

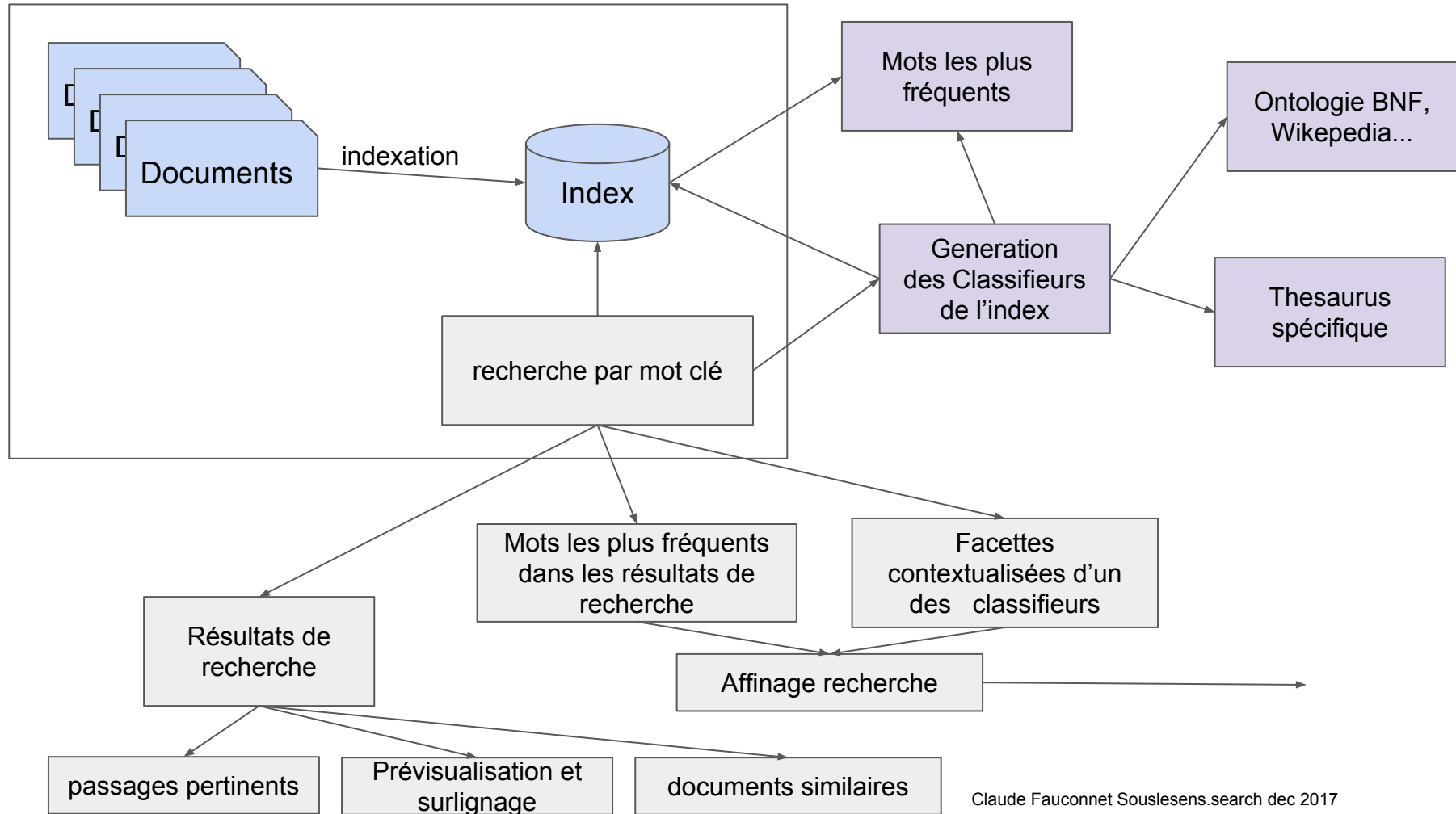
# SouslesensSearch

portail Elastic search inclus dans la suite de valorisation de données souslesens

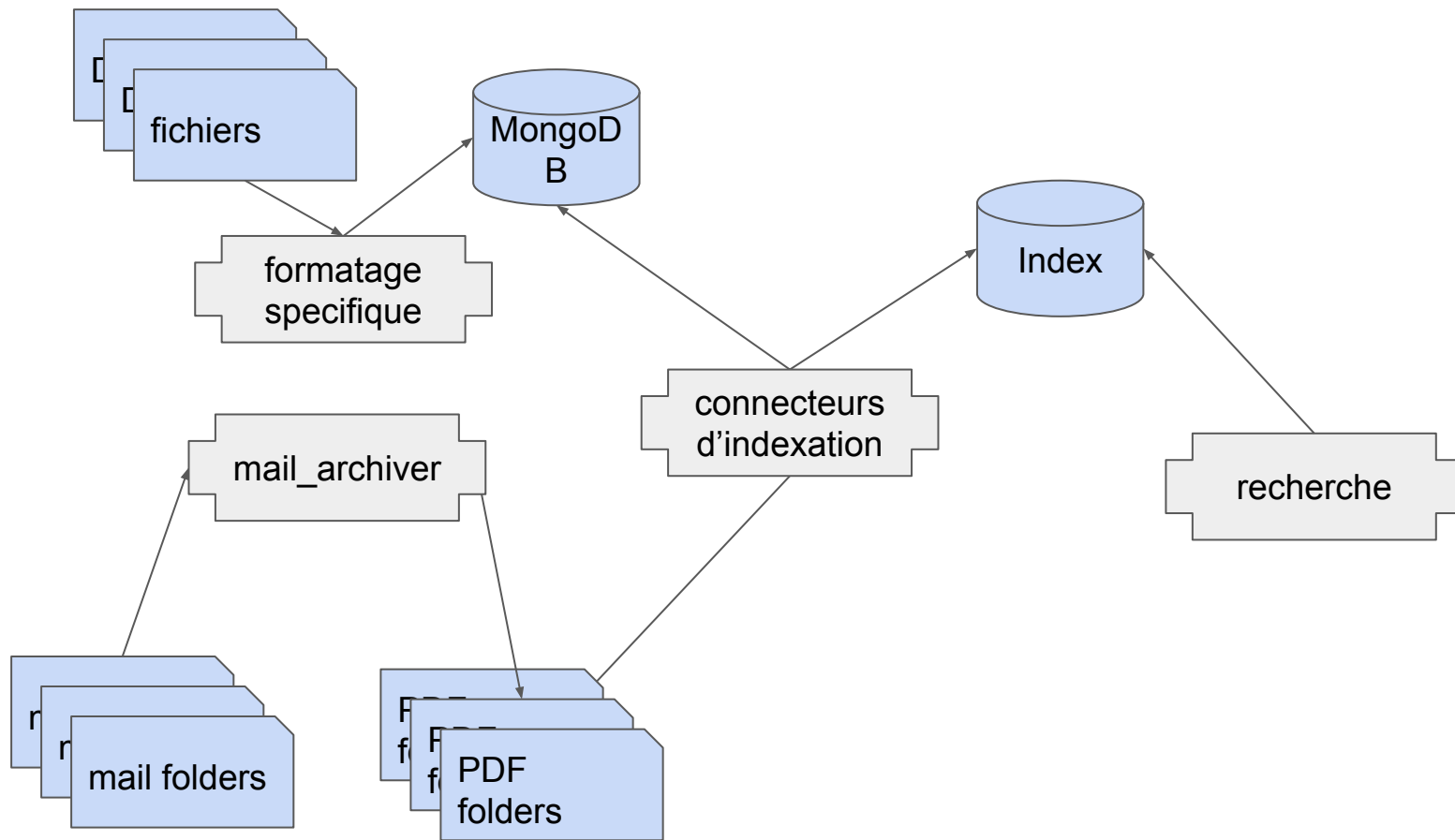
licence MIT <https://github.com/claudefaucconnet/souslesens>

Claude Fauconnet Dec 2017

# Architecture Souslesens.search



# Souslesens.search maquette ATD



# Souslesens.search processus d'enrichissement sémantique

*Indexation*

*Sinequa ElasticSearch*

*Mots les plus frequents*

*Ou les plus significatifs*

*extraction des concepts basée sur une ontologie*

*Pour chaque Mot recherche de  
broaders, narrowers et related  
dans l'ontologie (SPARQL)*

*Plusieurs ontologies de référence  
possibles en fonction de la  
langue, du domaine ...*

*génération d'un thésaurus SKOS  
spécifique au corpus de l'index*

*enrichissement manuel du thésaurus*

*calcul du classifier*

*Pour chaque concept du  
thésaurus, recherche des  
documents qui contiennent ses  
différentes formes lexicales*

```
classifier = {  
  broaderNodes: {},  
  words: {},  
  docIds: {}  
}
```

Back office