

Errol

- whiteboard
- Neo4j
- api
- Todo
- classifier

whiteboard

- combien de sous me reste t'il ?
- combien coute une baguette de pain ?
- combien y a t'il de proteines dans 100g de pomme de terre ?

1. Analyse de la grammaire. On cherche une quantité de "protéine" contenu dans "100g de pomme de terre"
2. Recherche de la definition de "proteine" et de "100g de pomme de terre"

I - Simplification des phrase complexe en phrase simple

I.2 - "Traduction" des verbes compliqué en verbe simple grace au cnrtl ou autre moteur de synonyme:

- import dans neo4j ?
- parcours de graph -> trouver le synonyme connu le plus proche ?

II - conversion de phrase simple en requête RDF ou en requête neo4j ?

I.1 découper en mot et phrase

Lien en vrac:

- <https://crisco2.unicaen.fr/des/synonymes/contenir>
- <https://www.atilf.fr/ressources/tlfi/>
- <http://www.nooj4nlp.org/index.html>
- <https://www.modyco.fr/fr/base-documentaire/ressources/jean-dubois.html> => dictionnaires francais pour machines
- talep.lif.univ-mrs.fr/FondamenTAL/ => même dictionnaire
- <http://rali.iro.umontreal.ca/rali/?q=fr/versions-informatisees-lvf-dem>
- <https://dumps.wikimedia.org/frwiktionary/>
- <http://infolingua.univ-mlv.fr/DonneesLinguistiques/Dictionnaires/telechargement.html>

Neo4j

Recherche:

```
MATCH (v1:verbe {label:'réfréner'}), (v2:verbe {label:'contenir'})  
RETURN v1,v2
```

Creer un index

```
CREATE INDEX FOR (verbe:label) ON (verbe.label)
```

Creer un noeud

```
CREATE (contenir:verbe {label:"contenir"})  
return contenir
```

Faire un relation plus un nouveau noeud

```
MATCH (contenir:verbe {label:"contenir"})  
CREATE (contenir)-[synonyme:SYNONYME {poid:3}]->(refrener:verbe {label:"réfréner"})  
RETURN contenir,synonyme,refrener
```

Faire une relation entre deux noeuds existants

```
MATCH (u:User {username:'admin'}), (r:Role {name:'ROLE_WEB_USER'})  
CREATE (u)-[:HAS_ROLE]->(r)
```

api

<https://www.wikidata.org/w/api.php?action=wbsearchentities&language=fr&search=pr%C3%A9sident&type=property&format=json>

Todo

1. correcteur orthographique
2. traitement récursif des triplets (ex: Quelle est la taille du premier-ministre de la france ?)
3. améliorer la tokenisation sur les mots "courants" (special rules spacy)
4. vérifier le vrai sens (président de l'angleterre => ca donne le roi, donc verifier après coup que c'est bien un président et pas un roi)
5. compréhension des verbes ?
6. gérer les mots type 'français' 'anglais' synonyme de "de la France" "de l'Angleterre"
7. gérer le cas du premier ministre anglais (qui n'a pas de relation dans wikidata)
8. gérer le cas où la recherche est nulle (la chine =>
<https://www.wikidata.org/wiki/Q113473434>) (peut être un soucis du point 3)
9. réfléchir à l'adapter pour un usage actionnel 'allume la lampe du salon'
10. gérer les synonymes

Qui est le président du pays qui a bâti la tour eiffel ?

Quand a été bâtie la tour eiffel

comment se nomme le président de la france ?

Quel est la fonction d'Emmanuel macron ? (ca fonctionne déjà :p)

Quelle est la couleur des yeux de Xi Jinping ? => ('couleur des yeux', 'xi Jinping')

classifier

<https://github.com/davidberenstein1957/classy-classification>