

# Acta de la reunió entre l'equip de Península i l'equip de recerca de HerStory, del 2/7/2025

## Participants

- Equip de Península: Victor Piles (organitzador), Thomas Gómez i Andrés Manso
- Equip de HerStory: Marina Salse i Miquel Centelles

## Desenvolupament i acords

Per part de l'equip de Península, s'hi incorpora el desenvolupador Thomas Gómez (thomas.gomez@peninsula.co).

Es revisa el document Sistema General, on s'exposa el sistema d'extracció de dades RDF a partir de formats relacionals, i la generació del sistema d'IAG per a l'explotació del graf de coneixement.

A partir de la revisió, s'incorporen les següents modificacions:

1. En relació a l'apartat "2. Análisis Estructural Previo": s'incorpora SQL com a format d'ingesta i anàlisi estructural juntament amb els dos ja previstos: CSV i RDF.
2. En relació a l'apartat "3. Extracción de Entidades y Relaciones", Península realitzarà dos mapatges a partir de la informació facilitada per l'equip de recerca:
  - 2a Un mapatge entre les regles d'extracció d'entitats i relacions dels fitxers font i les regles de transformació de l'esquema ER a Ontologia.
  - 2b Un mapatge entre les entitats i relacions extretes de forma automàtica, i les classes i propietats de l'ontologia desenvolupada per l'equip de recerca sobre la base de Wikidata. Aquest mapatge tindrà la forma de taula amb tres columnes: una columna amb les entitats i relacions procedents de l'extracció automàtica (origen), una taula amb la correspondència amb classes i propietats de l'ontologia de

HerStory (destinació), i una taula amb noves designacions de les classes i propietats identificades durant la fase d'extracció automàtica.

3. En relació a l'apartat "4. Almacenamiento en Neo4j".

3a L'equip de Península confirma que l'ús del model de graf de propietats Neo4j facilita la conversió en origen a RDF, de forma que no es dificulta (ni es ralentitza) la generació de respostes RDF a consultes i a consum de dades a través d'SPARQL.

3b A través de l'ontologia de HerStory, s'aplicaran restriccions semàntiques (axiomes) en les dades RDF del graf. Per exemple, les restriccions de domini i de rang en les propietats. Cal tenir present que la incorporació de restriccions semàntiques (axiomes) és una tasca específicament desenvolupada per un estudiant de doctorat vinculat al projecte (Matheus Jerevain).

4. En relació a l'apartat "Motor de Consultas". Aquesta valoració va especialment a dreçada a Elena Gómez, i al seu pla de recerca.

L'equip de Península explica com és el workflow de cerca semàntica a partir de la consulta (prompt) d'un usuari. Inclou cinc processos:

1. *Convierte tu pregunta en embedding vectorial*
2. *Busca chunks similares en el índice vectorial*
3. *Identifica entidades relevantes en esos chunks*
4. *Expande el contexto siguiendo relaciones en el grafo*
5. *Genera respuesta usando el contexto expandido*

De forma nativa, el sistema té en compte preguntes anteriors del mateix usuari, en la mateix sessió de consulta, per a ponderar i "personalitzar" la resposta. Aquest efecte es produeix durant el procés número 4. Això és el que permet Google Generative AI (Gemini). La incorporació de dades externes sobre les necessitats i preferències de l'usuari, més enllà de les dades de la sessió, exigirà un desenvolupament (evolutiu) adicional. És important treballar sobre aquest punt en comunicació amb l'equip de Península per tal d'alinear els desenvolupaments.

Finalment, es decideix el següent calendari:

- Divendres, 4/7, l'equip de Península enviarà a Miquel Centelles la revisió dels processos de l'Sprint actual per tal de reajustar-los a les modificacions 1 i 2(a i b).

- Dilluns, 7/7, Miquel Centelles confirmarà, si escau, la revisió dels processos de l'Sprint.