

PERFILADO DE AGENTES DE I+D+I

PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE

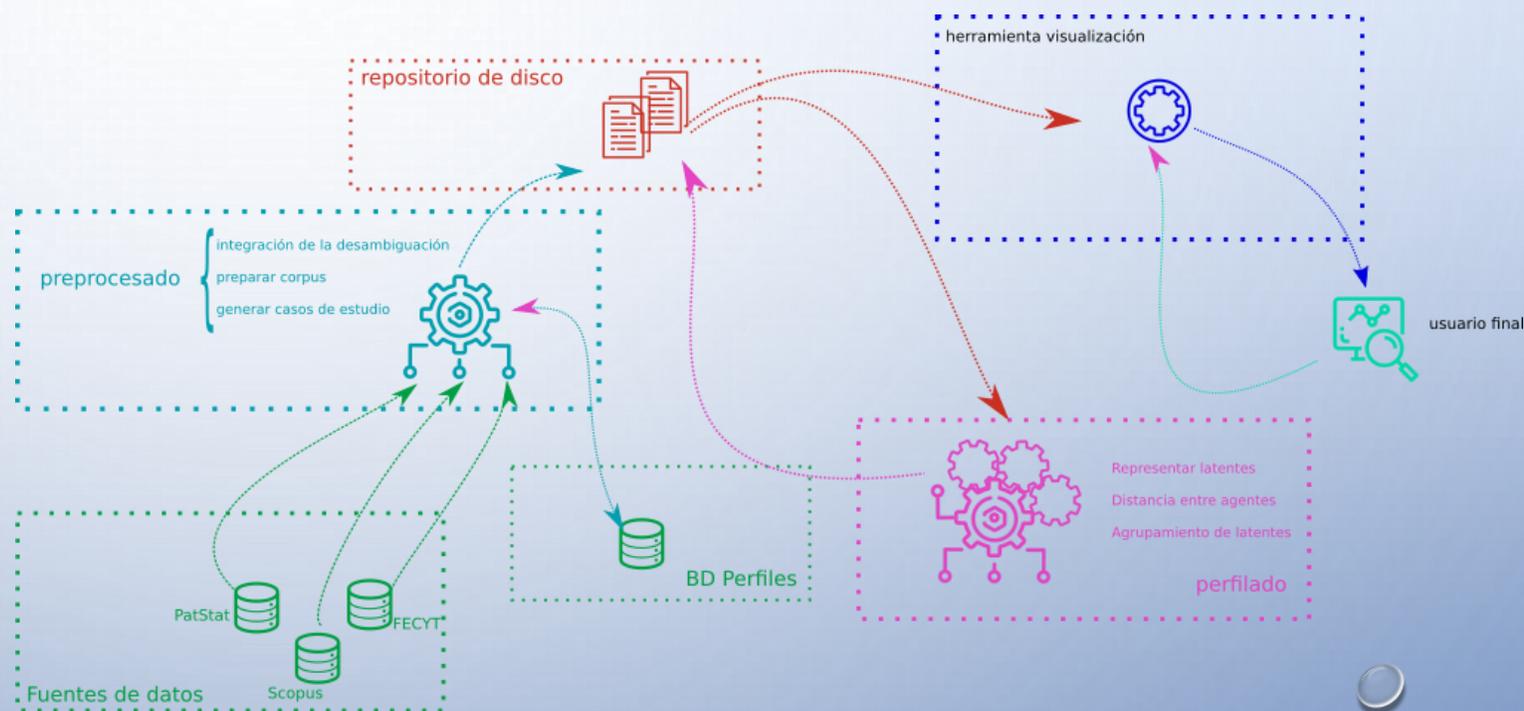
Madrid, 26 de junio de 2019

INTRODUCCIÓN

- **OBJETIVO:** obtener un perfilado de los diferentes agentes dentro del sector I+D+i mediante el análisis y caracterización de su actividad y productividad I+D+i.
- Los **agentes** serán los investigadores.
- La **actividad** de I+D+i que caracterizan la productividad científica de cada agente. Se extraen de tres fuentes:
 - Proyectos
 - Publicaciones científicas
 - Patentes

ARQUITECTURA DEL SISTEMA

1. Desambiguación de agentes
2. Representación vectorial de agentes
3. Perfilado de agentes
4. Herramienta de visualización: facilita el análisis de los perfiles de investigación



PREPROCESADO Y DESAMBIGUACIÓN

- **Unificar** tablas de las diferentes bases de datos → BD Perfiles

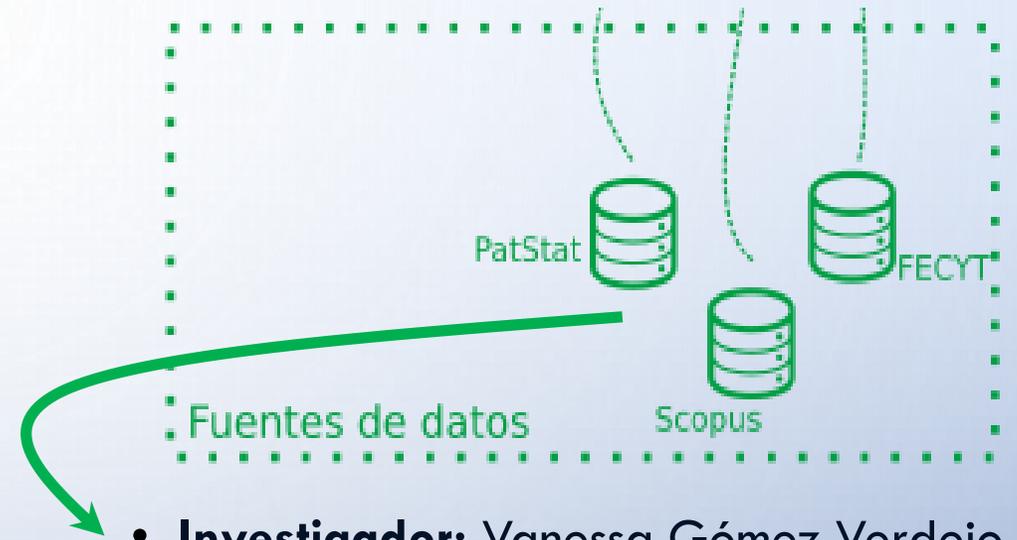
- Selección agentes bajo estudio
- Integrar desambiguación entre agentes
- Homogeneizar tablas

- Filtrar las actividades a una determinada temática (AI)

- Eliminar usuarios con pocas actividades

- SALIDA → Caso de estudio

- **Relaciones agente-actividad:** indica para cada agente las actividades en las que ha participado.
- **Relaciones entre las actividades:** informa de las relaciones entre las diferentes actividades a partir de un modelado de tópicos de los documentos y el cálculo de distancias entre ellos.
- **Metadatos** de utilidad para la visualización



- Patentes: V. Gómez
- Publicaciones:
 - Vanessa Gómez-Verdejo
 - V. Gomez-Verdejo, V. Gmez-Verdejo
- Proyectos: Vanessa Gómez Verdejo

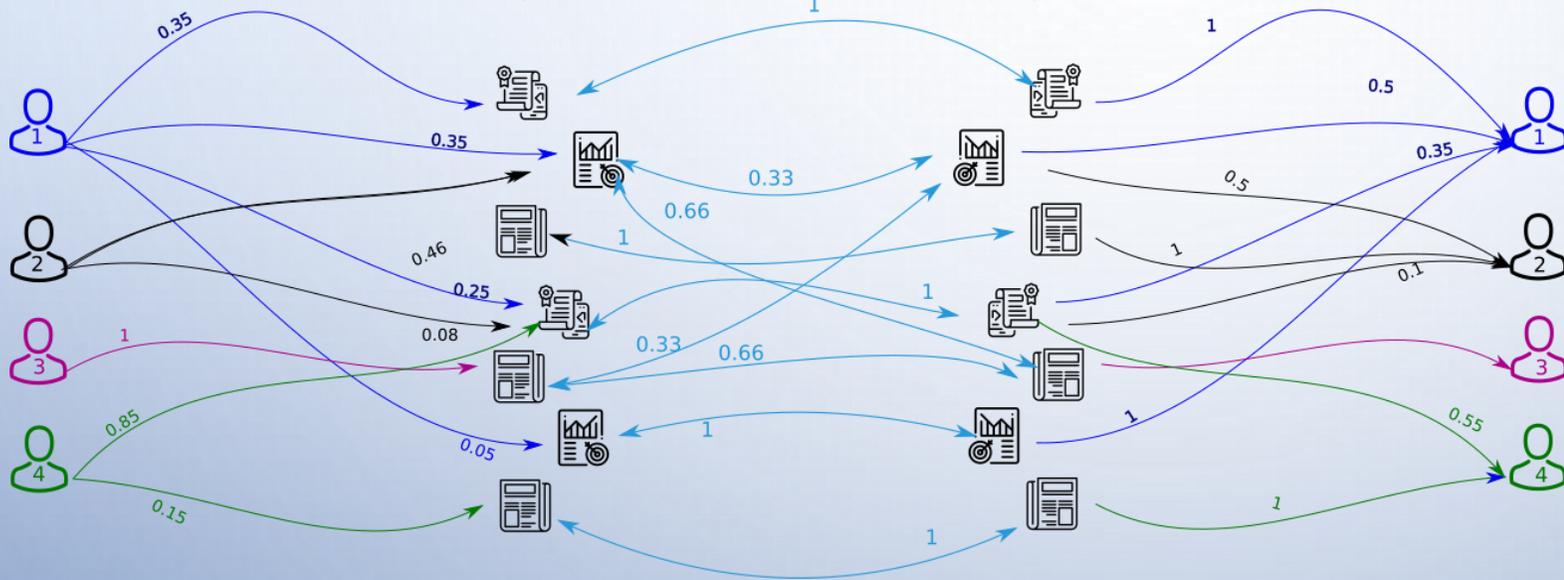
REPRESENTACIÓN VECTORIAL DE AGENTES

OBJETIVO: obtener una representación vectorial o latente de los agentes mediante adaptación de técnicas de word-embedding y deep learning.

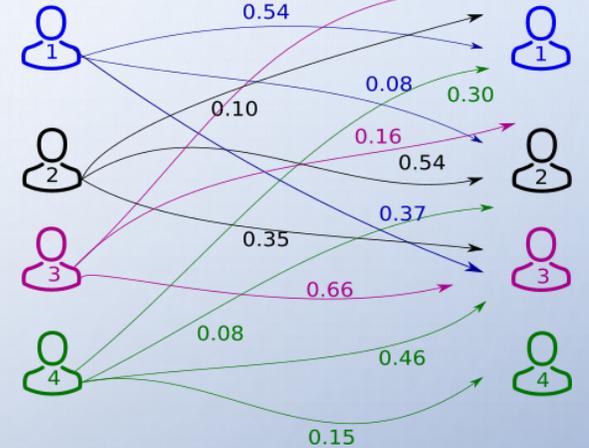
1. Generación de secuencias de interacción entre agentes
 1. Coautorías de artículos, software, participación conjunta en proyectos de investigación
 2. Citaciones de las mismas referencias (artículos, patentes)
 3. ... o, incluso, por participación en actividades de investigación con temáticas similares
2. Entrenar una red neuronal capaz de modelar estas interacciones
3. Usar la representación vectorial de su capa intermedia como una representación de los agentes
4. Identificación de perfiles: Agentes con representación vectorial parecida

GENERACIÓN DE SECUENCIAS DE AGENTES

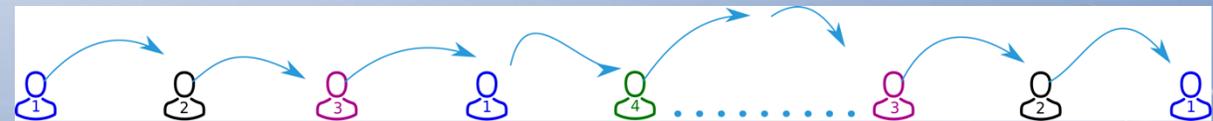
Agentes -> Actividades **×** Similitudes temáticas entre actividades **×** Actividades -> Agentes



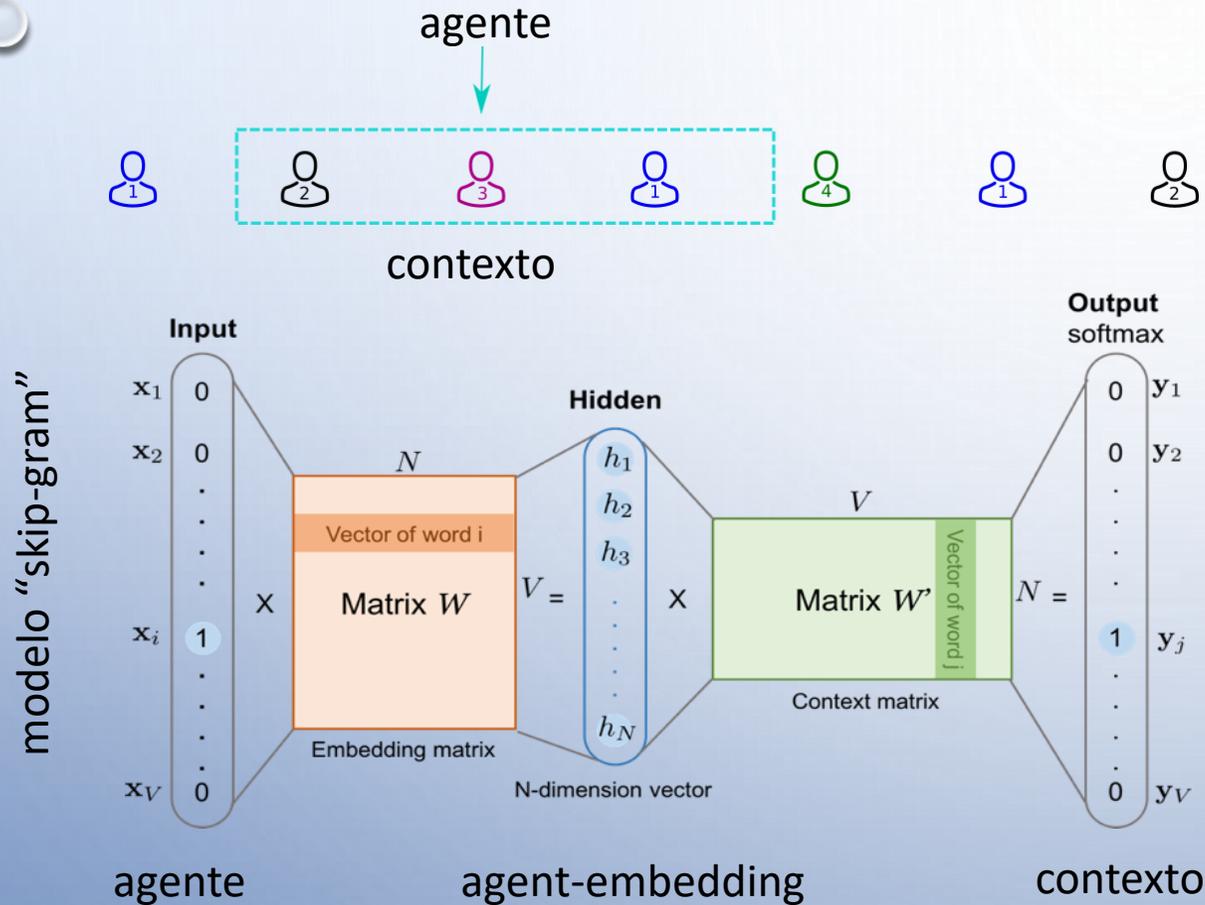
→ Agentes -> Agentes



Secuencia de agentes

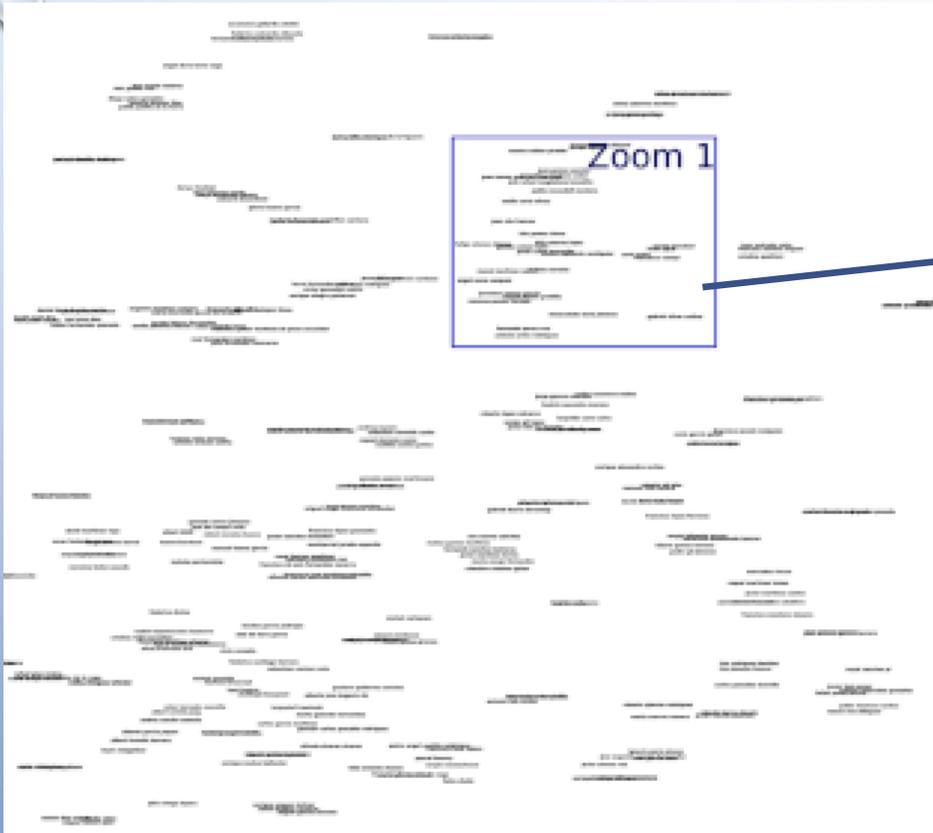


OBTENCIÓN DE AGENT-EMBEDDING



- Se entrena una red neuronal para predecir el "contexto" de cada agente
- Su capa intermedia proporciona una representación vectorial de cada agente

VENTAJAS DEL AGENT-EMBEDDING



- Agent-embedding permite:
 - Medir distancias entre agentes
 - Agrupar agentes -> perfiles de investigación
 - Visualización con una proyección en 2 dimensiones (t-SNE)

AGRUPAMIENTO DE AGENTES: PERFILES DE INVESTIGACIÓN

- **OBJETIVO:** el objetivo es detectar a los grupos de investigadores con intereses y productividad similares y asignarlos en grupos → **perfiles de investigación.**
- **PERFILADO A PARTIR DEL AGENT-EMBEDDING**
 - Se trata de agrupar agentes con representaciones similares
 - Se han evaluado varios algoritmos (K-means, Louvain, Fastgreedy, Walktrap ...).

HERRAMIENTA DE VISUALIZACIÓN

- **OBJETIVO:** cualquier usuario, accediendo desde un navegador web, sea capaz de analizar de manera interactiva los perfiles de investigación obtenidos.
- La implementación de la herramienta se ha hecho en JavaScript para facilitar su integración en páginas web, y ha hecho uso de librerías como D3.js, dc.js y sigma.js
- Planificada su integración en Corpus Viewer (finales 2019)