

Projet Web sémantique 4-IF-WS (2016–2017)

enrichir et explorer les résultats d'un moteur de recherche

Aimene Belfodil, Sylvie Calabretto, Mehdi Kaytoue

1 Préambule

Les séances prochaines seront consacrées entièrement au projet, à réaliser en hexanome. On propose ici une idée générale de projet, une trame principale. Au début de la séance suivante, chaque groupe devra s'être approprié cette trame. Il est tout à fait possible de s'en écarter et de proposer une variante du projet, ou même un autre projet qui vous tient à cœur. Dans ces derniers deux cas, vous veillerez à valider avec les enseignants le cahier des charges que vous proposerez (utilisation de RDF, SPARQL, ..., données liées, etc.).

2 Contexte

On cherche à proposer une méthode d'exploration des résultats d'un moteur de recherche classique du Web (Google, Bing, Yahoo, Duckduckgo, etc.). Pour une requête donnée, des éléments de la liste retournée peuvent représenter une même ressource, ou encore des choses bien différentes. Dans tous les cas, il s'agit d'une liste ordonnée, et les pages autres que les premières ne sont que rarement consultées. Être capable de caractériser chaque page, avec du sens, lier/regrouper les pages similaires afin de réorganiser ou naviguer à travers les résultats permet de faciliter la tâche de l'utilisateur et enrichit son expérience.

3 Description

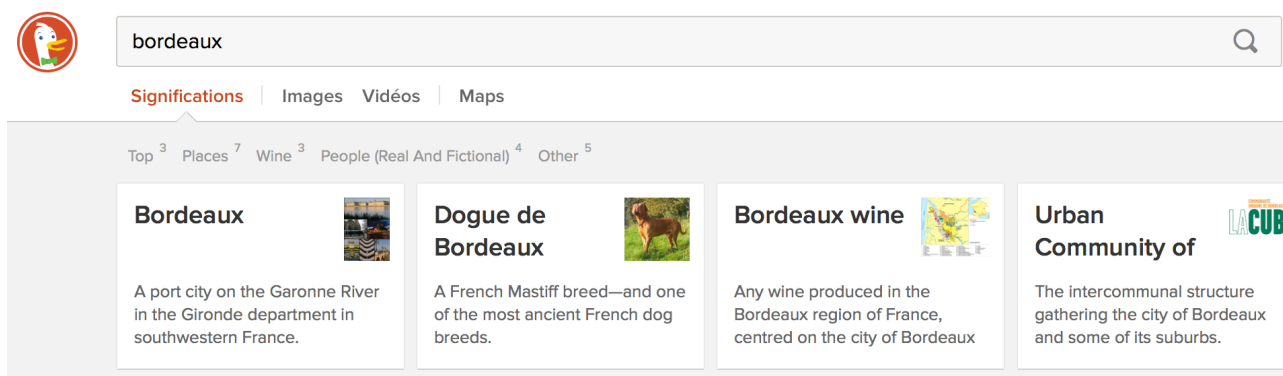
Il vous est alors demandé de retranscrire cette idée à travers une application dont quelques éléments sont proposés ci-dessous.

- Requête sur un moteur de recherche (dans une deuxième temps, mécanismes de guidage, suggestion, auto-complétion, qui utilisent les données liées)
- Stockage des x premiers résultats (page HTML en entier ou en partie de chaque résultat)
- Résolution d'entités dans chacun des textes (e.g. Dbpedia Spotlight)
- Construction du graphe RDF pour chaque résultat
- Relaxation, enrichissement du graphe (selon l'objectif fixé)
- Définir une mesure de similarité (e.g. Jaccard) entre graphes RDF et construire un graphe de similarité (e.g. dynamique, une arête n'apparaît que si la similarité est supérieure à un seuil donné par un composant graphique)
- Exploration et visualisation des résultats
- Et tout autre idée !

On comprendra un exemple en décryptant les figures 1, 2 et 3, ou tout simplement en testant le moteur <http://www.duckduckgo.com>. On utilisera JENA et DBPEDIA comme socle, mais tout autre choix pourra être discuté avec les enseignants.

4 Restitution et évaluation

Les projets seront évalués par groupe de 6 étudiants : vous **présenterez à l'oral** votre travail. Le DS vaut pour 60% de la note finale de l'UE, ce projet pour 40%. Vous veillerez aussi à déposer un **rapport de synthèse** sur moodle ainsi qu'une archive contenant le code source de votre plateforme ainsi que les instructions pour l'utiliser au plus tard une semaine après la dernière séance.



Bordeaux - Wikipedia, the free encyclopedia

Bordeaux (French pronunciation: ; Gascon: Bordèu; Basque: Bordele) is a port city on the Garonne River in the Gironde department in southwestern France.

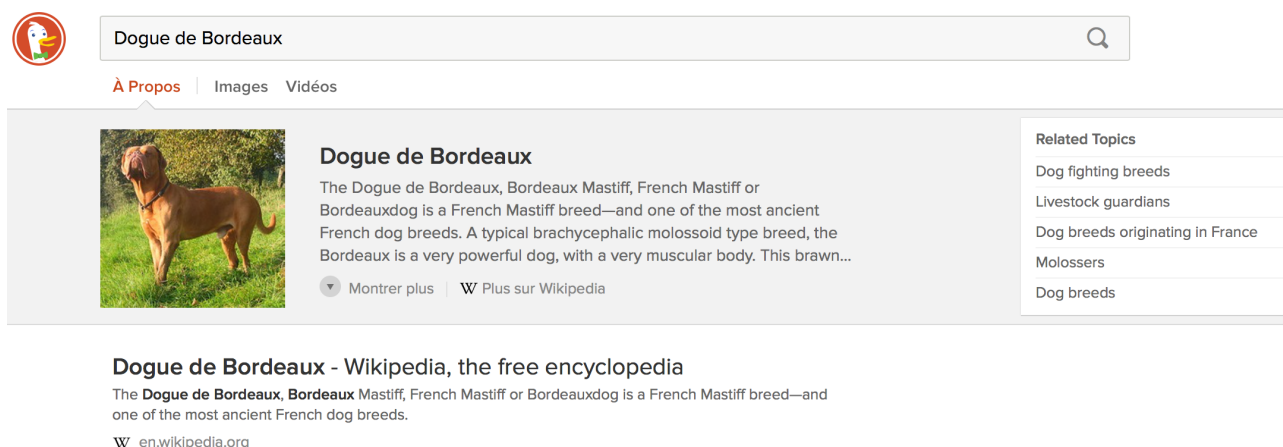
[W en.wikipedia.org](https://en.wikipedia.org)

Bordeaux wine - Wikipedia, the free encyclopedia

A **Bordeaux** wine is any wine produced in the **Bordeaux** region of France, centred on the city of **Bordeaux** and covering the whole area of the Gironde department, with a total vineyard area of over 120,000 hectares, making it the largest wine growing area in France. Average vintages produce over 700 ...

[W en.wikipedia.org](https://en.wikipedia.org)

FIGURE 1 – Exemple de requêtes avec identification de ressources possibles (moteur de recherche Web duck-duckgo)



Dogue de Bordeaux - Wikipedia, the free encyclopedia

The **Dogue de Bordeaux**, **Bordeaux** Mastiff, French Mastiff or Bordeauxdog is a French Mastiff breed—and one of the most ancient French dog breeds.

[W en.wikipedia.org](https://en.wikipedia.org)

FIGURE 2 – Concept identifié : suggestion des thèmes/concepts liés

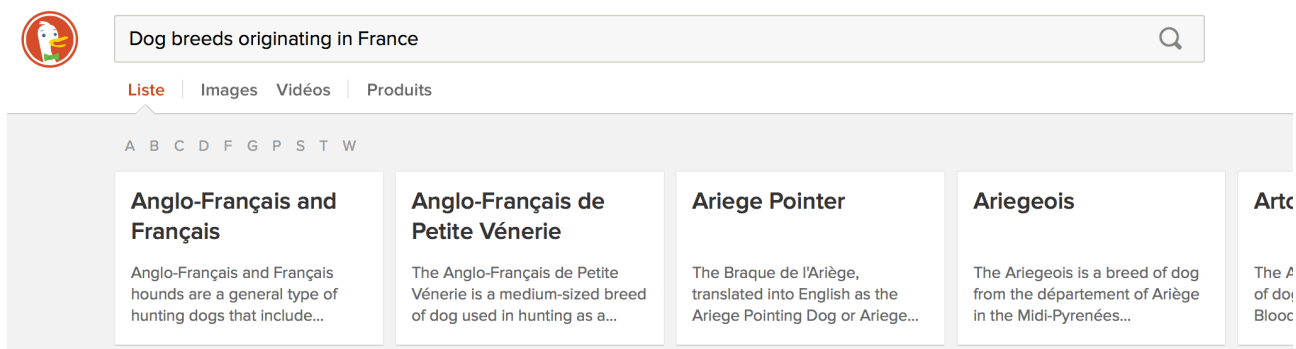


FIGURE 3 – Concept identifié, listes des catégories.