

Présentation finale

Vectorisation de texte

Hanane, Farouq, Walid, Asma, Steve

Université d'Avignon

6 janvier



Plan

- 1 Démo finale
- 2 Projet et méthode Scrum
- 3 Bilan

US2 – Fréquence des mots

- On colle un texte.
- On affiche les mots les plus fréquents.

← Retour

Fréquence des mots

Texte :

mot s'applique à l'entrelacement des fibres utilisées dans le tissage, voir par exemple Ovide : « Quo super [tolle]t textum rube baucis laevis » (un vêlage) sur lequel Baucis empressée avait jeté un tissu grossier »[2] ou au tressage (exemple chez Martial « Vincetum textum » panier d'osier tressé »). Le verbe a aussi le sens large de construire comme dans « baulican texere » construire une basilique » chez Cicéron[3].

Le sens figuré d'éléments de langage organisés et enchaînés apparaît avant l'Empire romain : Il désigne un agencement particulier du discours. Exemple : « glistodas texere » composer des épitres

Lancer l'analyse

Mot	Fréquence
sacile	6
mot	5
texte	4
chez	3
exemple	3
sen	3
texere	3
textum	3
agencement	2
applique	2
baucis	2
cicéron	2
comme	2
construire	2
ier	2

US4 – Similarité entre deux textes

- On colle deux textes.
- On obtient un score de similarité.

[← Retour](#)

Similarité entre deux textes

Texte 1 :

« verba in textu juxtapuncta » l'agencement des mots dans la phrase »[5].

Les formes anciennes du Moyen Age désignent au XIIIe siècle le volume qui contient le texte sacré des Évangiles, puis au XIIIe siècle, le texte original d'un livre saint ou des propos de quelqu'un. Au XVIIe siècle le mot s'applique au passage d'un ouvrage pris comme référence et au début du XIXe siècle le mot texte a son sens général d'« écrit ».

Texte 2 :

de texte brut dans un protocole de télécommunication ;
de document texte si le logiciel utilisé permet de formater le texte ;
ou encore de texte formaté ou riche lorsque les indications de formats sont données en texte brut (exemple : Rich Text Format) ;
de fichier texte dans un fichier qui ne contient que des caractères sans mise en forme autre que les sauts de lignes et le choix des caractères.

[Comparer](#)

Similarité : 3.13 %

US6 – Snippets : meilleur passage

- On donne un document + une requête.
- On affiche le meilleur passage (fenêtre).

Moteur de Recherche de Snippets

1. Document (Texte source)

Le sens figuré d'éléments de langage organisés et enchaînés apparaît avant l'Empire romain : il désigne un agencement particulier du discours. Exemple : « epistolas texere » = composer des épîtres » - Cicéron (Ier siècle av. J.-C.) [4] ou plus nettement chez Quintilien (Ier siècle apr. J.-C.) : « verba in textu iunguntur » = l'agencement des mots dans la phrase » [5].

Les formes anciennes du Moyen Âge désignent au XIIe siècle le volume qui contient le texte sacré des Évangiles, puis au XIIIe siècle, le texte original d'un livre saint ou des propos de quelqu'un. Au XVIIe siècle le mot s'applique au passage d'un ouvrage pris comme référence et au début du XIXe siècle le mot texte a son sens général d'« écrit » [6].

2. Requête (Mots clés)

texte

Taille Fenêtre (mots)

3

Nombre de phrases (Top K)

3++

☒ Activer la pondération TF-IDF (si décoché : comptage simple)

• Meilleur Passage (Fenêtre)

📄 Top Phrases Clés

Meilleur Passage (Contexte continu)

Score: 0.7631 TF-IDF | Position mot: 0

"« Texte »"

US6 – Snippets : Top phrases (Top K)

- On choisit Top K.
- On affiche les phrases les plus pertinentes.

Moteur de Recherche de Snippets

1. Document (Texte source)

Le sens figuré d'éléments de langage organisés et enchaînés apparaît avant l'Empire romain : il désigne un agencement particulier du discours. Exemple : « épistoles tenebre » composer des épîtres - « Cédron (lar) siècle au 2 - 4-16) au plus vintaines cher qualificatif (lar) siècle apr. 4-5-1) : « norme au texte tenebre » - l'agencement des mots dans la phrase «[3].

Les formes anciennes du Moyen Âge désignent au XIIIe siècle le volume qui contient le texte sacré des Évangiles, puis au XIIIe siècle, le texte original d'un livre saint ou des propos de quelque un. Au XIIIe siècle le mot s'applique au passage d'un ouvrage pris comme référence et au début du XIIIe siècle le mot

2. Requête (Mots clés)

texte

Taille Fenêtre (mots) : 3 Nombre de phrases (Top K) : 1

☒ Activer la pondération TF-IDF (si décoché : comptage simple)

Top N-1 Phrases Pertinentes

#1 | Score: 1.5282 | Index phrase: 7

"Les formes anciennes du Moyen Âge désignent au XIIIe siècle le volume qui contient le texte sacré des Évangiles, puis au XIIIe siècle, le texte original d'un livre saint ou des propos de quelque un."

#2 | Score: 0.7621 | Index phrase: 0

"Le Texte « est issu du mot latin « textum », dérivé du verbe « texere » qui signifie « tisser »."

#3 | Score: 0.7621 | Index phrase: 8

"Au XIIIe siècle le mot s'applique au passage d'un ouvrage pris comme référence et au début du XIIIe siècle le mot tente a son sens général d'écrit «[8]."

#4 | Score: 0 | Index phrase: 1

"Le mot s'applique à l'entrelacement des fibres utilisées dans le tissage, voir par exemple Ovide : « Quo ruper innect tentum rube vedula Baucis » (un siège) sur lequel Baucis empressée avait jeté un tissu grossier «[2] ou au tissage (exemple chez Martial) : « Viminum textum » panier d'osier tissé"

#5 | Score: 0 | Index phrase: 2

"Le verbe a aussi le sens large de construire comme dans « basilicum texere » construire une basilique - chez Cédron«[3]."

#6 | Score: 0 | Index phrase: 3

"Le sens figuré d'éléments de langage organisés et enchaînés apparaît avant l'Empire romain : il désigne un agencement particulier du discours."

#7 | Score: 0 | Index phrase: 4

"Exemple : « epistolae texere » composer des épîtres - « Cédron (lar) siècle au 2 - 4-16) au plus vintaines cher qualificatif (lar) siècle apr. 4-5-1) : « norme au texte tenebre » - l'agencement des mots dans la phrase «[3]."

US5 – Résumé automatique

- On choisit un pourcentage à garder.
- On génère le résumé avec une méthode.

[← Retour](#)

Résumé automatique

Texte à résumer :

« Texte grossier »[2] ou du tressage (exemple chez Martial « vimineum tessum = panier d'osier tressé »). Le verbe a aussi le sens large de construire comme dans « basilicam texere = construire une basilique » chez Cicéron[3].

Le sens figuré d'éléments de langage organisés et enchaînés apparaît avant l'Empire romain : il désigne un agencement particulier du discours. Exemple : « epistolas texere = composer des épîtres » - Cicéron (Ier siècle av. J.-C.)[4] ou plus nettement chez Quintilien (Ier siècle apr. J.-C.) : « verba in textu iungantur = l'agencement des mots dans la phrase »[5].

Les formes anciennes du Moyen Âge désignent au XIIe siècle le volume qui contient le texte sacré des Évangiles, puis au XIIIe siècle, le texte original d'un livre saint ou des propos de quelqu'un. Au XVIIe siècle le mot s'applique au passage d'un ouvrage pris comme référence et au début du XIXe siècle le mot texte a son sens général d'« écrit »[6].

Pourcentage à garder : %

Méthode de résumé :

Premières phrases

Générer le résumé

« Texte » est issu du mot latin « textum », dérivé du verbe « texere » qui signifie « tisser ». Le mot s'applique à l'entrelacement des fibres utilisées dans le tissage, voir par exemple Ovide : « Quo super iniecit textum rude sedula Baucis = (un siège) sur lequel Baucis empressée avait jeté un tissu grossier »[2] ou au tressage (exemple chez Martial « Vimineum textum = panier d'osier tressé »).

US7 – Ponctuation automatique

- On donne un texte sans ponctuation.
- On obtient une version plus lisible.

[← Retour au menu](#)

Ponctuation automatique

Texte sans ponctuation :

il était une fois dans un pays lointain un roi sage et juste
il gouvernait avec bonté son peuple était heureux et prospère
un jour un dragon terrifiant apparut dans les montagnes il semait la terreur

Ajouter la ponctuation

Résultat :

Il était une fois. Dans un pays lointain. Tain. Un roi sage et juste, il gouvernait avec bonté, son peuple était heureux et prospère, un jour. Un dragon terrifiant apparut dans les montagnes, il semait la terreur, ur,.

Sujet du projet

- Thème : application web de traitement de texte.
- On a développé : TF-IDF, similarité, résumé, snippets, ponctuation.
- Objectif : livrer petit à petit en appliquant Scrum.

Product Backlog

- On a écrit nos besoins en **User Stories** : interface, TF-IDF, similarité, résumé, snippets, ponctuation, idée LLM.
- Pourquoi : pour **prioriser** ce qui est important.
- Exemple : l'US **LLM** est restée à *faire* car on a d'abord livré les modules de base.

D	User Story	Definition of Done (DoD)	Description / Fonctionnalité	Sprint	Statut
US1	En tant qu'utilisateur, je veux une page d'accueil web claire avec les 4 modules	- Page responsive- 4 boutons fonctionnels- Navigation fluide	Interface de navigation pour rentrer le texte à résumer + la taille , un bouton pour afficher un les mots les plus fréquents en haut dans le texte rentré	1,2,3,4	In progress
US2	En tant qu'utilisateur, je veux afficher les vecteurs de mots (mots les plus fréquents en haut)	- Upload texte- Calcul fréquence- Top 10 mots affichés	Visualisation statistique des mots importants	1	Fait
US3	En tant qu'utilisateur, je veux calculer de TF et IDF	- Saisie texte- Calcul TF/IDF- Affichage tableau	- Saisie texte- Calcul TF/IDF- Affichage tableau		Fait
US4	En tant qu'utilisateur, je veux comparer deux textes pour détecter une similitude / plagiat	- 2 zones texte- Score 0-1- Visualisation	Calcul de similarité entre textes	2,3	Pas commencé
US5	En tant qu'utilisateur, je veux coller un texte et obtenir un résumé	- Choix % réduction- Résumé cohérent- Mots clés concernés	Module de résumé automatique avec choix de la taille du résumé	2, 3	Pas commencé

Sprints

- On a fait **6 sprints** au total.
- Un sprint = **une séance de TP** (sprint court).
- Pourquoi : en 3h, on vise un **résultat visible** (page, score, module utilisable).

Sprint Backlog + Kanban

- Au début du sprint : on choisit les US du sprint.
- On les découpe en **tâches** et on les met sur un tableau **Kanban**.
- On attribue les tâches selon la difficulté et les capacités de chacun.

timing

Aa tâches	📅 Date	👤 hana	⚙ État
📄 rectification de TF-IDF et des stopwor	4 décembre 2025 16:00 → 17	👤 Hana Boud	● Terminé
📄 résumé les 3 methodes	4 décembre 2025 16:00 → 17	👤 Asma Nihal Boukraa	● Terminé
📄 page du résumé + affichage du résult	4 décembre 2025 16:00 → 17	👤 FARAH Mohamed Walid	● Terminé

Méthodes d'équipe

- **Complexity matching (Sprint 3)** : tâches NLP difficiles aux personnes les plus à l'aise.
- **Pair programming** : utile sur similarité / résumé (Driver–Navigator).
- But : réduire les erreurs et livrer plus vite dans un sprint court.

Rétrospective : déroulement du projet

- Début : on a sous-estimé certaines tâches (ex. TF-IDF Wikipédia).
- Milieu : meilleure organisation (répartition + binômes) \Rightarrow meilleure stabilité.
- Fin : on a consolidé snippets + ponctuation.

Difficultés rencontrées

- Estimation : TF-IDF Wikipédia plus long que prévu.
- Intégration front/back : calcul OK mais affichage pas prêt (page/menu).
- Tests : peu de tests bout-en-bout (temps de séance).
- Organisation : absences et répartition parfois déséquilibrée.

Leçons apprises

- Garder 10 min fin de sprint : **merge + test complet + mini démo.**
- Une DoD claire aide : *affiché et testable* > *juste calculé.*
- Complexity matching + pair programming : bon choix pour livrer vite.

Perspectives

- Stabiliser l'app : version propre et intégrée (pages + modules).
- Ajouter des tests simples bout-en-bout : texte → résultat affiché.
- Améliorer la ponctuation