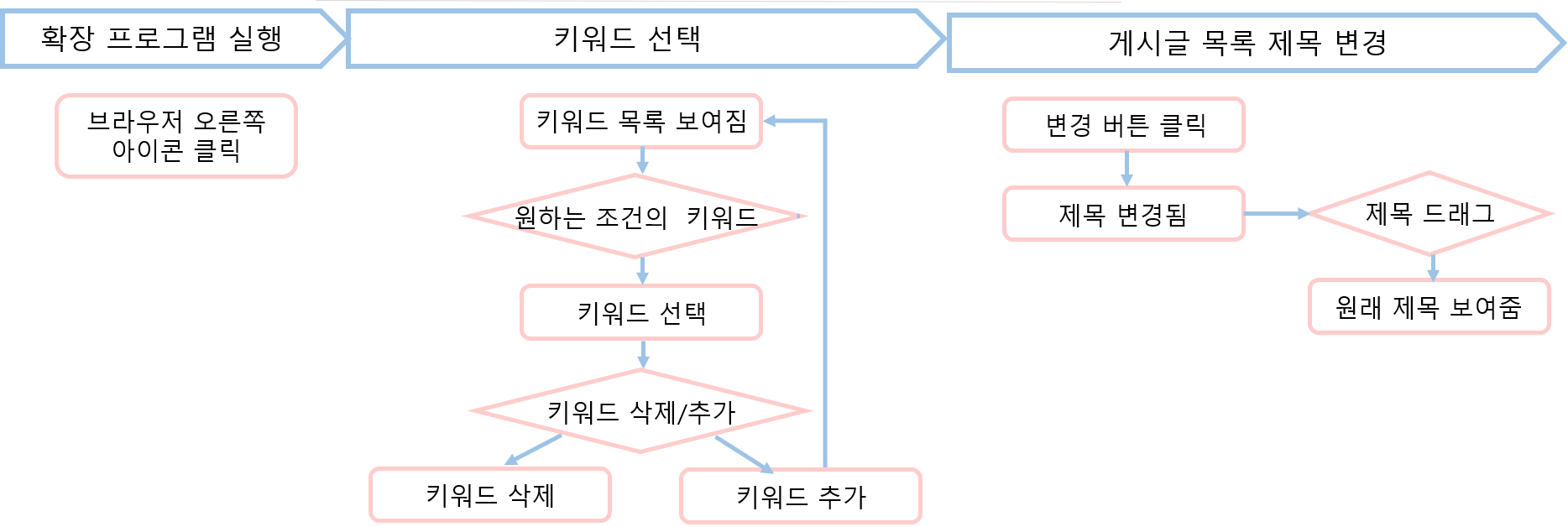
**주제: 웹 브라우저 상 커뮤니티 거래를 위한 제목 변경 확장 프로그램**

**1. 주제 설명** (제목변경 확장프로그램-화면 예시.pptx 같이 봐주세요!)

* 커뮤니티에서 거래나 중고거래 게시판은 대부분 제목의 형식이 통일되어 있지 않다. 제목으로 상품의 가격, 색, 상태 등을 알 수 없으며 심지어는 상품명조차 없는 경우도 있다.
* 우리 프로젝트는 이를 보완하기 위한 웹 브라우저 용 확장 프로그램이다. (웨일 브라우저 예정)
* 사용자는 제목에 있길 원하는 키워드(예: 상품명, 가격, 거래방식, 색상 등)를 선택한다.
* 게시글 제목과 본문에서 선택된 키워드에 대한 정보를 얻은 후 제목을 변경해준다.

**2. 프로세스**



**3. 대상 사용자**

* 웹 브라우저를 사용해 커뮤니티 상에서 상품의 거래(혹은 중고거래)를 하는 사람.
* 웹 브라우저와 브라우저 확장 프로그램을 사용할 수 있는 사람

**4. 페르소나**

* 윤연경 (나이: 22, 여, 직업: 대학생, 웹이용능력: 상) : 대학교 수업을 들을 때 필기용으로 사용할 아이패드 구매를 예정에 두고 있다. 용돈을 받는 대학생이라 새 상품보다 저렴한 중고 제품으로 구매하고 싶어 중고거래를 하는 커뮤니티를 탐색 중이다. 그런데, 게시판에 “아이패드”를 검색했을 때 게시글 제목만 보고 아이패드와 관련 없는 글, 광고성 글, 원하는 조건이 아닌 글이 선택되는 경험을 많이 했다. 또, 게시글마다 원하는 조건의 정보들을 다 따로 기록해야 해 시간이 오래 걸려 구매가 하루하루 늦춰지고 있는 상황이다.
  + 원하는 키워드로 보이게 변경하면 게시글 제목만 보고 본문을 읽을 가치가 있는 지(광고이거나 상품에 대한 정보가 있는지)를 더 정확하게 판단할 수 있게 된다.
  + 같은 상품에 대한 조건을 제목만 보고도 알 수 있어 게시글 목록을 보고 상품을 비교할 수 있다.
* 김민철(나이: 56, 남, 직업: 은행원, 웹이용능력: 중): 매일 앉아서 근무하는 김민철씨는 나이가 들수록 운동의 필요성을 느낀다. 퇴근 후에도 집에서 쉽게 운동하며 습관을 들이기 위해 집에 둘 만한 운동기구를 찾고 있다. 운동기구가 생각보다 비싸 중고로 알아보기로 결심을 하고 중고 거래에서 제일 유명한 커뮤니티라 할 수 있는 중고OO에서 알아보기 시작했다. 운동기구에 대해 잘 알지 못해 각 운동기구 종류가 가격대가 어느 정도인지 알지 못해 게시글에 들어갔다가 가격을 보고 놀라 나오기를 반복했다. 운동기구명이나 종류와 함께 가격도 제목에 같이 써준다면 이런 수고를 덜어줄 것 같다고 생각해 ‘가격’이라는 단어로 검색했지만 ‘가격’이란 단어가 제목이나 본문에 있는 게시글 수가 적어 선택의 폭이 좁아지는 점도 마음에 들지 않는다.
  + 검색을 상세히 하지 않고도 제목을 보고 판단이 가능해 검색어 추가, 상세조건 검색으로 잘못 필터링되는 글들을 볼 수 있다.
  + 제목을 통해 원하는 조건에 맞는지 판단할 수 있다.
* 이준호(나이: 30, 남, 직업: 공무원, 웹이용능력: 상): 최근 공무원 시험을 통과한 이준호씨는 발령지로의 이사를 앞두고 불필요한 물건 정리에 나섰다. 시간이 촉박해 여러 커뮤니티에 중고 거래 글을 게시하고 있는데도 관심을 보이는 사람이 없다. 사람들이 관심을 가질 만하게 제목을 써보려 하지만 과장된 제목은 제제를 받을 수 있고, 하루에 같은 글을 몇 번 이상 올려도 제제를 받는 규칙이 있는 커뮤니티도 있어 혼란스럽다. 사람들이 원하는 조건을 제목에 넣고 싶지만, 사람들마다 원하는 조건이 다 다르고 그것을 모두 넣기엔 제목이 길어져 가독성이 떨어지는 것이 난관이다. 어떻게 제목을 결정해야 하는지 중고판매의 고수의 특별 과외라도 받고 싶은 심정이다.
  + 판매자가 고려하지 않고도, 구매자는 원하는 조건이 담긴 제목을 볼 수 있다.
  + 판매자는 본문에 정보만 다 넣어 놓는다면, 제목을 고민하지 않을 수 있다.

**5. 기존 커뮤니티 사용자의 불편함**

* 제목이 너무 중구난방이다.
* 제목만 보고 게시글 내용을 판단하기 힘들다. 🡪 본문의 내용이 맘에 들지 않는 경우가 훨씬 많았다.(헛수고가 많다.)
* 작성자가 다양하기 때문에 게시글 제목 형식이 같기 힘들다.
* 커뮤니티에서 제목 형식을 규정하기엔 기존에 있는 글에 대해선 적용되지 않고, 커뮤니티 이용자(작성자)들이 불편하며, 커뮤니티 이용자(구매희망자)들 마다 원하는 조건을 제목 형식에 다 넣을 수 없다는 문제점이 있다.

**6. 구현 기능**

1. 사용자가 제목에 보여지길 원하는 키워드 선택
2. 검색된 페이지 크롤링 및 자연어처리 진행
3. 키워드를 포함한 새로운 형식의 제목 나타내기
4. 제목 드래그 시 이전 제목 나타내기

**7. 구현 방법**

웹크롤링

* HTML 페이지를 가져와서 HTML/CSS를 파싱한 후, 필요한 콘텐츠를 수집
* BeautifulSoup 라이브러리 사용
* <https://www.fun-coding.org/crawl_basic2.html>

자연어처리 과정

1. BeautifulSoup를 사용해서 웹사이트 크롤링 후 텍스트 추출
2. 데이터 전처리
   1. 토큰화 (Tokenization) – 문자열을 단어로 분리
   2. 불용어 제거 (Stopword) – 문서의 특징을 표현하는데 불필요한 단어 삭제
   3. 어간 추출 (Stemming) – 단어의 기본 형태 추출
   4. 문서 표현 (Representation) – 단어의 빈도수로 문서를 표현
3. 텍스트 파싱

자연어처리 서비스

* 아마존 컴프리헨드
  + 비정형 텍스트에서 주요 문구, 장소, 사람들의 이름, 브랜드, 이벤트, 정서를 추출하는 자연어처리 서비스
* 구글 클라우드의 오토머신러닝
  + 텍스트에서 구문, 엔티티, 정서를 찾아내고 텍스트를 사전 정의된 범주 집합으로 분류
* 코엔엘파이 (KoNLPy)
  + 한국어 정보처리를 위한 파이썬 패키지
  + <https://devtimes.com/bigdata/2019/04/18/konlpy/>
* NLTK
  + 영어 텍스트의 자연어처리를 위한 파이썬 라이브러리
* Gensim
  + 한국어 및 다양한 언어 지원
  + Topic modeling, corpus, word embedding 모델 지원
  + <https://aileen93.tistory.com/128>

**8. 사용 언어**

* Python 3.7

비고

수업 조교님들로부터 필요성 측면에서 이전 재알람 어플리케이션 아이디어보다 더 좋게 피드백을 받아서 이 주제로 선택하였습니다.

자연어 처리에 대한 부분은 기술적으로 어느 정도까지 가능할 지 잘 모르겠습니다. 저희가 회의한 결과로는 “모든 상품에 적용 가능한 키워드(예: 상품명, 브랜드, 가격, 거래방식, 색상 등)”로 일단 한정해서 구현할 생각입니다. 바꿔 말하자면, 핸드폰의 용량, 컴퓨터 본체의 부품(CPU, RAM 등)의 사양과 같은 부분은 여유가 되면 추가할 계획입니다.

웹 브라우저 Whale의 확장 프로그램으로 구현할 계획입니다. Whale 확장 앱 콘테스트에도 활용할 생각으로 Whale로 개발하려고 합니다.