|  |
| --- |
|  |
| 『졸업작품 프로젝트』과제제안서 |
|  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **프로젝트 내용** | **희망하는**  **연구개발의 제목** | 통합 리뷰를 기반으로 한 제품 안내 어플 | | | |
| **지도교수** | 이기훈 | | | |
| **팀원** | 2 명 | | 컴퓨터공학과 2015722084 한승주  컴퓨터공학과 2015722083 김성종 | |
| **교육 신청관련 개인정보 수집․활용 및 제공에 대한 동의** | | | | | 동의여부 |
| **수집하는 개인정보 항목** | | | • 소속, 성명, 학년, 학번, 연락처, E-mail | | O |
| **개인정보의 수집 및 이용목적** | | | • “졸업작품 프로젝트” 안내를 위함. | | O |
| **개인정보의 보유 및 이용기간** | | | - 수집한 고객의 개인정보를 수요조사 기간까지만 보유하며 수요조사 완료 후 관련법규에 의거하여 안전하게 파기합니다. (개인정보보호법 시행령 제 16조)  - 정보제공자가 개인정보 수집·이용에 대한 동의를 철회할 경우 수집한 개인정보를 **즉시 파기**합니다. | | O |
| **개인정보 제공 동의 거부 권리 및 동의 거부 따른 불이익 내용 또는 제한사항** | | | 귀하는 **개인정보 제공 동의를 거부할 권리가 있으며, 동의 거부에 따른 불이익은 없음.** 다만, 추가적인 교육 서비스를 받을 수 없음. | | O |
| **※ 개인정보 제공자가 동의한 내용외의 다른 목적으로 활용하지 않으며, 제공된 개인정보의 이용을 거부하고자 할 때에는 개인정보 관리책임자를 통해 삭제를 요청 할 수 있음.**  **「개인정보보호법」등 관련 법규에 의거하여 상기 본인은 위와 같이 개인정보 수집 및 활용에 동의함.**  **2020 년  01  월  10  일**  **성명 :   한 승 주 , 김 성 종 (한승주, 김성종)** | | | | | |

|  |
| --- |
|  |
| 『졸업작품 프로젝트』 과제제안서 |
|  |

|  |  |
| --- | --- |
| **제작동기** | 1. **기존의 리뷰방식의 문제점**  * 각각의 플랫폼이나 웹사이트별로 사용하는 연령대나, 나이대 등이 다르기 때문에 실제로 제품의 정보를 하나의 사이트만 보고는 정확한 정보를 구할 수 없고, 구매했더라도 자신이 원하는 바와 다른 경우가 많다. * 실제로 국내의 경우 페이스북은 남성유저가 여성유저대비 14% 많고, 인스타그램은 여성유저가 남성유저 대비 4% 많은 비율을 가지고 있다. 또한 페이스북은 연령대가 고른 반면에 인스타그램은 20~30대 비율이 상대적으로 높다.  1. **통합 리뷰방식의 장점**  * 정확성   여러 플랫폼의 리뷰를 데이터베이스로 구축하여, 핵심 키워드만으로 해당 키워드에 맞는 제품들을 소개할 수 있고, 통합적인 리뷰를 보여줌으로써 신뢰성 있는 정보를 제공할 수 있다.   * 접근의 용이성   이를 어플리케이션(혹은 웹사이트)으로 추가적인 구현을 한다면 접근의 용이성이 증가한다.   1. **국내 현황**  * 플랫폼인 페이스북, 인스타그램등에서 볼 수 있는 광고나 리뷰, 그리고 실제로 제품을 주매하는 곳인 쇼핑몰인 G마켓, 옥션 등 판매하는 제품의 가격이나 제품의 정보를 한눈에 볼 수 있는 다나와, 네이버 비교쇼핑 등은 여러 존재하지만, 각 사이트의 리뷰를 통합적으로 파악할 수 있는 것은 없다. * 키워드를 이용하여 키워드에 맞는 항목을 표현해주는 것은 대부분의 쇼핑몰이나 플랫폼에서 구축한 DB에서 해당 항목을 보여주는 것이 대부분이고, 리뷰에 대한 키워드를 보여주는곳은 상대적으로 많지 않다. * 따라서 다양한 플랫폼에서의 제품에 대한 리뷰를 한눈에 볼 수 있다면, 또한 특정 키워드만 추출해서 간단하게 표현할 수 있다면 소비자의 입장에서 제품을 더 정확하고 손쉽게 판단하여 선택하고 구매할 수 있을 것이다. |
| **Flowchart** |  |
| **수행일정** |  |

|  |  |
| --- | --- |
| **개념설계** | 1. **플랫폼의 웹사이트 코드 분석 및 크롤링(Beautiful soup)**  * **DB를 구축하기 위해서 해당 웹사이트에 접속하여, 실제 리뷰나 정보가 담겨있는 웹사이트 코드를 분석** * **BS4를 이용한 페이지의 데이터를 호출하고, 제품의 이름, 가격, 리뷰 등의 정보의 tag를 찾아 추출한다.** * **각 사이트 별로, 여러 페이지가 존재한다면, 링크를 재귀적으로 검색하여 데이터를 추출한다.**   **Ex)**  … 생략 …  def get\_product\_info(box):      …  def get\_page\_products(url):      …  urls = [      “http://joise.com/category/tonermist/43/?page=1”,      “http://joise.com/category/tonermist/43/?page=2”      “http://joise.com/category/tonermist/43/?page=3”      “http://joise.com/category/tonermist/43/?page=4”      “http://joise.com/category/tonermist/43/?page=5”  ]  for page\_number in range(0,5):      page\_products= get\_page\_products(urls[page\_number])      print(page\_products)  결과  -------------------------------------------------------------  [{‘name’: “It’s skin Cera Routine Moisturizer 150ml”, pri … 생략 …  [{‘name’: ‘may COOP Raw Sauce 150ml’, ‘price’: ‘USD 37.65’, … 생략 …  [{‘name’: ‘The FACE SHOP Dr.Belmeur Daily Repair Toner 200ml’, … 생략 …  …   1. **크롤링된 정보를 이용한 DB구축(Mysql)**  * **MySQL서버에 접속하여 CREATE명령어를 이용하여 데이터베이스 생성** * **Cursor를 추출하여 execute 메서드로 SQL을 실행하여 테이블을 생성** * **Execute 메서드의 INSERT 구문을 이용하여 데이터를 계속 확장**   **Ex)**  import MySQLdb  #데이터베이스 연결하기  connection = MySQLdb.connect(  user=”scrapingman”,  passwd=”myPassword-1”,  host=”localhost”,  db=”scrapingdata”,  charset=”utf8”)  # 커서 생성 및 실행할 때마다 다른결과가 나오지 않게 테이블 제거  cursor=connection.cursor()  cursor.execute(“DROP TABLE IF EXISTS user\_id”)  … 생략 …  **Ex)**     1. **Kkma, Hannanum을 이용한 KoNLP(키워드 생성)**  * Kkma나 Hannanum 모듈을 이용하여, 해당 모듈에 맞추어 입력된 문자열에서 키워드로 표현할 품사를 추출하고, 가장 빈도수가 높은 단어를 DB에 저장하여, 이를 키워드로 표현하고자 한다. * **문자열의 형태소 분류(Hannanum)**   N : 체언 (고유명사 대명사 의존명사 수사 등이 포함)  J : 관계언 (격조사가 포함)  X : 접사 (접두사 접미사 등이 포함)  E : 어미 (연결어미 종결어미 등이 포함)  P : 용언 (동사 형용사 등이 포함)  S : 기호   * **Pos를 이용한 문장 분할 및 추출 및 def를 이용한 단어 카운트**   **Ex)**  0 대한민국헌법 유구한 역사와  9 대한국민은 3·1운동으로 건립된 대한민국임시정부의 법통과 불의에  98 총강 제1조 ① 대한민국은 민주공화국이다. ②대한민국의  100 ① 대한민국은 민주공화국이다. ②대한민국의 주권은 국민에게  110 나온다. 제2조 ① 대한민국의 국민이 되는  126 의무를 진다. 제3조 대한민국의 영토는 한반도와  133 부속도서로 한다. 제4조 대한민국은 통일을 지향하며,  147 추진한다. 제5조 ① 대한민국은 국제평화의 유지에  787 군무원이 아닌 국민은 대한민국의 영역안에서는 중대한  1836 파견 또는 외국군대의 대한민국 영역안에서의 주류에  3620 경제 제119조 ① 대한민국의 경제질서는 개인과  [0, 9, 98, 100, 110, 126, 133, 147, 787, 1836, 3620]  가장 빈도수 높은 단어 10개 가져오기 (키워드 생성)  [('저', 155),  ('법률', 121),  ('수', 88),  ('대통령', 84),  ('국가', 73),  ('국민', 69),  ('헌법', 69),  ('조', 58),  ('국회', 55),  ('때', 55)]   1. **Android UI제작**  * **자신이 구매하고자 하는 제품의 리뷰를 보기 위한 제품의 검색창과 위에서 나타낸 키워드 중 전체 제품에서 가장 많은 비중을 차지하는 몇 개의 키워드를 다음과 같이 제품 검색창 아래에 표현하여, 키워드를 통한 제품을 검색할 수 있게 추가적인 구현**      * **만약, 제품에 대한 정보가 DB에 존재할 때, 다음과 같이 제품의 가격정보와 리뷰의 최다 빈도 키워드를 작성하고, 실제 리뷰를 사이트 별로 보임** |
| **기대효과** | **[기업적 측면]**   * 제품에 대한 좋은 리뷰를 많이 보이는 제품의 경우에는 신뢰성 있는 리뷰를 통해 구매력이 증가할 것이다. * 해당 제품의 실제 이용고객이 제품을 사용하고 실제로 좋은 제품이라면 회사의 신뢰성이 증가하여 차기 출시될 제품의 신규 구매력도 증가할 것이다.   **[사용자 측면]**   * 여러 플랫폼을 찾아 검색하지 않더라도 제품을 검색했을 때 여러사이트의 리뷰를 한번에 파악할 수 있다.   리뷰에서 자주 언급된 단어를 통하여 신뢰성 있는 리뷰를 파악할 수 있다. 또한, 이러한 키워드에 맞는 제품을 추가적으로 검색할 수 있다. |
| **참고문헌** | https://kiwiprnad.blog.me/221458922512 [[페이스북과 인스타그램 유저의 인구통계학적 특성](https://blog.naver.com/kiwiprnad/221458922512)]  https://blog.naver.com/zzang9ha/221692276477 [Python] MySQL DB연동(기본 crud 테스트)  https://blog.naver.com/meibee55/221703098261 [Python] KoNLP 자연어 처리 (Kkma, Hannanum)  파이썬을 활용한 클로러 개발과 스크레이핑 입문, 2019, 카토 카츠야, 요코야마 유우키, 위키북스  파이썬 데이터 수집 자동화 한방에 끝내기 한입에 웹크롤링, 2018, 김경록, 서영덕, 비제이퍼블릭 |

**본인은 상기의 내용이 모두 사실임을 확인하며 『졸업작품 프로젝트』를 제안합니다.**

2020년   01  월 10  일

신청인    성명    한승주, 김성종          (한승주, 김성종)