**TextRank 알고리즘**

* 글의 핵심적인 내용을 추출하여 보다 적은 문장 수의 글로 간추리는 것
* 텍스트 요약
* 추출적 요약(Extractive Summarization)
* 추상적 요약(Abstractive Summarization)
* 추출적 요약
* 기존의 글에서 중요도가 높거나 핵심이 되는 문장 그대로 추출해서 요약글을 만드는것
* 새로운 단어가 생겨나거나 새로운 문장이 생성되지 않는 요약 방법
* 추상적 요약
* 새로운 단어와 새로운 문장을 생성해서 요약을 하는 방법
* TextRank = 추출적 요약법
* TextRank Model
* **PageRank 논문 – The PageRank Citation Ranking: Bringing Order to the Web을 기반으로 한 알고리즘**
* 하이퍼링크를 가지는 웹 페이지에 대해서 **얼마나 참조가 됐는지, 또는 유입이 되었는지** 등으로 페이지의 순위를 매기는 알고리즘
* PageRank에서는 **링크를 클릭할 확률로 그 순위를 매김**.
* TextRank : Bringing Order into Texts 논문에 따르면, TextRank는 그래프 기반의 랭킹 모델(graph – based ranking model)로, 이러한 순위를 매기는 방법이 문단의 추출적 요약에 매우 효과적일 것이라 생각되어 개발됨.

폰트, 텍스트, 친필, 화이트이(가) 표시된 사진

자동 생성된 설명

* S(Vi) = 최종적으로 구하여는 TextRank 값
* Wji = 해당 문장이나 단어 i와 j 사이의 가중치
* d = damping facter, 한 vertex에서 다른 vertex로 연결될 확률(0~1)사이 값

(논문에서는 0.85로 지정)

* vertex마다 위 식을 통해 TextRank 값을 구하고 나면, 그 크기대로 정렬해서 순위는 매기게 됨
* TextRank 파이썬으로 구현

1. summarize를 import

* from genism.summarization.summarizer import summarize

1. 요약하고자 하는 텍스트를 불러옴.
2. 끝 (사실 매우 간단함)

* summarze() 모듈
* summarize(text) : 기본값에 따라 요약
* summarize(text, ratio=0.05) : ratio 값은 0~1 사이의 실수로, 전체 문장 수에 대한 요약할 문장 수의 비율을 입력하면 그 만큼 요약.

(ex. 10문장 글에서 2문장만 요약하고 싶다면 0.2를 입력)

* summarize(text, word\_count=300) : 요약 텍스트로 구성할 단어 수 지정.

(word\_count를 사용하면 ratio설정이 무시됨)

* summarize(text, split=True)
* True : 문장이 ‘list’형태로 반환
* False(기본값) : 문자열이 합쳐져 출력