WEB OF LINKED DATA (PATIENT AND CANCER)

* Create Maven project and import Apache Jena dependencies (<https://www.youtube.com/watch?v=N_GpbwMHeMQ>)
* mapping sql to rdf whith d2r server or **create a rdf file where we have to describe instances using classes and property**
* create rdf schemas : describe classes and property
* indexation and search motor
* Expansion de requête : thésaurus WordNet, co-occurence

Expansion de requête : utilisation RDF et SPARQL + calcul similarité requête initiale et requête étendue, utilisation du thésaurus (RDF issu de Neo4j) pour définir les termes liés à la requête pour faire l’expansion

?? recherche synonyme du mot

Utilisation Edit Distance correction erreur syntaxique commises par l’utilisateur en se basant sur une collection de termes issus de la tokenisation

Utiliser FrenchAnalyzer pour indexation des documents

Télécharger un corpus sur les cancers

Calcul cosine-similarité ou BM25 permet de classer les documents par ordre de pertinence

Création documents avec lucene <https://howtodoinjava.com/lucene/lucene-index-search-examples/> et indexation ++++

Expansion par synonyme https://stackoverflow.com/questions/36806978/query-expansion-lucene

Création classe SplitDocument

Bases :

Information RDF : Sujet - Prédicat – Objet

Sémantique = étude du sens, de la signification des signes, notamment du langage

Ontologie : spécification (description exhaustive d’un domaine, en définit les termes) formelle (sans ambiguïté) et explicite (aucune connaissance ne doit être sous-entendue) d’une conceptualisation (représentation du monde) partagée (doit faire consensus au sein du domaine qu’il décrit).