* 자카드 유사도 (Jaccard Similarity)
* 2개의 집합 A, B가 있을 때 두 집합의 합집합 중 교집합의 비율
* 즉, 두 집합이 완전히 같을 때 자카드 유사도는 ‘1’이며, 두 집합에 교집합이 없는 경우는 ‘0’

텍스트, 폰트, 화이트, 번호이(가) 표시된 사진

자동 생성된 설명

* 코사인 유사도 (Cosine Similarity)
* 두 벡터 사이의 각도를 계산하여 두 벡터가 얼마나 유사한지 측적하는 척도
* 즉, DTM, TF-IDF, Word2Vec 등과 같이 단어를 수치화하여 표현하는 것과 다르게 문서 간 유사도를 비교하는 게 가능.
* 1에 가까울수록 두 벡터가 유사하다고 해석, 문서의 길이가 다른 경우에도 비교적 공정하게 비교할 수 있음
* 두 문서(벡터)의 유사도를 비교하는 것이므로 여러 문서 중에 가장 유사한 것을 비교하거나 가장 유사한 것들끼리 묶는 클러스터링Clustering을 수행하기 위해서는 n개의 문서가 있을 때 (n(n-1))2 만큼의 반복 연산을 수행해서 각각의 모든 문서쌍끼리의 유사도를 구해야한다.

텍스트, 폰트, 화이트, 라인이(가) 표시된 사진

자동 생성된 설명

* 피어슨 유사도(Pearson Similarity)
* 두 벡터의 상관계수(Pearson correlation coefficient)를 의미
* 유사도가 가장 높은 경우 값 = 1, 가장 낮을 경우 값 = -1
* 특정 인물의 점수 기준이 극단적으로 너무 낮거나 높을 경우 유사도에 영향을 크게 주기 때문에, 이를 막기 위해 상관계수를 사용하는 방법
* 사용자 u 와 사용자 v 간의 Pearson Similarity

텍스트, 폰트, 라인, 화이트이(가) 표시된 사진

자동 생성된 설명

* KMP 알고리즘 (Knuth Morris Partt Algorithm)
* 불일치가 발생하기 직전까지 같았던 부분은 다시 비교하지 않고 패턴 매칭(검색)을 진행하는 방법
* 접두사 / 접미사

1. 유사도 기법을 활용해 사용자와 기업의 채용 공고를 비교
2. 비교 후 계산된 비율을 활용해 취업 성공률에 대해 사용자에게 보여짐