

Metodología para el desarrollo de vocabularios multilingües

Diego Ferreyra

El presente trabajo forma parte de la memoria del trabajo final realizado en el marco del Master II professionnel sur la «Gestion de l'Information et de la Connaissance de la Université Paul Valéry - Montpellier III.

Objetivos del proyecto

Definición del escenario:

Contextos en los que ya se dispone de un vocabulario controlado monolingüe y se desea desarrollar su versión multilingüe.

Objetivo:

Definir una metodología formal para el desarrollo de vocabularios multilingües utilizando como base un vocabulario monolingüe existente.

Tesauros Multilingües

IFLA Guidelines for Multilingual Thesauri (2009)

- 1. Equivalencia exacta:** sinónimos entre lenguas. El término preferido X e Y son semántica y culturalmente equivalentes.
- 2. Equivalencia cercana o inexacta:** quasi-sinónimos entre lenguas (diferenciado por el punto de vista).

Los términos preferidos en X e Y expresan el mismo concepto general, pero el significado de los términos en X e Y tiene diferencias de connotación o apreciación.

Tesauros Multilingües

IFLA Guidelines for Multilingual Thesauri (2009)

- 3. Equivalencia parcial** (quasi-sinónimos entre lenguas, diferenciado por nivel de especificidad). Un término preferido en una lengua tiene un significado más abarcativo o general que el término definido en la otra lengua.
- 4. Equivalencia uno-a-muchos:** Para expresar el significado de un término preferido en la lengua de origen, son necesarios 2 o más términos en la lengua de referencia.
- 5. No equivalencia:** no existen términos preferidos en la lengua de referencia con un significado equivalente al del término preferido de la lengua de origen

Tesauros Multilingües

IFLA Guidelines for Multilingual Thesauri (2009)

Se desarrollará una metodología para implementar vocabularios controlados multilingües simétricos

Son vocabularios en los que cada término preferido de cada lengua de origen puede relacionarse con uno y sólo un término preferido de cada lengua de destino.

Selección del vocabulario controlado

- Tesauro de géneros de ficción elaborado por Mónica Gabriela Pené
 - Liberado bajo licencia creativecommons
 - Tamaño reducido (483 términos): permite una escala de experimentación
 - Géneros literarios: un campo diferenciado dentro de los procesos de clasificación asistida por medios informáticos

Selección de herramienta informática

- TemaTres: servidor de vocabularios controlados
 - Gestión distribuida y multiusuario del vocabulario
 - Soporte para vocabularios multilingües
 - Condiciones de portabilidad de datos: soporte para esquemas de metadatos: BS8723-5, DC, MADS, SKOS-Core, VDEX, XTM, Zthes
 - Gestión federada de vocabularios controlados: se trabajará con editores en distintas lenguas en distintos países.
 - Mecanismos para alerta y sincronización de cambios en los términos: notifica a cada editor de cambios en los otros vocabularios.
 - Control de calidad, mecanismos de auditoría
facilidades de gestión especializadas en gestión de vocabularios controlados

Etapas del proyecto

1. Incorporación del vocabulario en español a TemaTres.
2. Creación de las versiones en inglés, francés e italiano del vocabulario y generación de set de términos correspondientes
3. Generación de las relaciones terminológicas en inglés, francés e italiano.
4. Homologación de relaciones de traducción entre todos los idiomas.

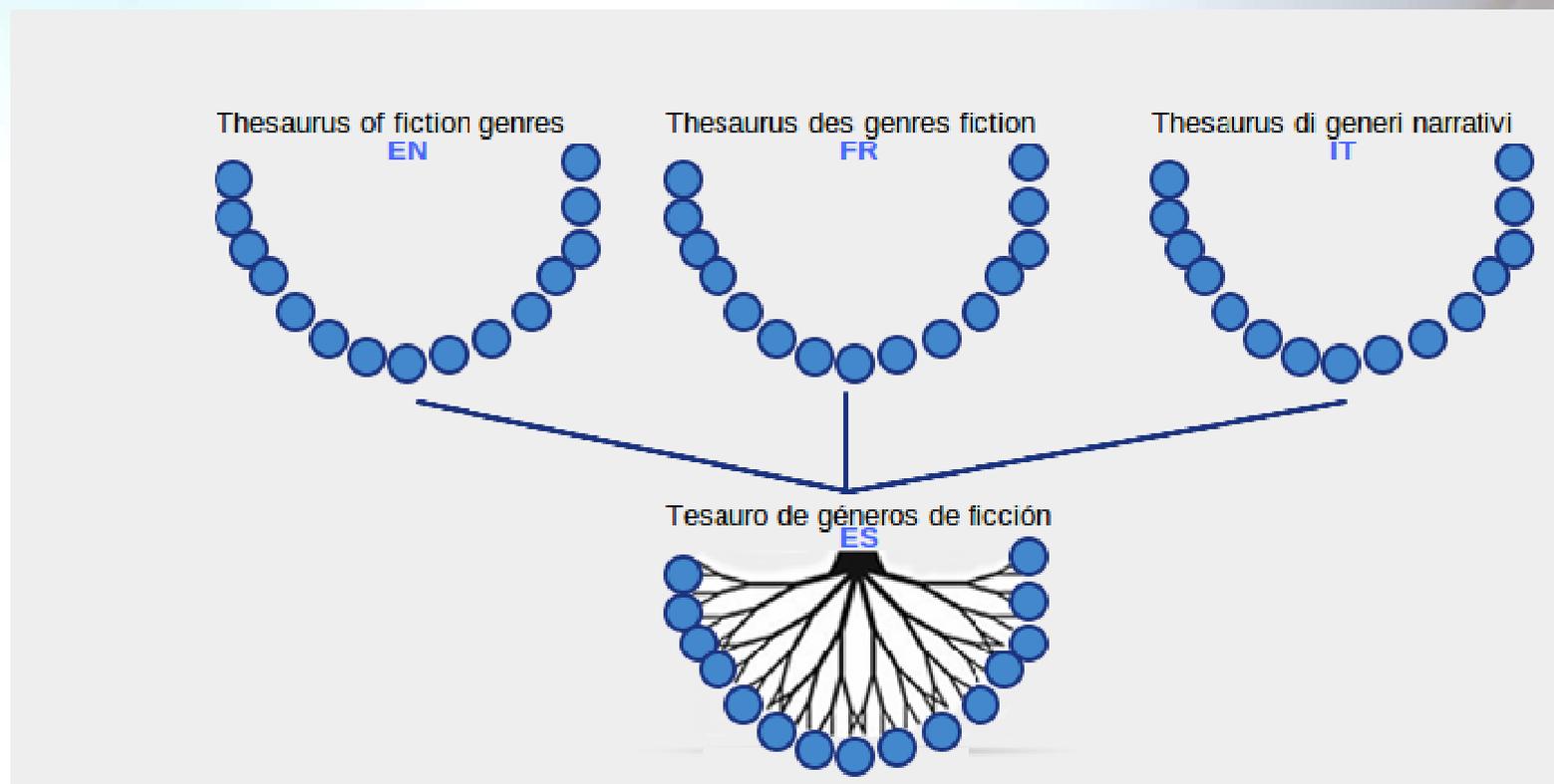
Etapas del proyecto: Incorporación del vocabulario en español a TemaTres.

Migración de los términos y las relaciones terminológicas establecidas en la versión impresa del Tesouro de géneros de ficción



Etapas del proyecto: versiones en inglés, francés e italiano

A partir de la versión inicial en español, se realizó la traducción de todos los términos a los idiomas previstos.

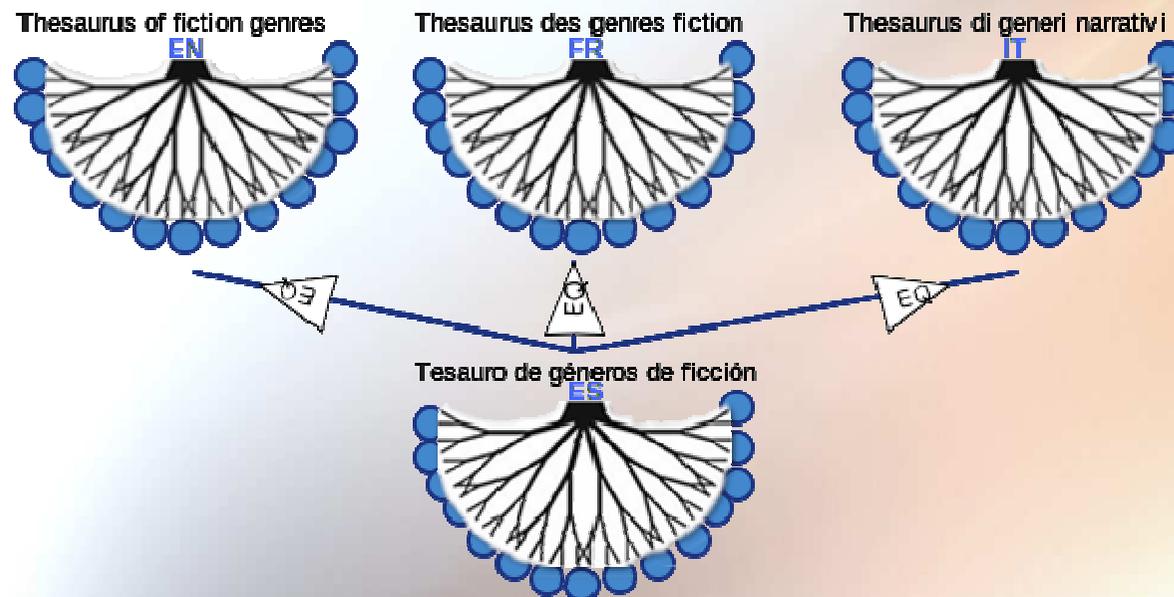


Etapas del proyecto: relaciones terminológicas en inglés, francés e italiano

Algoritmo de replica de relaciones terminológicas:

Si el término A del Vocabulario 1 (Av1) es igual a el término A de V2 (A v2); entonces las relaciones del término Av2 deberían ser las mismas que las del término Av1.

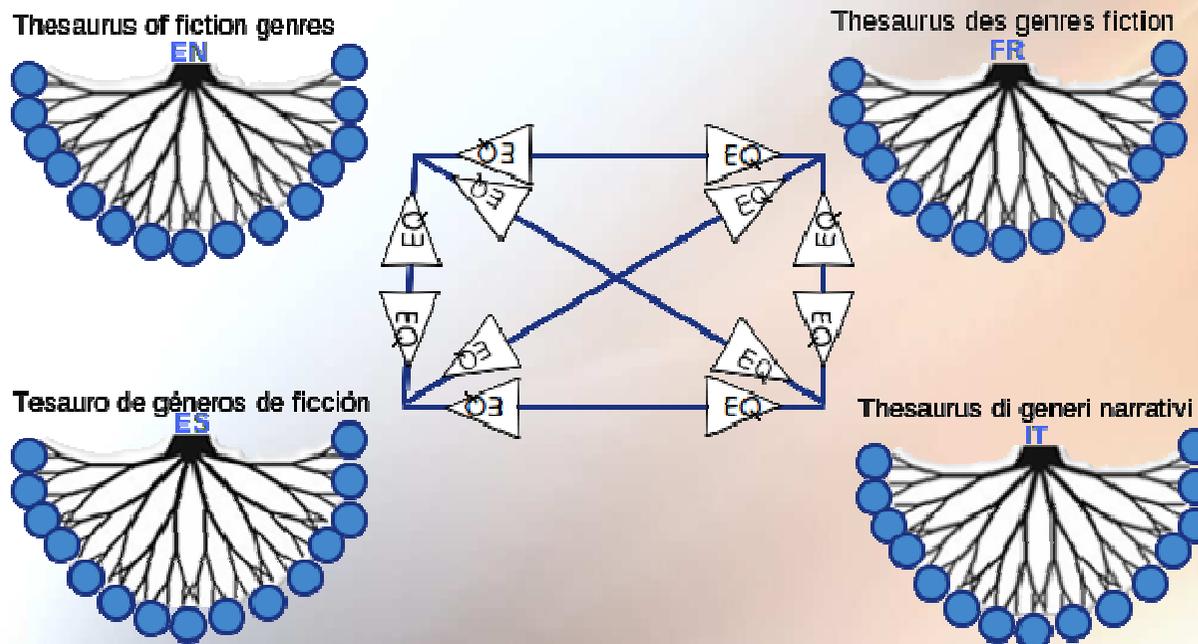
Ejemplo: si en la versión en español **Relato histórico** es un término relacionado (**TR**) de **Relato corto**; entonces en la versión en inglés **Historical relate** será un término relacionado (**TR**) de **Short story**



Etapas del proyecto: Homologación de relaciones de traducción

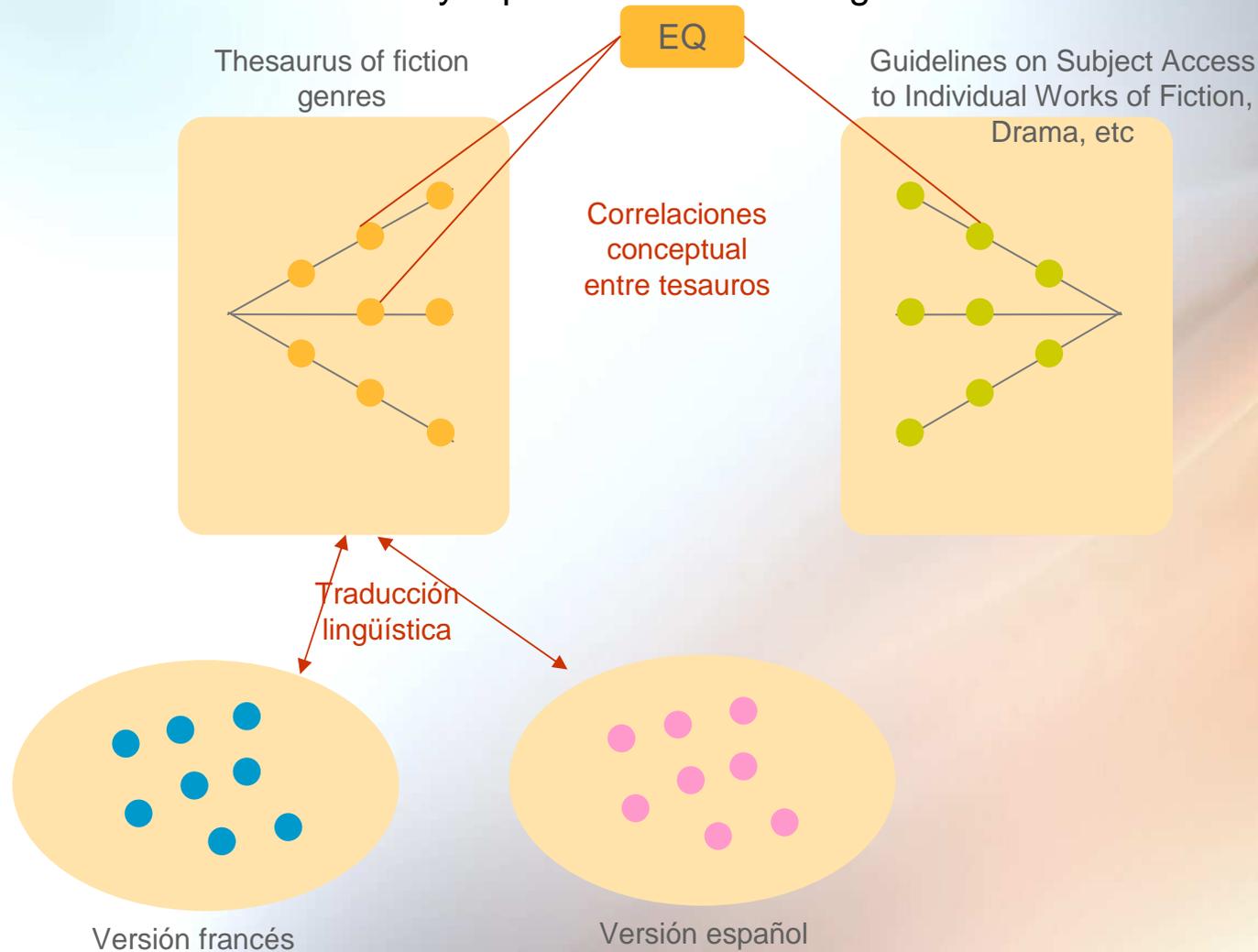
Los vocabularios se referenciaban con respecto al vocabulario controlado inicial en español. En este paso se generan las relaciones directas entre todos los idiomas. El algoritmo para homologar las relaciones de traducción se diseñó según el siguiente esquema lógico:

Si el término A del Vocabulario 1 (Av1) es igual a el término A de V2 (Av2); y el término A del Vocabulario 1 (Av1) es igual a el término A de V3 (Av3); entonces el término Av2 es igual al término Av3. Por ejemplo, la traducción de **Relato histórico** al inglés es **Historical relate**; y la traducción de **Relato histórico** al francés es **Chronique**; entonces la traducción de **Historical relate** al francés será **Chronique**



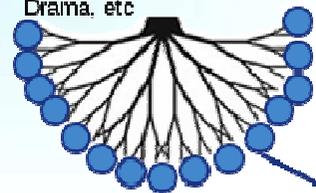
Etapas del proyecto: replica y extensión

Se repitieron los pasos y procedimientos con otro vocabulario para evaluar las condiciones de reuso y réplica de la metodología

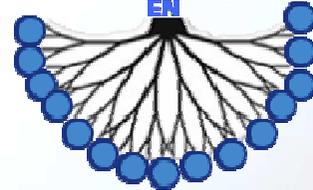


Etapas del proyecto: replica y extensión

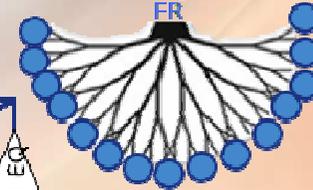
EN
Guidelines on Subject Access
to Individual Works of Fiction.
Drama, etc



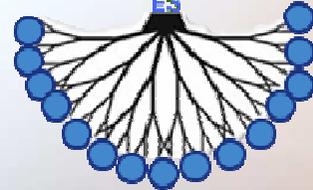
EN
Thesaurus of fiction genres



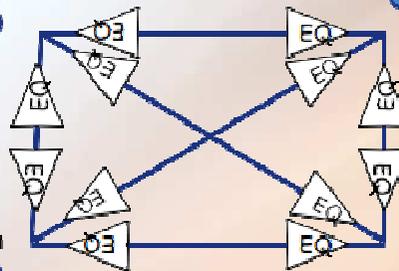
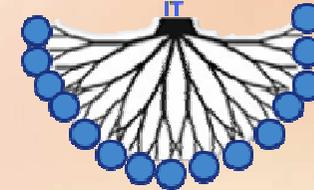
FR
Thesaurus des genres fiction



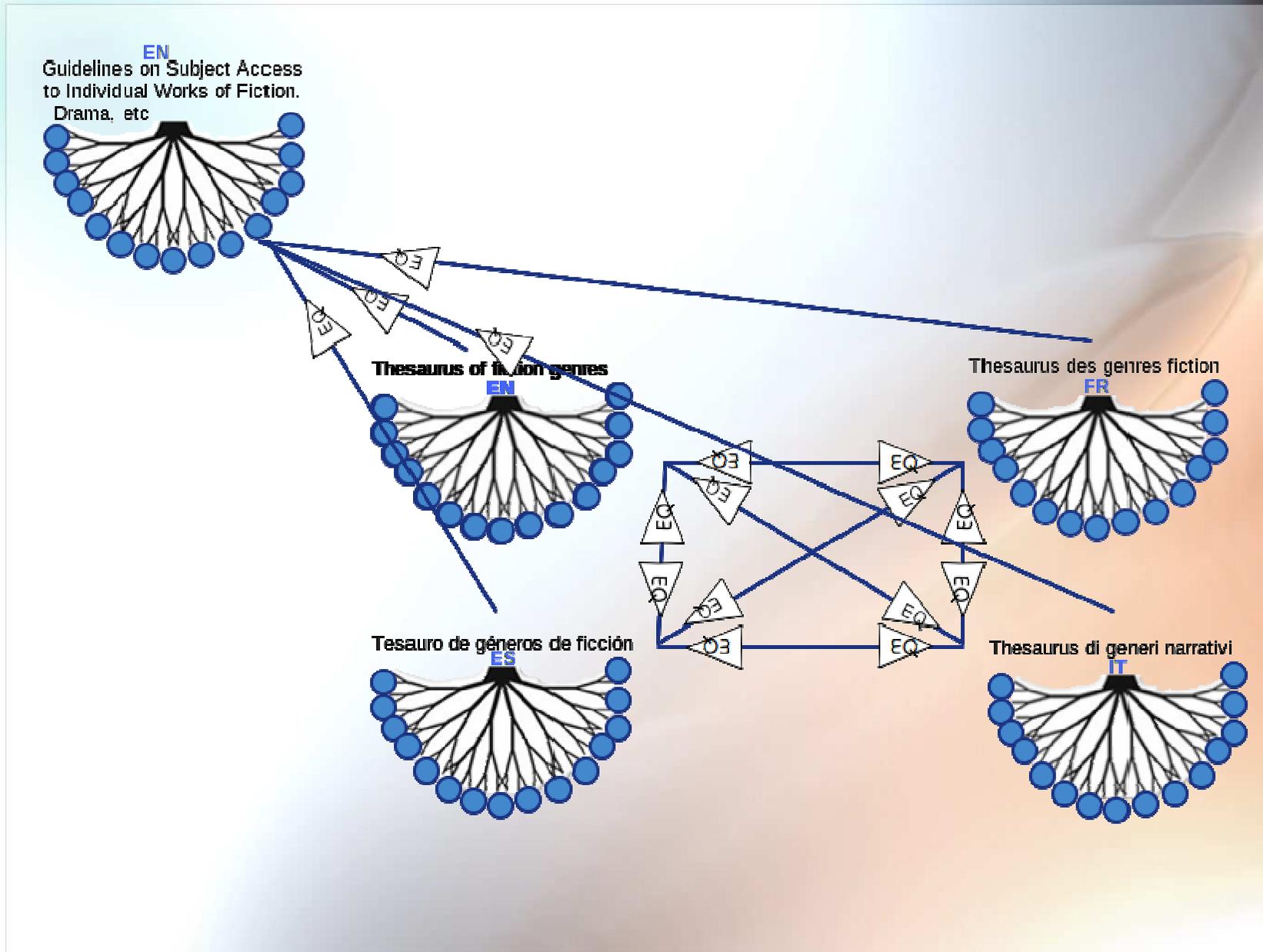
ES
Tesauro de géneros de ficción



IT
Thesaurus di generi narrativi



Etapas del proyecto: replica y extensión



Conclusiones

Resultados alcanzados

Desarrollo e implementación de un tesoro multilingüe de géneros de ficción: <http://www.vocabularyserver.com/ficcion/>

Condiciones de viabilidad de la metodología

A través de metodologías replicables y aplicables a diversos escenarios y en diferentes escalas, fue posible generar un vocabulario multilingüe a partir de un vocabulario inicial en un único idioma.

Al día de hoy esta metodología ha sido aplicada en varios proyectos. La metodología ha podido ser replicada en otros escenarios y por otros profesionales:

Central European Flood Risk Assessment and Management in CENTROPE (CEframe) <http://www.ceframe.eu/thesaurus/>

Conclusiones

Algoritmo de homologación: problemas detectados

Problemas semánticos en el caso de términos que sólo podían ser conceptualizados en el contexto de la lengua de origen y cuyas traducciones no pueden ser contextualizadas en el esquema de significado conformado por la estructura de relaciones terminológicas de la lengua de destino.