알고리즘 적용 기획서

HappyHouse14_Algorithm_YoonJongMyeong_LeeJaeUng

#1

내용

KMP 알고리즘을 활용한 검색 서비스

적용 알고리즘

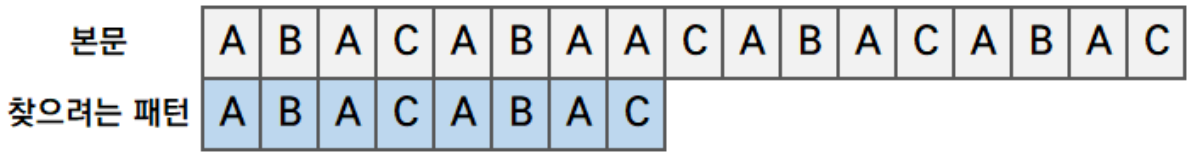
KMP 알고리즘

알고리즘 개요

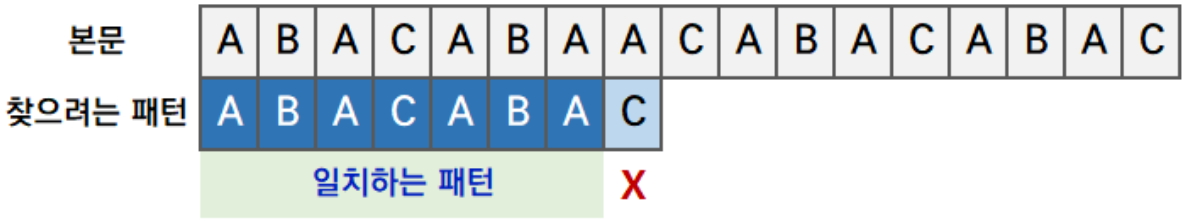
KMP 알고리즘은 비교한 문자열을 오른쪽으로 한 칸씩 옮기지 않습니다. 비교하는 도중에 얻은 정보를 활용해 패턴의 몇 번째 문자까지 텍스트와 일치했는지 확인한 후, 텍스트를 기준으로 패턴의 위치를 몇 칸 옮길지 정합니다.

패턴에 포함된 각 문자와 텍스트가 일치하지 않을 때 패턴을 몇 칸 옮기는 것이 최선인지 조사하고 그 결과를 테이블로 만듭니다. 이 테이블을 **이동 테이블**이라고 부릅니다. 그래서 검색할 때는 이동 테이블을 바탕으로 패턴을 몇 칸 옮겨야 할지 정합니다.

> (1) 본문이 되는 문자열(비교할 대상)이 있고, 찾으려는 패턴을 처음에 둡니다.



> (2) 일부가 일치하지 않는다면, 불일치한 부분을 제외한 직전까지 일치하는 패턴에서 최대 길이의 경계를 찾습니다.



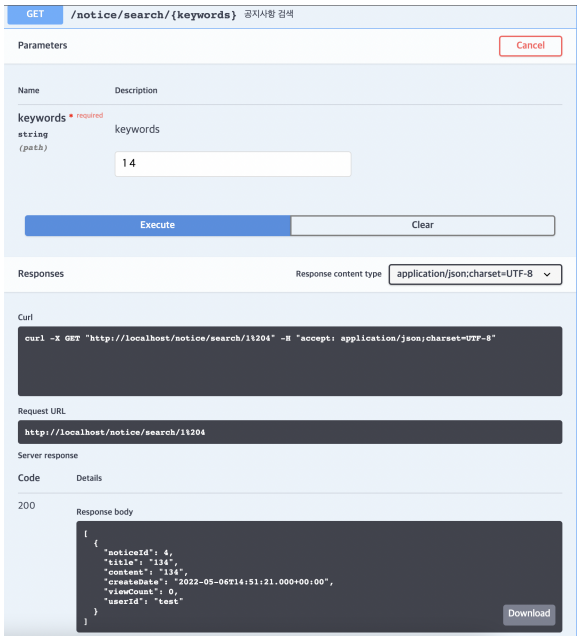
시간 복잡도는 검색할 문자열의 길이를 N, 탐색할 문자열의 길이를 M이라고 할때 O(N+M)이 소요됩니다.

적용 서비스

- 지역, 아파트 검색 서비스
- 공지사항 검색 서비스

적용 서비스 개발 개요

- 검색창에 키워드 검색
  1. search/{keyword} api 호출
  2. keyword 파라미터를 split 메소드를 이용해 공백으로 구분하여 키워드 분리
  3. 각 키워드의 이동 테이블 생성
  4. 테이블의 모든 이름과 id들을 리스트에 저장
     - 모든 데이터들을 불러올 시 메모리가 부족할 수 있기 때문에, 검색에 필요한 데이터만 불러온다.
  5. 리스트의 각 원소를 돌면서 일치하는 부분 문자열이 모두 있는지(AND 연산) 확인
  6. 일치하는 부분이 있는 데이터를 새롭게 쿼리하여 결과 리스트 생성
  7. 결과 리스트 반환



#2

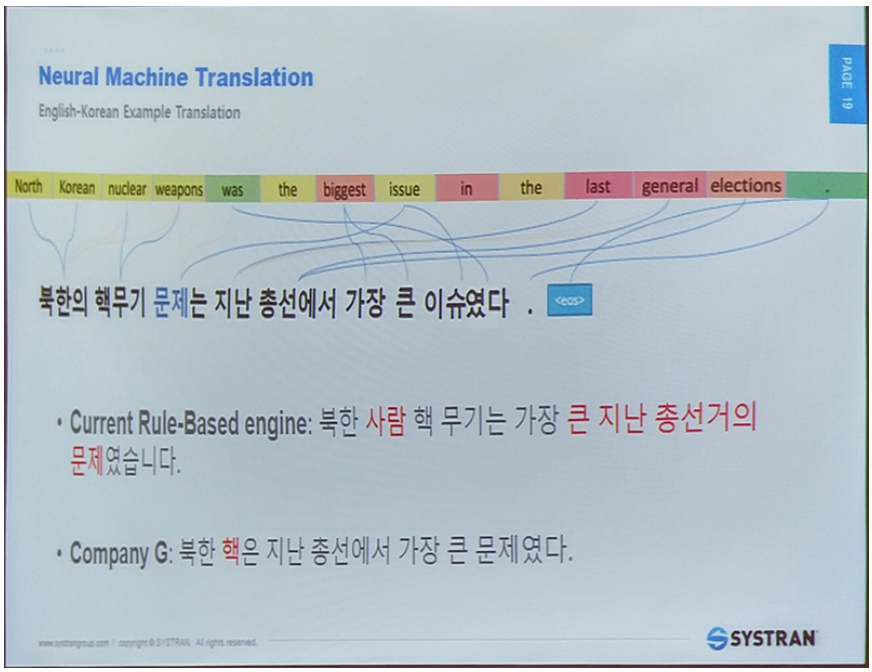
내용

NMT를 활용한 영어 번역 서비스

알고리즘 개요

NMT(Neural Machine Translation)은 인공지능 분야 중 하나인 인공신경망을 번역 기술에 적용한 것으로, 인공신경망은 인간이 생각을 하는 최소 단위인 뉴련의 집합체(신경망)를 소프트웨어적으로 구현한 것을 말합니다. 즉 인간의 뇌와 유사한 구조를 소프트웨어로 만들고, 이 뇌에 번역 능력을 학습시키는 것입니다.

NMT는 기존 통계기반 자동번역의 미흡한 점을 보완하는 번역 기술입니다. 인공지능을 통해 문장을 분석합니다. 단어나 구 단위로 구분하는 것이 아니기 때문에 문장의 전체 맥락을 더 잘 이해하고, 인간과 비슷한 수준의 번역 결과를 낼 수 있습니다. 예를 들어 한국어와 영어를 변환할 경우 기존 SMT는 오역이 심합니다. 두 언어는 어순이 다르기 때문에 단순한 단어나 구를 번역한다고 해서 우리가 이해할 수 있는 문장을 만들기는 어렵다는 의미입니다. 이와 달리 NMT는 인공지능을 통한 학습이 가능하기 때문에 새로운 언어를 배우는 인간의 뇌처럼 문장의 의미를 학습하고 자연스러운 번역 결과를 만듭니다.



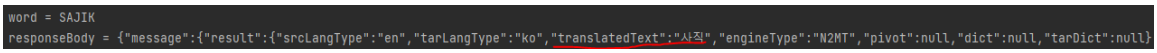
적용 서비스

아파트, 동 이름 영어 검색 시 한글 번역 서비스

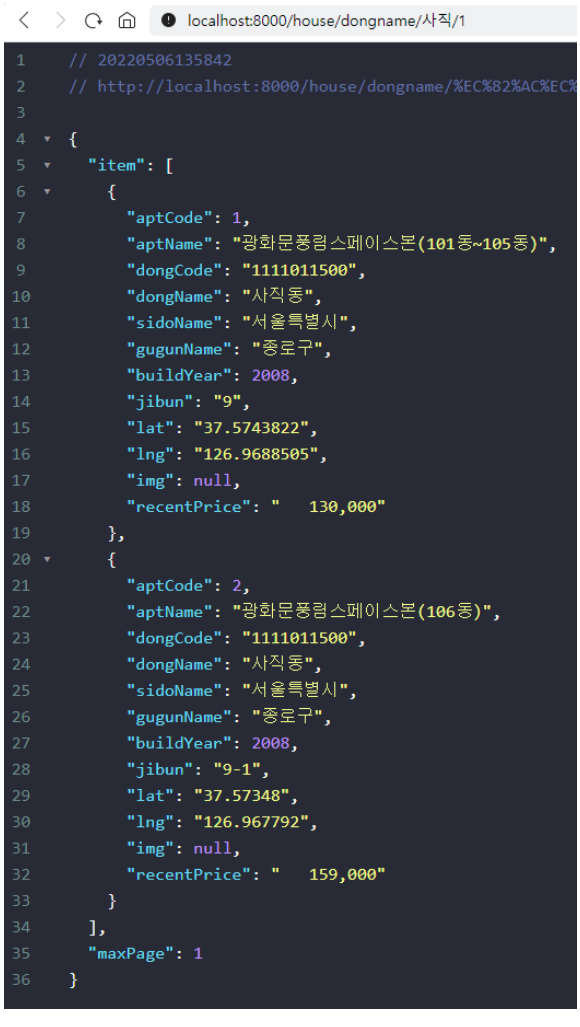
적용 서비스 개발 개요

💡 인공신경망을 직접 구현하기는 시간상 어렵기 때문에 파파고 API를 활용하여 번역을 적용하였습니다.

1. 검색 키워드를 HttpUrlConnection을 이용하여 파파고 API로 request 전송
   ex) http://localhost:8000/house/dongname/sajik/1
   a. {sajik} 검색시 영어를 판별하여 한글로 변환
2. API Response로 받은 번역결과를 활용하여 데이터베이스 쿼리 실행



3. 결과를 검색 결과창에 출력



한글 검색, 영어 검색 결과가 동일한 것을 확인할 수 있습니다.

#3

내용

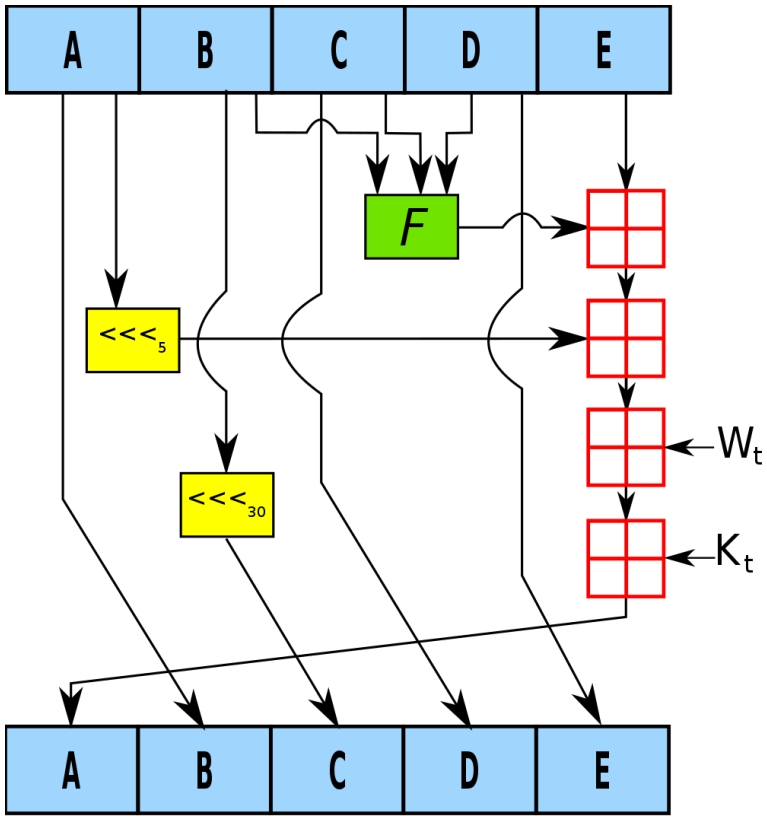
해시 알고리즘을 활용한 비밀번호 암호화

적용 알고리즘

SHA-256 암호화 알고리즘

알고리즘 개요

SHA 암호 해시 함수는 디지털 데이터 상에서 수학적으로 동작하며 알려져 있고
예측된 해시값에 대해 계산된 해시(알고리즘의 실행 출력)를 비교함으로써 사람이 데이터의 무결성을 파악할 수 있게 됩니다.
이를테면 다운로드한 파일의 해시를 계산한 다음 이전에 게시한 해시 결과물의 결과와 비교하면 다운로드한 파일이 수정 또는 조작되었는지 알 수 있습니다.
또한 비밀번호의 해시를 계산하여 저장하면 데이터베이스가 공격을 받아 유출이 되어도 해시값으로 되어있어 원문을 유추할 수 없습니다.
로그인 과정에서 비밀번호를 확인 할 경우 입력된 비밀번호가 같다면 해시값이 동일한 값이 나오기 때문에 해시값 비교를 통해 사용자의 비밀번호를 안전하게 확인할 수 있습니다.



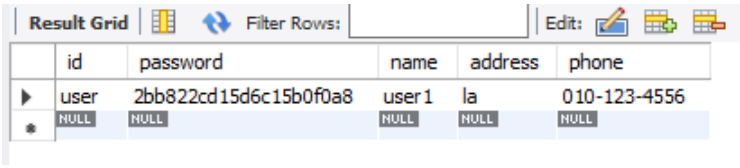
적용 서비스

- 회원가입시 비밀번호 Hash 알고리즘 적용하여 비밀번호 저장
- 로그인 시 비밀번호 Hash 알고리즘 적용하여 비밀번호 일치하는지 확인
- 사용자 개인정보 보안 강화

적용 서비스 개발 개요

1. 해쉬 함수 구현
    a. 입력값을 바이트어레이로 변환
    b. 변환된 값을 통해 SHA-256 암호화 진행
    c. 암호화된 바이트어레이를 16진수 스트링으로 변환
2. 회원가입시 해시함수를 호출하여 해싱된 비밀번호를 DB에 저장
3. 로그인 시 입력한 값에 해쉬함수를 적용하여 DB에 저장된 값과 비교

평문값 전달



암호화 값 저장