

# LoGE: Expansion Locale- Globale de document non supervisee avec un moteur de recherche Extensible

Oussama Ayoub, Ludovic Li, Christophe  
Rodrigues, Nicolas Travers

Atelier TextMine – EGC2023

# Contexte



**Recherche d'informations parmi des documents hétérogènes et potentiellement très longs**

**Contraintes de facilité de mise en œuvre au sein d'un SI existant**

**Temps de réponse le plus réduits possible**

# Travaux connexes

- Nombreuses approches par apprentissage profond supervisées :
  - supposent un grand nombre de couples requête/document à disposition
  - supposent des ressources de calcul conséquentes
- Quelques approches non-supervisées comme UDEG, (NAACL,2021)
- Approche classique, BM25, Robertson et Jones, 1976 :

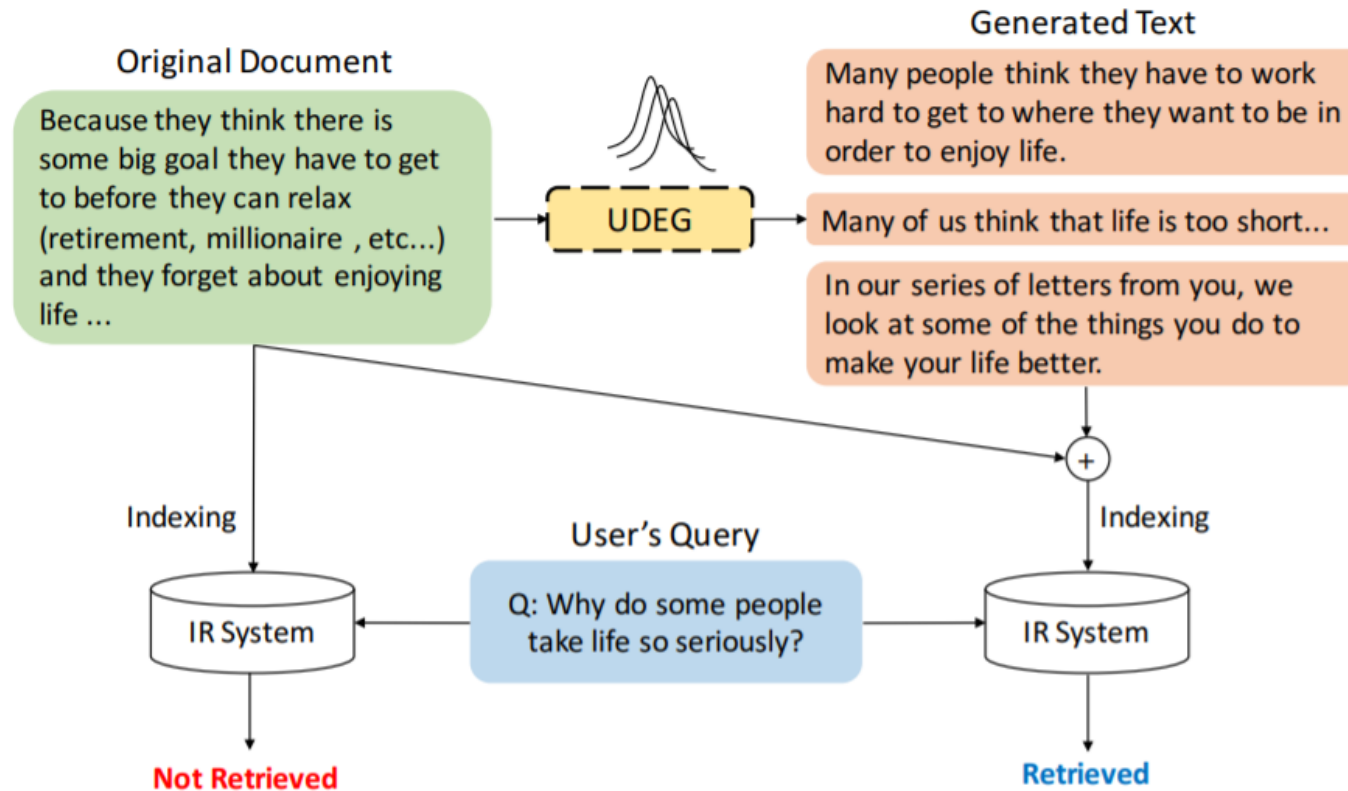
$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{\text{avgdl}})}$$

$$\text{IDF}(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}$$

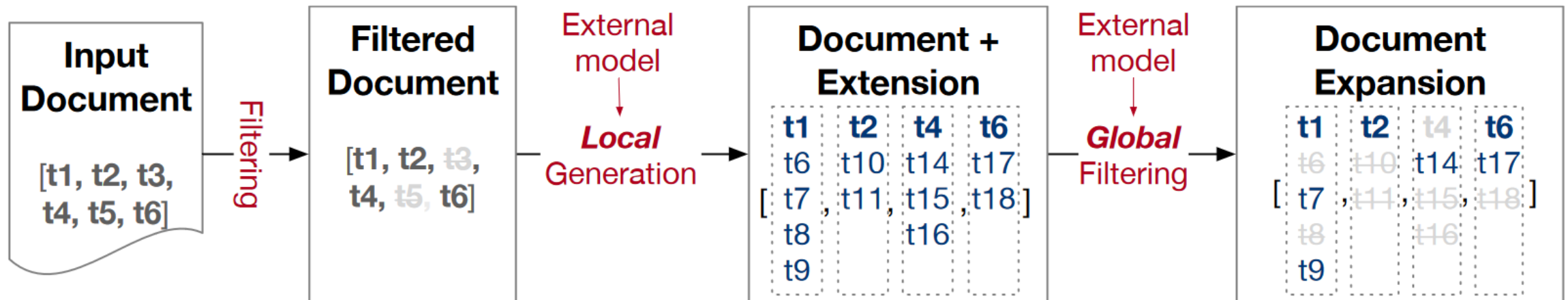
# Enrichissement de documents

\* UDEG, Jeong & al. (2021)

Résumés Abstractifs Supervisés / Extensions / BM25



# Notre Proposition : Processus d'expansion Locale-globale(LoGE)



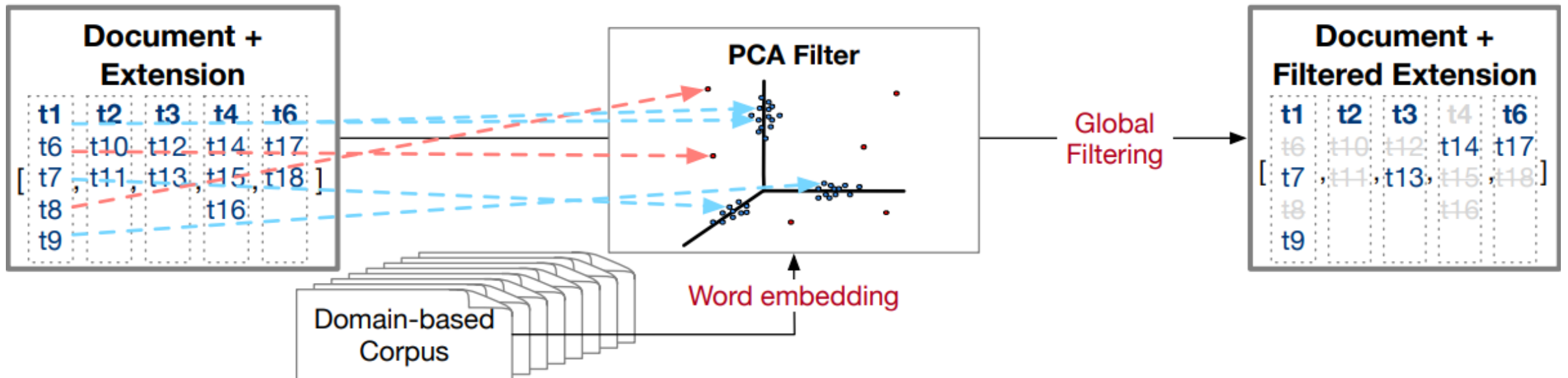
# Génération Locale



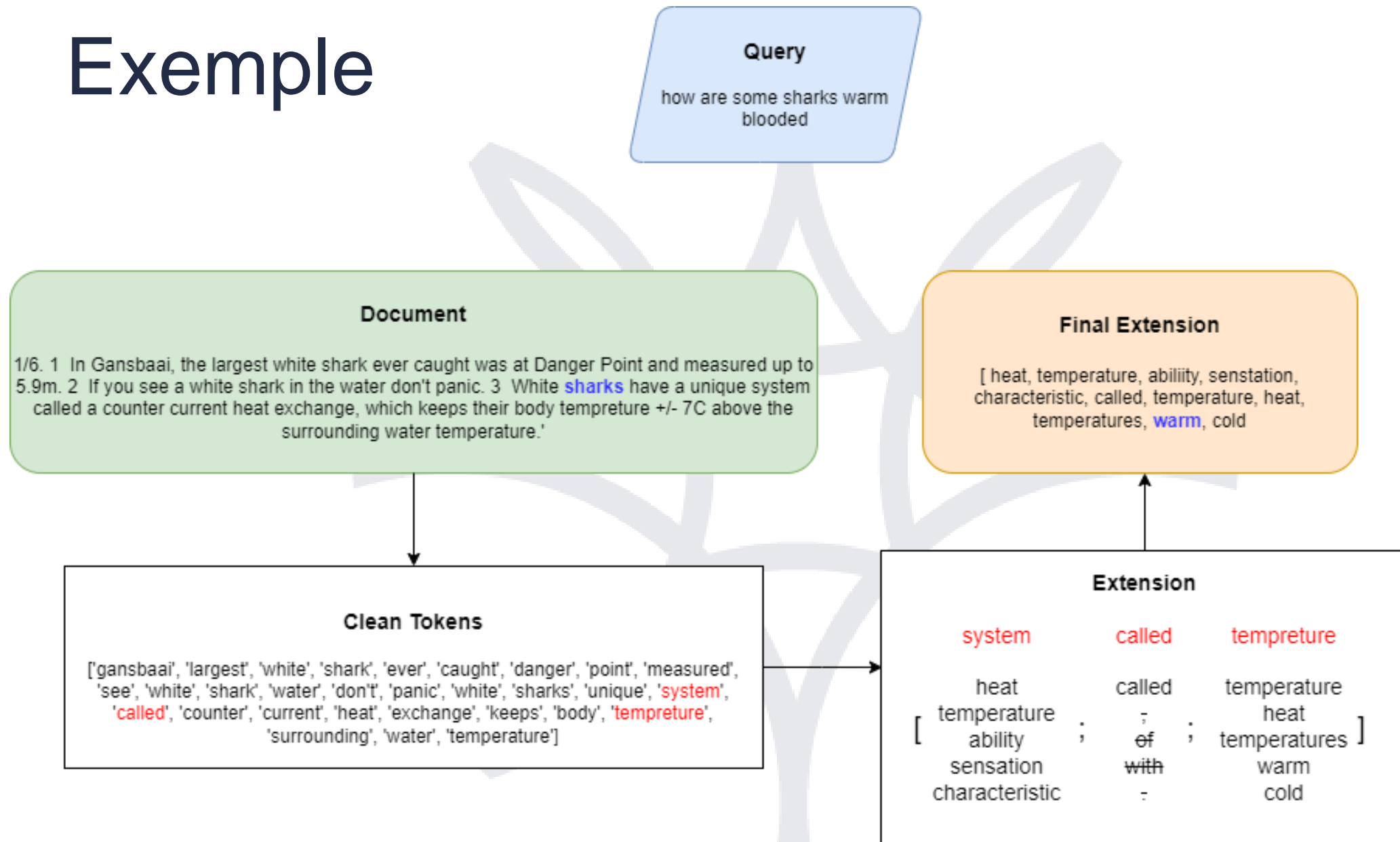
- Inspiré par les travaux : *Beyond Accuracy: Behavioral Testing of NLP Models with CheckList*, Ribeiro & al., ACL 2020
- Enrichissement de documents en utilisant des modèles de langage pré-entraînés (ne nécessitant que des corpus volumineux) pour générer des extensions.
- Possibilité d'utiliser des modèles spécialisés (LegalBERT, ClinicalBERT,...)

# Filtrage Global

- Méthode adaptée de Gong & al., (ACL 2018)
- A partir des plongements vectoriels de chaque mots, les mots les plus proches des composantes principales du documents sont gardés

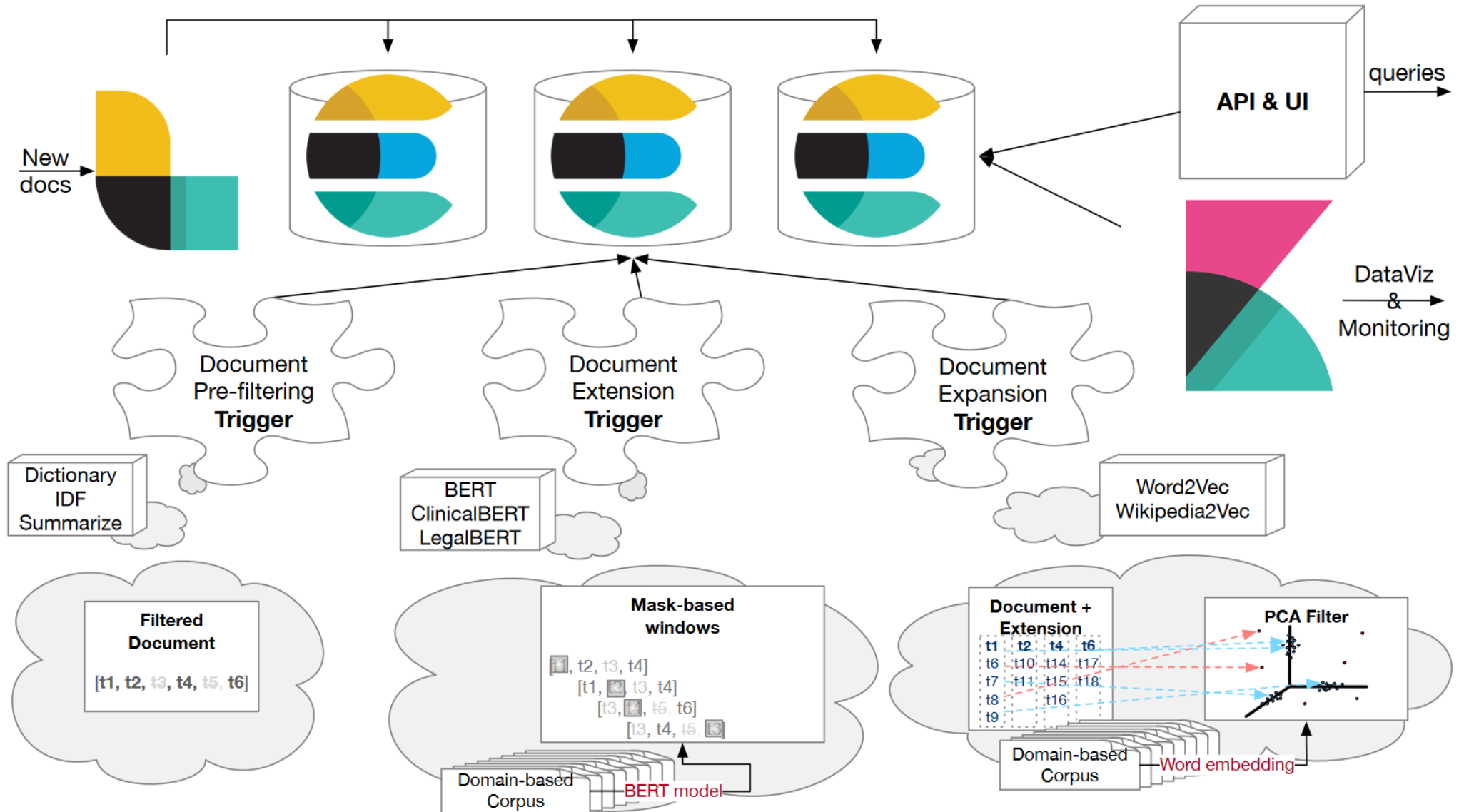


# Exemple





# Architecture de LoGE



# Quelques résultats

- EU/UK Legislation Dataset [Ilias et al. (2021)]
- 50K documents/2K requêtes/7K mots en moyennes

Model	NDCG@10	Precision@10	Recall@10	MAP@10	F1@10
Base	0,5609	0,0706	0,6814	0,1607	0,1275
Pegasus	0,6058	0,0853	0,8186	0,1719	0,1538
UDEG (sans Base)	0,2772	0,0382	0,3652	0,0777	0,0689
Loge 5	<b>0,7642</b>	<b>0,0926</b>	<b>0,902</b>	<b>0,2196</b>	<b>0,1676</b>



**Merci de votre attention  
Des questions?**