

Fait par : Nemin Aichetou & EL-kaissi Hind

Encadré par : Pr. Mahmoudi

IDDL0 2020-2021

NLP

Mini projet

## **NLP-- Text summarization**

### **o Introduction:**

L'objectif du projet est de comprendre les concepts de traitement et de création du langage naturel .

Le projet se concentre sur créer un outil qui résume automatiquement les documents.

Pour le réaliser ,nous avons donc créé une application Web “ Text Summarizer “ qui génère des résumés automatiquement avec le choix de la langue (arabe, Français ou anglais), idéal pour ceux qui n'ont pas le temps d'analyser tout le contenu auquel ils ont accès, des documents de recherche ou tout texte long.

Cela s'est fait en html , css et python à l'aide de **Flask** et des packages NLP tels que **NLTK** et en utilisant aussi le framework PureCss.

#### o **NLTK :**

NLTK est une bibliothèque essentielle qui prend en charge des tâches telles que la classification, la recherche de racines, le balisage, l'analyse, le raisonnement sémantique et la tokenisation en Python.

#### o **Flask :**

Flask est un framework web open-source de développement web en python . Cela signifie que Flask fournit des outils, des bibliothèques et des technologies qui vous permettent de créer une application Web.

#### o **Pure css**

Pure css est un framework front-end (HTML5 et CSS) spécialement conçu pour le développement d'applications web responsives.

## o Technique utilisée :

**Résumé basé sur l'extraction** : cette technique consiste à extraire des phrases clés du texte source et à les combiner pour en faire un résumé. C'est pourquoi, identifier les bonnes phrases pour la synthèse est important dans une méthode extractive.

## o Algorithme

- ✓ Enlever les espaces , les majuscules et les retours à la ligne du texte
- ✓ Enlever les stop words et la ponctuation puis faire la Tokenization en mots et en phrases du texte
- ✓ Calculer les fréquences des mots dans un dictionnaire (clé - mot, valeur - la fréquence)
- ✓ Calculer les fréquences pondérées et le score de chaque phrase
- ✓ Choisir Les n meilleures phrases pour le résumé
- ✓ Mettre en ordre les phrases du résumé suivant le contexte

## o Les langues utilisées :

- ✓ Français
- ✓ Anglais
- ✓ Arabe

## Exécution :

sur netbeans :

```
PS C:\Users\pc_hind\Documents\Flask> py "summarize.py"
* Serving Flask app 'summarize' (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
* Restarting with stat
* Debugger is active!
* Debugger PIN: 127-444-135
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

ou juste sur la ligne de commande :

```
C:\Users\pc_hind\Documents\Flask>py "summarize.py"
* Serving Flask app 'summarize' (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
* Restarting with stat
* Debugger is active!
* Debugger PIN: 127-444-135
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

## Web app :

La langue par défaut est l'arabe mais on peut choisir l'anglais par exemple en cliquant sur le bouton associé et la page s'affiche comme suit :

SUMMARIZERReset

TEXT  
SUMMARIZATION

LANGUAGE : ENGLISH

French

English

Arabic

Text to summarize

SUMMARIZERReset

Text to summarize

Number of sentences wanted :  Summarize

Summary

Après avoir choisi la langue ,on fait entrer notre texte , on choisit le nombre de phrases à avoir dans le résumé et on clique sur le bouton pour afficher le résultat .  
 Pour relancer le site a zero on clique sur 'Reset'.  
 Pour tester , résumons un long texte arabe pour voir le résultat :

SUMMARIZER

Reset

النص

خلال هذه الفترة، ظهرت الندية بين القوتين العظيمتين خلال التحالفات العسكرية والدعاية وتطوير الأسلحة والتقدم الصناعي وتطوير التكنولوجيا والتسابق الفضائي. ولقد اشتركت القوتان في الإنفاق الضخم على الدفاع والترسانات النووية والحروب غير المباشرة – باستخدام وكلاء.

في ظل غياب حرب ممتدة بين الولايات المتحدة والاتحاد السوفيتي قامت القوتان بالانشراك في عمليات بناء عسكرية وصراعات سياسية من أجل المساهمة على الرغم من أن الولايات المتحدة والاتحاد السوفيتي كانا حلفاء ضد قوات المحور إلا أن القوتين اختلفتا في كيفية إدارة ما بعد الحرب وإعادة بناء العالم. خلال السنوات التالية للحرب، انتشرت الحرب الباردة خارج أوروبا إلى كل مكان في العالم. حيث سعت الولايات المتحدة إلى سياسات المحاصرة والاستئصال للشيوعية وحشد الحلفاء خاصة في أوروبا الغربية والشرق الأوسط. خلال هذه الأثناء، دعم الاتحاد السوفيتي الحركات الشيوعية حول العالم خاصة في أوروبا الشرقية وأمريكا اللاتينية ودول جنوب شرق آسيا.

صاحبت فترة الحرب الباردة عدة أزمات دولية مثل أزمة حصار برلين 1948-1949 والحرب الكورية 1950-1953 وأزمة برلين عام 1961 وحرب فيتنام 1956-1975 والغزو السوفيتي لأفغانستان وخاصة أزمة الصواريخ الكوبية 1962 عندما شمر العالم أنه على حافة الانجراف إلى حرب عالمية ثالثة. آخر أزمة حدثت خلال تدريبات قوات الناتو 1983. شهدت الحرب الباردة أيضاً فترات من التهدئة عندما كانت القوتين تسميان دهر التهدئة. كما تم تجنب المواجهات العسكرية المباشرة لأن حدوثها كان سيؤدي إلى دمار محتم لكلا الطرفين بسبب الأسلحة النووية.

عدد جمل الملخص: 3

تلخيص

الملخص

الحرب الباردة هو مصطلح يُستخدم لوصف حالة الصراع والتوتر والتنافس التي كانت توجد بين الولايات المتحدة والاتحاد السوفيتي وحلفائهم من فترة منتصف الأربعينيات حتى أوائل التسعينيات. على الرغم من أن الولايات المتحدة والاتحاد السوفيتي كانا حلفاء ضد قوات المحور إلا أن القوتين اختلفتا في كيفية إدارة ما بعد الحرب وإعادة بناء العالم. صاحبت فترة الحرب الباردة عدة أزمات دولية مثل أزمة حصار برلين 1948-1949 والحرب الكورية 1950-1953 وأزمة برلين عام 1961 وحرب فيتنام 1956-1975 والغزو السوفيتي لأفغانستان وخاصة أزمة الصواريخ الكوبية 1962 عندما شمر العالم أنه على حافة الانجراف إلى حرب عالمية ثالثة.

Exemple d'un texte en français :

## texte à résumer

Une perte brutale de l'odorat (anosmie), associée ou non à une perte du goût (agueusie), est une manifestation relativement fréquente et parfois révélatrice de l'infection par le SARS-CoV-2.

Le taux de formes asymptomatiques est estimé à près de 20 % des personnes infectées. La transmission interhumaine se fait surtout par gouttelettes respiratoires et aérosolisation lorsqu'une personne respire dans un même lieu clos ou à proximité immédiate d'autres personnes. La transmission augmente en intérieur mal ventilé et lorsque la personne infectée tousse, éternue, parle ou chante. La transmission indirecte par surface contaminée, bien que possible, n'a pas été formellement démontrée. La période d'incubation est en moyenne de 5 à 6 jours, avec des extrêmes pouvant aller de deux à quatorze jours.

Une proportion importante de personnes infectées, notamment d'enfants, ne présente que peu ou pas de symptômes mais peut néanmoins transmettre la maladie. De plus, il a été établi que, chez les personnes symptomatiques, la contagiosité peut précéder les symptômes de 2 à 3 jours et qu'elle serait maximale

Le nombre de phrases souhaité :  Résumer

## Résumé

dans les formes les plus graves, l'apparition d'un syndrome de détresse respiratoire aiguë peut entraîner la mort, notamment chez les personnes plus fragiles du fait de leur âge ou en cas de comorbidités. de plus, il a été établi que, chez les personnes symptomatiques, la contagiosité peut précéder les symptômes de 2 à 3 jours et qu'elle serait maximale durant les premiers jours de la maladie. le taux de létalité, estimé à environ 0,5 % en moyenne, varie de 0,01 % pour les moins de 15 ans à 17,4 % pour les plus de 90 ans.

## En anglais :

DOMELRE was the first domestic electrical refrigerator, invented by Frederick William Wolf Jr (1879–1954) in 1913 and produced starting in 1914 by Wolf's Mechanical Refrigerator Company in Chicago. Several hundred units were made. The unit replaced the block of ice in the icebox with an electrical-powered cooling device, and was completely automatic.

Labelled as the "first electrical refrigerator", It has been described as "revolutionary" in the history of domestic refrigeration.

DOLMERE was invented by Frederick William Wolf Jr, a charter member of the American Society of Refrigerating Engineers, in 1913. An estimated several hundred to thousands of units were produced starting in 1914 by his Mechanical Refrigerator Company in Chicago. Fred Heideman was also involved in the unit's design. In 1916 Wolf sold the rights to the invention to Henry Joy, president of Packard Motor Car Company in Detroit, which released an upgraded version under the name ISKO. Having sold about a thousand more models, Wolf's company nonetheless went bankrupt in 1922.

Number of sentences wanted :  Summarize

## Summary

dolmere was invented by frederick william wolf jr, a charter member of the american society of refrigerating engineers, in 1913. an estimated several hundred to thousands of units were produced starting in 1914 by his mechanical refrigerator company in chicago. the unit was considered relatively inexpensive for its time the original model was sold for \$900 (\$24,450 in 2021 dollars); the 1916 model was priced at \$385 in 1916 (\$9,600 in 2021 dollars), later dropping to \$275 (\$6,850 in 2021 dollars).