

Для побудови Summary тексту (функція summary) необхідно:

1. Токенізувати текст за реченнями
2. Підрахувати вагу кожного речення наступним чином:
 - 2.1 Токенізувати кожне речення за словами
 - 2.2 Видалити стоп слова
 - 2.3 Видалити знаки пунктуації
 - 2.3 Підрахувати вагу кожного слова (tf, idf, tf_idf в залежності від ситуації можуть давати різні рівні ефективності)
 - 2.4 Для підрахунку ваги речення просумувати ваги всіх інформативних слів в даному реченні
3. Сформувати Summary з речень з максимальною вагою (в реалізації кількість речень в Summary визначена заздалегідь або вводиться користувачем).

Для пошуку ключових слів (функція keywords_searching) необхідно:

1. Токенізувати текст за словами
2. Видалити стоп слова
3. Видалити знаки пунктуації
4. Підрахувати вагу кожного інформативного слова (тут найдоцільніше використовувати tf, оскільки ключові слова характеризуються найбільшою частотою появ у тексті)
5. Застосувати до всіх слів стемінг або лемітизацію
6. Видалити дублікати, але вага слова, що залишиться, буде сумою ваги даного слова і ваг дублікатів
7. Сформувати список ключових слів з слів з максимальними вагами (в реалізації кількість ключових слів визначена заздалегідь або вводиться користувачем)

Використання стемінгу та/або лемітизації перед обчисленням ваг речень призведе до невірних обчислень, оскільки не будуть враховуватись різні форми одного й того ж слова, що є необхідним для пошуку ключових слів (слова може зустрічатись дуже багато разів, але кожного в різних формах, а отже найімовірніше воно не буде включене до списку ключових слів, хоча вона таким є).

Не використання стемінгу та/або лемітизації призведе до розгляду кожної форми слова як окремого слова, що також призведе до невірних обчислень.