|  |
| --- |
| **1. 주제**  **실시간 검색어 차트가 사라진 요즘 새로운 방식의 검색어 차트**  **(가)반, 6번, 20180376 안제준** |

|  |  |
| --- | --- |
| **2. 요약**  네이버의 실시간 차트 서비스가 사라져 빠르게 변화하는 요즘 시대와 사회적 이슈를 인지하지 못해 사회에서 도태되고 불편을 느끼는 사람들이 많아졌습니다. 그래서 저는 새로운 방식의 검색어 차트를 만들고자 합니다. 실시간으로 올라오는 인터넷 기사의 제목을 크롤링 하여 키워드를 뽑아 해당 키워드가 들어간 기사의 수로 순위를 매겨 그것을 차트로 보여주는 방식의 차트를 만들고자 합니다. 또한 해당 차트는 위젯 방식으로 스마트폰에 보이며 접근성을 높일 생각입니다. 기존에 있던 실시간 차트의 문제점이었고 사라진 원인 중 하나인 검색어 조작을 해결할 수 있고 사람들이 사회적 문제와 이슈를 빠르게 인지하여 현재 느끼고 있는 불편함을 해결하고자 합니다. | **3. 대표 그림 (1개 이상, 10점)**  실시간 차트 |

\* 표지 없이 1(주제), 2(요약), 3(대표 그림), 6번(결론) 합하여 1장 이내

|  |
| --- |
| **4. 서론 (1장 이내)**  **최근 일어난 kt 인터넷 먹통와 페이스북, 인스타그램 먹통등 여러가지 문제가 발생하면서 즉각적으로 현재 발생한 이슈를 빠르게 인식하지 못해 실시간 검색어 차트를 부활하자는 여론과 실시간 검색어 폐지 후 불편을 느끼는 사람이 많아졌습니다. 또한 바쁘게 돌아가는 사회에서 능동적으로 정보를 검색하고 찾는 것에 불편을 느끼고 정보를 제공받기를 원하는 사람이 많다. 그런 사람들은 실검을 대체하는 사이트를 찾기도 합니다. 하지만, 현재 유지되고 있는 대부분의 실시간 차트는 기사 하나와 링크되어 실시간으로 많이 보는 그 기사만 보고 판단하는 일이 주위에서 빈번하게 발생합니다.**  **실시간 검색어 차트가 사라진 이유들을 살펴보면 실검을 조작하거나, 정치적으로 악용, 신뢰성 논란, 기업의 광고성 문구가 자주 등장하면서 광고판 같다는 이유들이 있습니다. 하지만 실시간 검색어 차트가 사라진 현시점에서 보면 전반적인 세상돌아가는 것을 알 방법이 줄어 불편을 느끼거나 그나마 정보를 얻기 쉽고 접하기 쉬운 페이스북이나 인스타 같은 sns에서도 이슈의 진위여부도 파악되지 않은 그저 빠르게 가져온 잘못된 기사나 흔히 말하는 알고리즘으로 편향된 정보를 접하는 일이 발생합니다.**  **그렇다면 차트를 구성하는 방식을 기사의 제목에서 포함되어 있는 키워드들을 뽑아 해당 키워드를 포함한 기사의 수를 바탕으로 랭킹을 매긴다면 기사의 수라는 직관적인 숫자가 포함되어 신뢰성은 보다 높아지고 조작 가능성은 줄일 수 있을 것입니다. 그리고 해당 키워드를 포함하는 모든 기사를 보여준다면, 사람들은 다양한 시각에서 그 이슈를 접할 수 있고 조금 더 다양한 정보를 받을 수 있을 것입니다. 또한 차트를 위젯으로 스마트폰에 보여주면 보다 쉽게 접근할 수 있게 접근성도 확보할 생각입니다.** |

|  |
| --- |
| **5. 본론 (1장 이내)**  **back- end는 우선 이전 지난 1년 정도의 기사들 헤드라인만 모아 크롤링해서 데이터 데이스를 구성합니다. 데이터가 많으면 많을수록 머신러닝한 프로그램의 정확도가 올라갈 것 같아 일단 1년간의 기사들을 이용해서 Embedding layer로 임베딩을 시키고 LSTM 모델을 쌓고 Dense layer을 통해서 모델을 만듭니다. 그리고 TfidVectorized와 LogisticRegressin으로 머신러닝 프로그램을 구성합니다. 이제 머신러닝이 끝나고 어느 정도 정확도가 맞는다면 파이썬의 beautiful soup 크롤링 패키지를 이용해 기사의 헤드라인을 크롤링해서 tsv파일로 저장합니다. 그 후 머신러닝된 프로그램을 이용해 결과값을 추출해 데이터 프레임에 저장한 뒤 파이썬 딕셔너리로 해당 단어가 포함된 기사의 횟수를 카운트해서 랭킹을 매기는 구조입니다.**  **front-end는 안드로이드 스튜디오를 이용해 어플을 만듭니다. 랭킹이 매겨진 각 키워드에는 하이퍼텍스트 링크를 걸어 해당 키워드로 포털 사이트에 검색하고 해당 키워드에 관한 정보를 제공합니다. 그리고 위젯 형식으로도 개발하여 어플을 실행시키지 않아도 보여지는 방식으로 어플을 만듭니다.**  **해당 검색어 차트는 매일 00시에 리셋되며 포털 사이트에 인터넷 뉴스 기사가 올라오면 즉각적으로 프로그램이 반응하여 해당 기사의 제목을 크롤링하고 키워드를 추출한 뒤 순위를 보이는 것까지 개발해 보고 싶습니다. 여기서 조작을 방지하기 위해 같은 기자가 같은 제목으로 기사를 쓴다면 그 기사는 카운팅 되지 않습니다.** |

|  |
| --- |
| **6. 결론**  **현재 실시간 검색어가 사라지고 본인 포함 불편을 느끼는 사람이 많아졌습니다. 실시간 검색어가 사라진 이유를 알아보면 조작에 관한 신뢰성 문제 그리고 기업들이 광고문구를 실검에 올려 광고판 같아졌다는 이유들이 있었습니다. 그렇다면 실검의 데이터 베이스를 얼마나 많이 그 검색어를 검색했는지가 아닌 실제 올라온 기사들을 바탕으로 실검을 구성하고 그것과 관련된 기사들의 수를 직관적으로 표현하고 그 키워드가 포함되어 있는 기사들을 모두 보여준다면 기존 실검의 문제들을 해결할 것 같습니다. 물론 기자들이 내용이 없고 양만 채우기용 기사를 마구잡이로 쓰고 기사의 질이 떨어진다면 또 다른 문제가 생길 수 있습니다. 이것은 향후 토론 후 그런 기사들은 어떻게 처리할 것인가 생각해 보면 좋을 것 같습니다.** |

**7. 출처**

https://lsjsj92.tistory.com/409