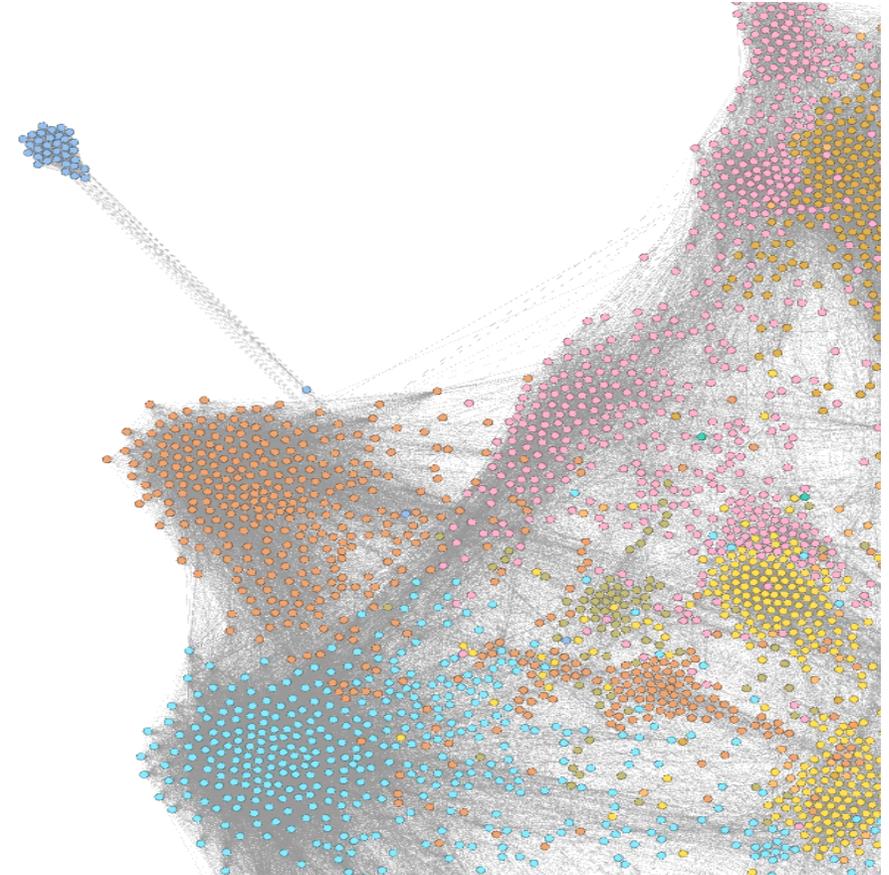
<https://www.elsevier.com/connect/resource-center/artificial-intelligence#research-clusters>

<https://www.elsevier.com/connect/resource-center/artificial-intelligence>

6

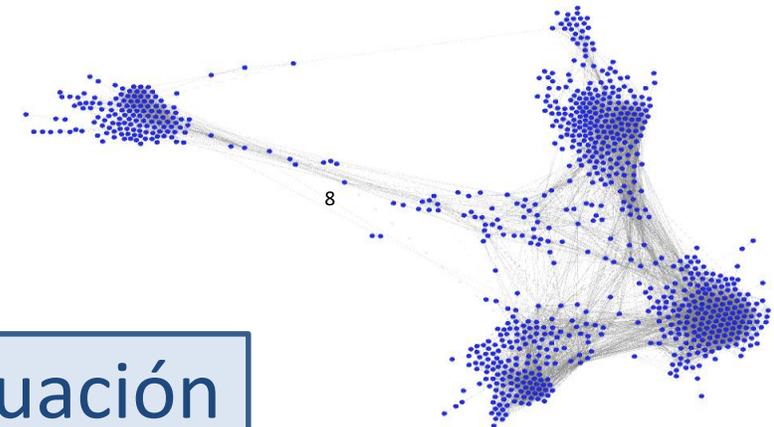
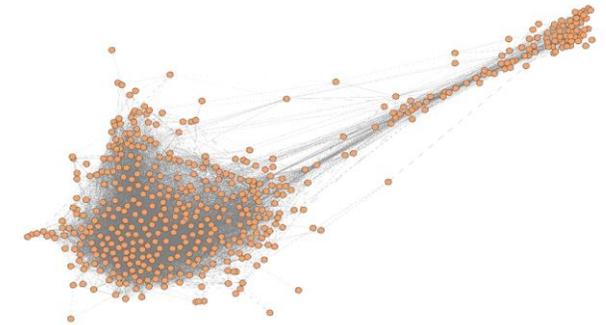
Permite compartir modelos y comparar datos nacionales e internacionales.

- **Multicorpus.** Adaptación stopwords y tópicos vacíos del dominio (citas papers-patentes)
- Procesado **NLP** (PoS, NERC, sinónimos, ...)
- **Distancias** interdocumentales. Selección de métrica. Parámetros modelos y selección de representación documentos. (citas, cocitas)



- **Metadato** automático (clasificadores)
- **Agrupaciones semánticas**
Selección algoritmo y parámetros.
(dispersión metadatos clasificación)
- **Relaciones** entre
agrupaciones semánticas.
Metagrafos, TRLs
- **Perfilado** de agentes

Validación científica = Campañas de evaluación



- Integración **pilotos**
 - Nuevos corpus (contratación pública, empleo, GitHub, ...)
 - Desambigüación
 - Perfilado agentes
 - Nuevos modelos de análisis temporal
 - Redes de colaboración
 - Clasificadores automáticos
- Visualización básica de **grafos** semánticos

- Modelo distribuido: **federación** de servicios
 - Niveles de acceso a la información, control de acceso (metadatos, texto, BoW, tópicos ...)
 - Creación distribuida de modelos o explotación de grandes modelos de referencia.
 - Consultas distribuidas, agregación de resultados
 - **Respositorios locales, modelos globales** compartidos
 - Integración métricas múltiples, dependiendo dominio/problema

- Esquema información:
 - Integración múltiples corpus, no silos de información (CRIS)
 - Información tópicos, texto y metadatos integrada
 - Creación áreas temáticas (IA ...) por el usuario. Validación por expertos.
- Plataforma
 - Integración HPC/GPU en el cálculo de determinados modelos (plataforma NLP, grafos)
 - Integración de motores de traducción (eng, lenguas cooficiales)

- Técnicas:
 - Separación **técnicas, funcionalidades, dominios** de aplicación (otros niveles de análisis NLP, tópicos mayor precisión, ...)
 - Lead-Lag. Grafo completo **temporal**
 - Detección temprana de áreas emergentes