

Encyclopedia2GeoKG : un outil pour l'extraction d'informations et la génération de graphes de connaissances géo-historiques à partir d'articles encyclopédiques



Bin Yang, Ludovic Moncla, Fabien Duchateau, Frédérique Laforest (prenom.nom@liris.cnrs.fr)
INSA Lyon, CNRS, UCBL, LIRIS, UMR5205, F-69622 Villeurbanne, France

1. CONTEXTE & MÉTHODOLOGIE

Le projet **ECODA** a pour objectif de proposer une modélisation et une chaîne de traitement pour la construction automatique d'un **graphe de connaissances** géo-historiques à partir d'**articles encyclopédiques**. Le corpus d'étude est l'Encyclopédie de Diderot et d'Alembert (1751-1772).

Modélisation

Une **ontologie spatiale** est proposée pour modéliser les **concepts géographiques** (e.g., ville, pays, rivière, etc.), **les relations** (e.g., inclusion, adjacence, etc.) et **les propriétés spatiales** (e.g., latitude, longitude, surface, etc.).

Extraction d'informations et peuplement

Notre proposition repose sur une chaîne de traitement hybride (ci-contre) composée de différents types de **modèles de langue** (BERT, T5, GPT) fine-tunés pour des tâches de **repérage et classification** d'entités et de relations.

2. CHAÎNE DE TRAITEMENT

Article encyclopédique

Classification article

Classification du type de lieu de la vedette

Repérage des entités géo-sémantiques

Classification des entités nommées de lieux

Extraction des relations spatiales

Graphes de connaissances

ou ce vaie iert a mettre le vin qu'on donne a cha- que plat de matelots. C'est une espee de broc de bois relié de cercles de fer. *Voyez* BROc. (G).

* BIDOUZE, (Géog.) petite rivière du bas Languedoc, qui se jette dans la mer Méditerranée.

* BIDOUZE, (Géog.) rivière de la Gascogne, qui se jette dans la Gave près de Bayonne.

* BIECZ (Géog.), petite ville de Pologne, dans le Palatinat de Cracovie, sur la rivière de Wisłoka; elle est remarquable par ses mines de Vitriol. Long.

Vol. 2, page 242

Géographie / Place

River

Named Entity Recognition

* BIDOUZE HEAD, (Géog. DOMAIN_MARK) rivière NC_SPATIAL de la Gascogne NP_SPATIAL, qui se jette dans la Gave NP_SPATIAL près de Bayonne NP_SPATIAL.

la Gascogne : Region

la Gave : River

Bayonne : City

rivière de la : inclusion

se jette dans : mouvement

près de : distance-orientation

(E0, type, River)

(E1, type, Region)

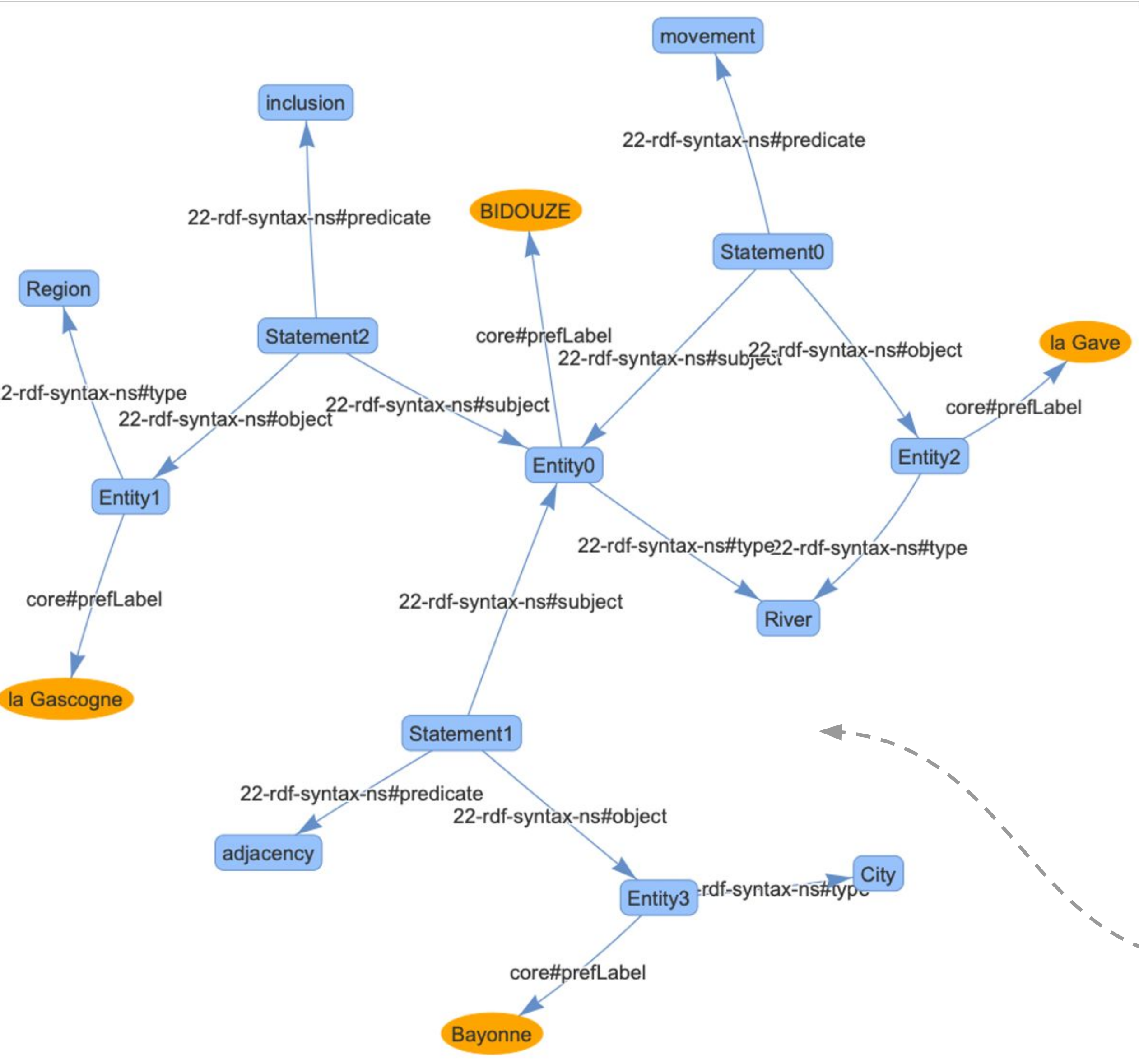
(E2, type, River)

(E3, type, City)

(E1, inclusion, E2)

(E1, movement, E3)

(E1, distance, E4)



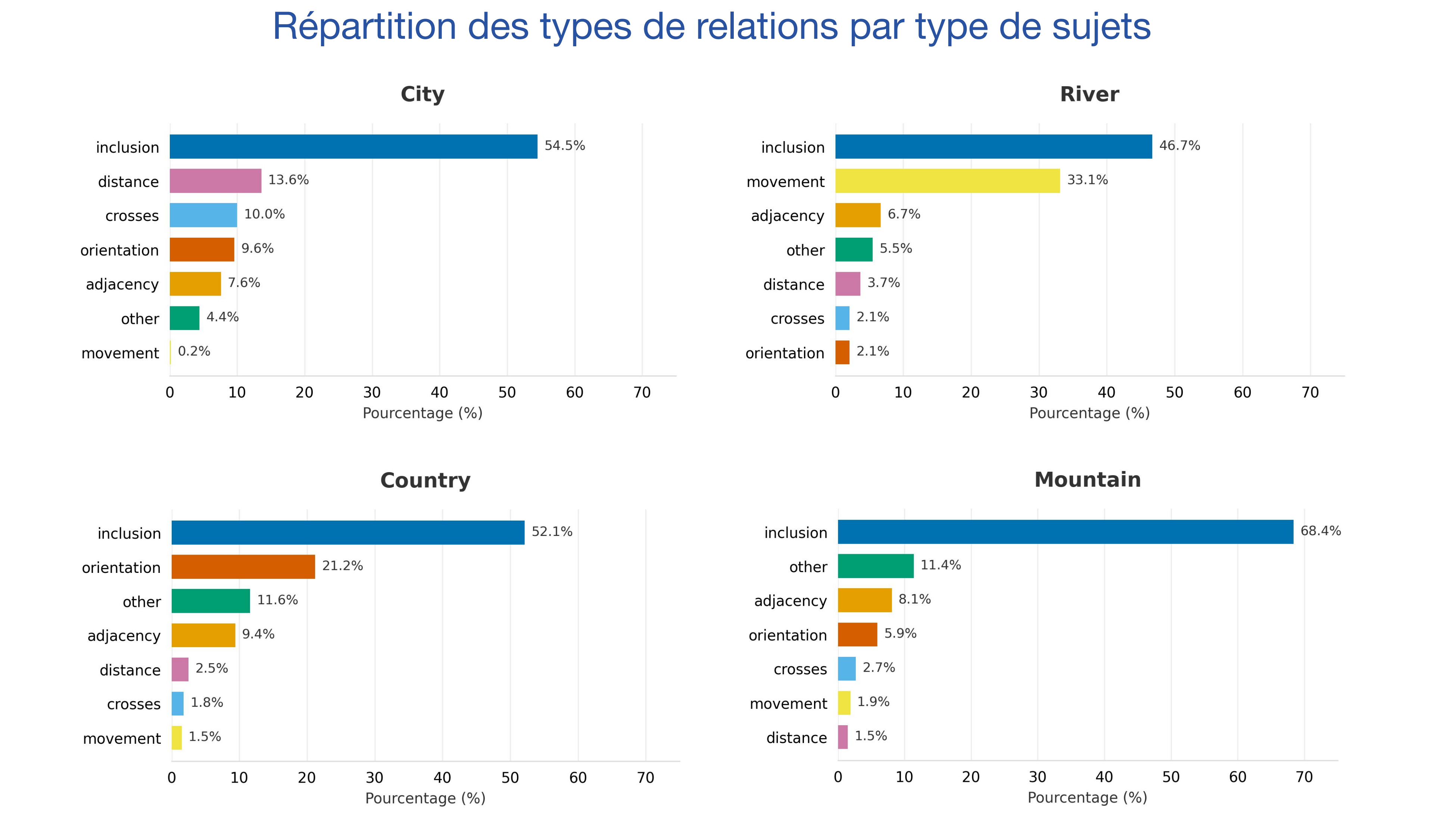
3. RÉSULTATS

Quelques chiffres

- 15 384 articles (domaine Géographie)
- 35 552 entités
- 46 584 relations

Distribution des principales entités

City	17 159	48.3%
Region	5 167	14.5%
River	4 930	13.9%
Island	2 541	7.1%



RESSOURCES

Modèles & jeux de données

Code

Démonstration