

СОФИЙСКИ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА

КУРСОВ ПРОЕКТ

ПО ИЗКУСТВЕН ИНТЕЛЕКТ

Тема:

(№13) Резюме на текст

Студенти:

Иван Руменов Йочев, група 3, фак. № 45552

София, юни 2022 г.

1. Формулировка на задачата

Да се напише програма, която по даден текст, прави резюме.

1. Използвани алгоритми

Използваният подход за целта е Extractive Text Summarization. За финалната идея съм използвал статия на Alfrick Opidi [1].

1. Описание на програмната реализация

Следният алгоритъм е използван за целта:  
1. Разделяме текста на отделни изречения  
2. От всяко изречение, които не са от значение за смисъла, като например “the”, “a”, “an”  
3. Разделяме остатъка от изречението на отделни думи, като поддържаме списък от думи за всяко изречение  
4. Според броят на срещане в целия текст, на всяка от получените думи, изчисляваме тегло за всяка от тях.  
5. Нормализираме теглата на думите в интервал от 0 до 1  
6. Изчисляваме тегло за всяко от изреченията, като сумираме теглата на думите в него  
6. Определяме брой изречения - k, като дължина на резюмето. Използваме процент от общия брой изречения  
7. Сортираме изреченията по тегло

8. Конкатенираме първите k на брой изречения от текста в оригиналния им вид  
9. Полученият низ е резюме на началния текст

1. Примери, илюстриращи работата на програмната система  
     
   Примерен текст:  
   With the present explosion of data circulating the digital space, which is mostly non-structured textual data, there is a need to develop automatic text summarization tools that allow people to get insights from them easily. Currently, we enjoy quick access to enormous amounts of information. However, most of this information is redundant, insignificant, and may not convey the intended meaning. For example, if you are looking for specific information from an online news article, you may have to dig through its content and spend a lot of time weeding out the unnecessary stuff before getting the information you want. Therefore, using automatic text summarizers capable of extracting useful information that leaves out inessential and insignificant data is becoming vital. Implementing summarization can enhance the readability of documents, reduce the time spent in researching for information, and allow for more information to be fitted in a particular area.  
     
   Генерирано резюме:  
   With the present explosion of data circulating the digital space, which is mostly non-structured textual data, there is a need to develop automatic text summarization tools that allow people to get insights from them easily.Therefore, using automatic text summarizers capable of extracting useful information that leaves out inessential and insignificant data is becoming vital.
2. Литература  
   [1] Статия на Alfrick Opidi: <https://blog.floydhub.com/gentle-introduction-to-text-summarization-in-machine-learning/>