

유사 발음 상표명 분석을 위한 영문 동음어 생성 방법

¹조상현, ¹권혁철, ²문경혜, ²정상영, ³탁원준, ³이정훈*

부산대학교, ²인텔라이즈, ³동의대학교

delosycho@gmail.com, hckwon@pusan.ac.kr, iromoon@gmail.com, mycampground001@gmail.com, taku@office.deu.ac.kr, *it_leejh@deu.ac.kr

Methods for Generating English Homonyms for Analysis of Trademark Names with Similar Pronunciation

Sanghyun Cho¹, Hyuk-Chul Kwon¹, ²Kyung-hye Moon, ²Sang-young Jeong, ³Won-Jun

Tak, ³Jung-Hun Lee*

¹Pusan National Univ., ²Intelights *³Dong-Eui Univ.

요약

상표명의 등록 가능성을 분석하기 위해서는 다양한 상황을 고려해야 하지만, 기등록된 유사 발음의 상표를 피하는 것은 가장 우선적인 문제라고 할 수 있다. 한글은 다양한 표현으로 동음어가 생성되는 경우가 비교적 드물지만, 영어의 경우 같은 발음으로 다양한 표현의 단어가 발생할 수 있으므로 상표 등록 과정에서 이를 모두 파악하는 것이 까다로울 수 있다. 본 논문에서는 한글 상표명을 영어 상표명으로 변환하고 변환된 영어 상표명의 발생할 수 있는 다양한 형태의 동음어를 생성하는 방법을 제안한다. 유사 발음을 가지는 영어 상표명을 생성하기 위해서 2가지 방법을 적용하였는데, 첫 번째는 생성 모델을 이용해서 한글 상표명을 입력하면 영어 상표명을 생성하도록 학습하고, Beam Search를 이용해서 생성 모델이 다양한 형태의 영어 상표명을 생성하도록 하는 것이다. 두 번째는 모델이 생성한 영어 상표명을 입력받고 유사 발음 변환 규칙을 활용해서 유사한 발음을 가지는 다양한 형태의 영어 상표명을 생성하도록 하는 것이다. 본 논문의 실험을 통해서 다양한 자연어처리 기법이 유사 발음 상표명 생성에 적용될 수 있음을 보였다.

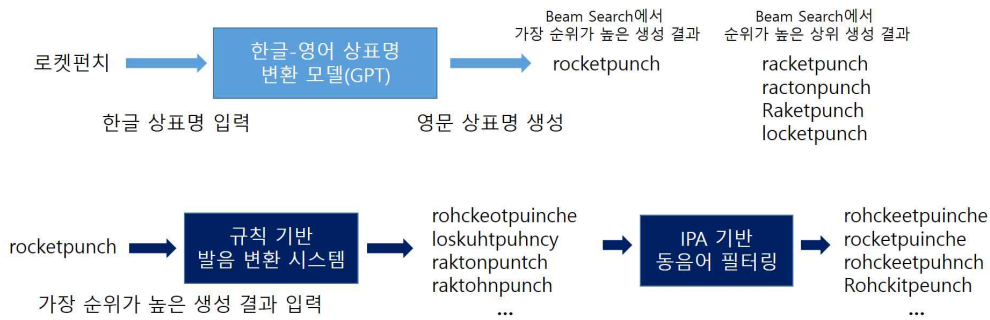


그림 1. 영문 유사 발음 상표 생성 과정

I. 서론

상표 등록은 매년 약 25만 건 이상 등록되고 있으며, 이는 특허, 실용신안, 디자인 등록 수보다 많은 숫자이다. 이처럼 매우 많은 수의 상표가 등록되고 있지만, 상표 등록의 복잡한 과정과 등록 규칙과 같은 관련 지식의 부재로 여전히 상표 등록에 어려움을 겪는 소상공인들이 많다. 이처럼 상표 등록을 간소화하고, 등록하려는 상표의 등록 가능성을 분석하는 것에 대한 수요는 매우 크지만 이를 위한 공개된 솔루션이나 시스템은 많이 공개되어 있지 않은 실정이다.

본 논문에서는 이러한 상표 등록 과정에서 유사한 발음을 가진 상표를 자동으로 분석하기 위한 영문 상표의 동음어 생성 발음 방법을 제안하고자 한다. 유사 발음 생성 방법은 GPT를 이용한 생성 모델을 기반으로 한 생성 방법과 국제 발음 기호와 발음 변환 규칙을 이용한 변환 방법의 2가지 방법을 적용하고 비교해보았다.

II. 본론

그림 1은 본 논문에서 제안하는 동음어 생성 과정을 나타낸다. 발음 변환의 첫 번째 과정은 한글 상표명을 입력하면 영어로 된 상표명을 일차적으로 먼저 생성하는 것이다. 여기에는 GPT를 적용했으며, 한글 상표명을 입력하면 영어 상표명을 입력하도록 학습하였다. 이를 학습하기 위해서는 한글, 영어 상표명 쌍이 필요한데, 키프리스에서 1990년부터 2023년까지 등록된 상표명 데이터를 받아와서 한글과 영어 상표명이 같이 등록된 데이터를 선별하여 학습 데이터 및 테스트 데이터로 이용하였다.

두 번째 과정은 생성 모델이 생성한 영어 발음 외에도 해당 발음이랑 유사하게 나타날 수 있는 다양한 동음어들을 생성하는 과정이다. 이 과정에는 생성 모델의 Beam Search를 이용한 방법과 발음 변환 규칙을 이용한 규칙 기반 생성 방법의 두 가지 방법을 적용하였다.

표 1. 유사 모음 및 유사 자음 표 예시

유사 모음	유사 자음
ue, ui, oo, u, eu	j, z, g
oh, or, eau, o	b, p, v

표 1은 변환 가능한 자음 및 모음의 발음변환표 일부를 나타낸다. 해당 발음변환표는 국립국어원에서 공개한 한국어 로마자 표기법, 특허청에서 공개한 유사 발음 사례 분석을 통한 유사 자음 및 모음을 수집하고 종합하여 구축하였다. 동음어 생성 과정에서 해당 발음변환표를 이용해서 유사 발음에 있는 모음 자음들을 포함하는 상표명들을 생성하는 방법으로 동음어를 생성하였다

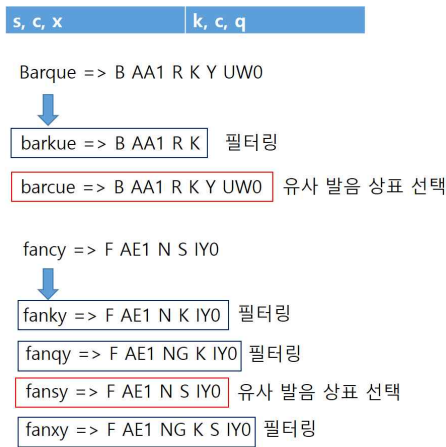


그림 2. IPA 기반 필터링 과정 예시

Beam Search 기반의 생성 방법과 비교하여, 규칙기반 발음 변환 방법의 문제점은 실제로는 비슷하지 않은 다른 발음을 가지는 상표명들이 동음어로 생성될 가능성이 크다는 것이다. 본 논문에서는 이러한 규칙기반 동음어 생성의 한계점을 극복하기 위해서 국제 발음 기호, IPA(International Phonetic Alphabet)를 이용하였다. 생성 모델이 생성한 영어 발음 상표는 비교적 높은 확률로 프롬프트로 입력한 한글 상표의 발음과 유사한 발음을 가지는 영어 상표를 생성하게 되는데 이 영어 상표의 발음을 기준으로 잡는다. 기준이 된 영어 단어를 IPA로 변환하고 규칙기반 발음 변환 방법으로 생성된 후보 단어들의 IPA와 유사도를 비교하여 유사도가 기준을 통과하는 단어들만 유사 발음 상표명으로 생성된 것으로 채택하는 방법을 적용했다. 그림 2는 IPA 유사도 비교를 통한 유사 발음 상표명 필터링 과정을 나타낸다.

III. 실험 결과 및 결론

표 2. 한글 상표명 영어 변환 성능

모델	성능(EM)
polyglot-1.3b	57.14
polyglot-3.8b	58.24

표 2는 polyglot 모델을 이용해서 한글 상표명을 입력하고 영어 상표명을 생성하도록 했을 때, 올바르게 변환한 성능을 나타낸다. 다만, 평가에 이용된 기준 자료는 키프리스에서 같이 등록된 영어 상표명과 일치하는지 아닌지를 보는 EM(Exact Match)로 평가했기 때문에 일치하지 않는 답변이

생성되었다더라도 올바른 답이 생성되었을 수 있다.

표 3. 유사 발음 방법 간의 비교 성능 결과

모델	Beam Search	IPA 필터링
polyglot-1.3b	38.88	44.45
polyglot-3.8b	31.91	44.68

표 3은 다양한 발음 변환을 했을 때의 각 방법의 성능 비교를 나타낸다. 해당 태스크는 평가를 위한 별도의 평가 방법이나 평가 데이터셋이 별도로 구축되어 있지 않기 때문에 테스트셋에서 랜덤하게 20개의 데이터를 뽑아서 직접 평가하는 방법을 적용하였다.

$$y_a = \frac{cor(word_a)}{word_a \cup word_b}, y_b = \frac{cor(word_b)}{word_a \cup word_b}$$

위의 식은 평가 메트릭을 나타낸다. Cor(Word)는 생성된 동음어 중 맞게 생성된 개수를 나타내며, $word_a \cup word_b$ 는 두 방법에서 생성한 동음어의 총 개수를 나타낸다.

1.3b 모델을 적용했을 때, Beam에서 비교 성능이 더 높게 나온 것은 IPA 필터링으로 입력되는 입력 영어 상표명을 잘못 생성한 케이스가 더 많아서 그런 것으로 생각된다.

본 논문에서는 상표 등록 과정에서 유사 발음 영어 상표명을 손쉽게 찾고 분석하기 위한 다양한 자연어처리 기법들을 활용한 유사 발음 영어 상표명 생성 방법을 제안하였다. 향후 연구에서는 해당 태스크를 위한 더 정확한 평가 방법 및 평가 데이터셋을 고안하고 제안할 것이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 지역지능화혁신인재양성사업(IITP-2024-2020-0-01791)

참고 문헌

- [1] Ko, Hyun AH, Kichang Yang, Minho Ryu, Taekyoon Choi, Seungmu Yang, Jiwung Hyun, Sung-Yong Park and Kyubyong Park. "A Technical Report for Polyglot-Ko: Open-Source Large-Scale Korean Language Models." ArXiv abs/2306.02254, 2023.
- [2] Black, Sid, Stella Biderman, Eric Hallahan, Quentin G. Anthony, Leo Gao, Laurence Golding, Horace He, Connor Leahy, Kyle McDonell, Jason Phang, Michael Martin Pieler, USVSN Sai Prashanth, Shivanshu Purohit, Laria Reynolds, Jonathan Tow, Benqi Wang and Samuel Weinbach. "GPT-NeoX-20B: An Open-Source Autoregressive Language Model." ArXiv abs/2204.06745, 2022.
- [3] 박수호, 김광영, Robert J. Fouser, 권혁철. 한국어 음운 현상과 단어 유형을 고려한 한글 - 로마자 변환 시스템 구현. 한국정보과학회 학술 발표논문집, 2001
- [4] 문성민, 김수한, 고연숙, 한국어 발음 변환기(G2P)의 현황과 성능 향상에 대한 언어학적 제안. 언어와 정보, 26(2), 27-46, 2022