**텍스트마이닝**

**아고다닷컴 호텔 리뷰 comment 분석**

|  |  |  |  |
| --- | --- | --- | --- |
| **버전** | **일자** | **수정내용** | **수정자** |
| **1.0** | **2024-02-11** | **Report 문서 추가** | **성호준** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. 파일 및 폴더 구성
   1. Main.py : 사용한 파이썬 환경, 패키지 정보 기록
   2. Crawling.py : agoda.com url 에서 crawling 진행하는 코드
      1. 상단 hotel\_comment\_url\_list 에서 호텔 명 및 url 목록 확인
      2. 10개 호텔, 최대 500개의 리뷰만 크롤링 수행
      3. Ex) 리뷰 800개 → 500개만 크롤링, 리뷰 30개 → 30개 크롤링
      4. Hotel\_comment\_url\_list 전체 호텔에 대해 크롤링 진행
         1. 브라우저 or 성능 이슈로 인해 에러 날 경우, 다른 호텔 주석 처리 후, 해당 호텔만 다시 수행
   3. frequencyAnalysis.py : csv 파일 로드하여 빈도 분석 진행하는 코드
      1. 매번 실행할 때마다 크롤링하는 경우, 에러 발생하거나 시간이 너무 오래 걸리므로, 크롤링하여 저장된 csv 파일을 불러와서 분석
      2. 호텔 1개 / 전체 호텔을 대상으로 분석
      3. 전체 리뷰, 긍정적 리뷰 (9점 이상), 부정적 리뷰(5점 미만) 으로 나누어 분석
      4. 호텔 1개 분석하려면 135행의 analyze\_single\_hotel('hotelData/hotel1.csv') 실행
      5. 전체 호텔 분석하려면 136행 analyze\_total\_hotel() 실행
   4. hotelData/ : 크롤링한 데이터 저장 directory

<이슈 사항 및 보고서에 들어가면 좋을 내용>

1. selenium 의 버튼 클릭 관련하여, overlay로 인한 click 불가능 에러
   1. scrolldown 이벤트 추가하여 해결
2. 다음 페이지 로드 시, 버튼 클릭 이후 데이터가 로드되지 않은 상태로 select 하여 data empty 현상 발생
   1. Selenium webdriver exception 처리
3. 분석 결과, 부정적 리뷰 분석에 장점 or 중립적인 내용이 많음
   1. 아래와 같이, 낮은 점수 + 긍정적 리뷰가 달린 경우가 많음

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

<분석 결과>

1. 호텔 전체 Comment 빈도 분석

텍스트, 스크린샷, 다채로움, 도표이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 도표, 다채로움이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 다채로움, 도표이(가) 표시된 사진

자동 생성된 설명

스크린샷, 텍스트, 다채로움이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 그래픽이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

2. 호텔 1건 Comment 빈도 분석 (켄싱턴호텔, hotel1.csv)

텍스트, 스크린샷, 도표, 다채로움이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 도표, 다채로움이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 다채로움, 직사각형이(가) 표시된 사진

자동 생성된 설명스크린샷, 그래픽, 폰트, 그래픽 디자인이(가) 표시된 사진

자동 생성된 설명스크린샷, 패턴, 그래픽, 다채로움이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 그래픽, 그래픽 디자인이(가) 표시된 사진

자동 생성된 설명