

대표 프로젝트를 소개해주세요.

프로젝트명	인기검색어 TOP10 배치 개발
프로젝트기간	2020.05 ~ 2021.02
참여인력	검색TF 개발자2명
사용한 기술스택	ElasticSearch, Elasticsearch Rest High Level Client Java, Spring Boot, Gradle, MyBatis, MySQL
프로젝트 소개 *	
<p>홈플러스 프론트에서 가장 많이 검색되어 지는 TOP10 키워드를 노출합니다.</p> <p>[설계]</p> <p>인기검색어는 실제 고객이 입력한 검색어를 대상으로 노출해야 합니다. 완전 자동화로 인기검색어가 적용되도록 하고자 하였지만, 화면에 노출돼서는 안되는 검색어(금칙어, 욕설 등)를 노출할 수는 없으므로 어드민에서 기본적으로 인기검색어를 컨트롤 할 수 있도록 확정어와, 제외어를 추가하여 처리하도록 하였습니다.</p> <p>또한 검색어를 형태소분석하여 토큰값의 포함 여부로 같은 키워드의 여부를 판별하려 했으나, 홈플러스의 검색 특성상 문장으로 검색하는 경우는 없으므로 띄어쓰기와 특수문자만을 필터링하여 검색어를 비교하는 것으로 하였습니다.</p> <p>[통계]</p> <p>어드민에서 일간, 주간, 월간으로 어떤 검색어가 상위로 노출되었는지를 알기 위해 인기검색어 배치에 통계기능을 추가하였습니다. 통계는 인기검색어 로직에서 확정어, 제외어의 처리를 제외한 순위를 DB에 넣음으로써 확인할 수 있도록 하였습니다.</p>	
전체 시스템구성 *	
<ol style="list-style-type: none">1. Jenkins의 timer에 의해 인기검색어 프로세스 동작2. (NOW - 2시간) ~ NOW의 범위로 검색되는 키워드를 ES 집계쿼리로 요청3. 띄어쓰기, 특수문자를 제외하였을 경우 중첩되는 키워드의 검색수 처리4. 요청 데이터에 제외어 필터링 및 순위 재설정5. 확정어 순위를 삽입함으로써 적용 인기검색어 완성6. DB에 저장된 이전의 인기검색어의 순위와 비교하여 인기검색어의 등락폭을 결정(이때 확정어로 들어온 등락폭은 빈값)7. 인기검색어를 ES에 색인하고 alias값을 교체8. 새로운 색인 인덱스로 서비스 시작	

프로젝트에 기여한 내용 *

[확정어, 제외어 DB 설계]

어드민 페이지에서 관리되도록 확정어, 제외어 테이블 설계. 인기검색어와, 인기검색어 통계로 Type을 나뉘어서 저장

[ES 인기검색어 Query작성]

성공 검색로그에 시간범위에 따른 집계를 요청하는 ES Query 작성

[통계 범위 로직 개발]

통계의 주간및 일간 처리시 시간범위가 이전월을 포함하는 경우, 대상 인덱스와 일정범위를 계산하는 방안 구현

프로젝트 성과 *

인기검색어의 확정어와 제외어의 기능은 운영 서비스가 시작될 때 빛을 발하였습니다. 최초 실행 시에는 쌓이는 로그가 얼마 없고, 제대로 된 통계를 낼 수 있는 검색어가 아니었습니다. 기존에 쌓였던 인기검색어 순위를 확정어에 등록함으로써, 운영 서비스 오픈 시에도 제대로 된 인기검색어를 서비스할 수 있었으며, 서비스의 점검이 있어서 2시간 전의 검색어를 가져올 수 없는 경우에도 확정어를 사용하여 서비스에 문제가 없도록 처리될 수 있었습니다. 통계에서는 기획자가 일간, 주간, 월간 검색어를 분석하여 상품의 검색결과가 더욱 잘 노출될 수 있도록 검색키워드를 사전에 추가하였습니다.

트러블 슈팅 경험 *

[문제정의]

기능추가 및 변경 이후 테스트가 원활히 이루어지지 않고 실패

[사실 수집]

테스트 시에 not_found_index 예외가 발생하면서 프로세스 종료

[원인추론]

테스트 시에는 개발서버에서 진행하므로 기간에 해당하는 index를 가져올 때, 검색어가 존재하지 않아 전월 또는 금월에 index가 만들어지지 않음

[조치 방안]

테스트 기능을 추가, 설정파일에 원하는 기간의 범위를 지정하고, Jenkins 빌드시에 -TEST를 붙여 수행하면, 현재시간 기준으로 검색어를 조회하지 않고, 설정파일에 정의한 시간을 기준으로 검색어를 조회가 가능하도록 변경

아쉬운 점 *

인기검색어 설계시, 형태소분석을 통한 토큰 분석 방식으로 결정이 되었다면, 해당 프로젝트를 통해 조금이나마 커스텀된 아리랑 형태소분석에 대해 좀 더 자세히 알 수 있었을 것으로 생각합니다.

참고자료

Github - <https://github.com/mertyn88>

Blog - <https://velog.io/@mertyn88>

Portfolio - <https://mertyn88.github.io/>