지식재산연구
Journal of Intellectual Property

지식재산연구 20권 1호, 2025년 3월, 63-88면
Journal of Intellectual Property Volume 20, Issue 1, March 2025, Pages 63-88.
https://doi.org/10.34122/jip.2025.20.1.63

Print ISSN 1975-5945
Online ISSN 2733-8487

RESEARCH ARTICLE

A Study on Judging Phonetic Similarities for Trademark Searches by Focusing on Enhancing the Search Precision of Korean and English Character Trademarks

Sol-Bin Hwang[1], Woo-Chul Shim[2,3], Bong-Soo Ko[4], Han-Sung Noh[5,6]

[1]Staff of Intelligent Information Strategy Dept. Korea Institute of Patent Information, Republic of Korea
[2]Assistant Manager of Intelligent Information Strategy Dept. Korea Institute of Patent, Republic of Korea
[3]M.S. Candidate in Department of Computer Engineering, Chungnam National University, Republic of Korea
[4]Manager of Intelligent Information Strategy Dept. Korea Institute of Patent Information, Republic of Korea
[5]Head of Intelligent Information Strategy Dept. Korea Institute of Patent Information, Republic of Korea
[6]Ph.D. Candidate, Dept. of Intellectual Property Convergence, Chungnam National University, Republic of Korea

Corresponding Author: Han-Sung Noh (neodream@kipi.or.kr, neodream@o.cnu.ac.kr)

ABSTRACT

Trademarks are marks (symbols, letters, figures, sounds, etc.) used to distinguish one's goods from those of others. Trademark rights are a type of intellectual property, that protect trademarks and maintain reliability, thereby safeguarding consumers' interests, and contributing to industrial development. To register a trademark, assessing its identity or similarity with previously applied trademarks is necessary to examine its distinctiveness, particularly since the phonetic similarity of trademarks plays a crucial role. The importance of a textual trademark pronunciation search system, that determines phonetic similarity has been emphasized to maintain the trademark system's efficacy. However, as trademarks are composed of both Korean and English characters, difficulties arise in converting the pronunciation of one language into another for assessing similarities in trademark pronunciation search systems.

This study proposes a search methodology to address these issues. It focuses on determining the phonetic similarities of textual trademarks composed of Korean and English characters. It utilizes AI-based transliteration and phonetic transcription models on trademarks written in Korean or English, to propose a methodology that unifies Korean and English into a single pronunciation, thereby converting them to closely match the phonetic similarity criteria for textual trademarks. Additionally, it verifies the performance of this methodology through actual cases of trademark phonetic similarities for confirming its effectiveness.

KEYWORDS

Textual Trademarks, Phonetic Similarities, Transliterations, Phonetic Transcriptions, AI-based Methodologies

Open Access

Received: November 19, 2024
Revised: January 06, 2025
Accepted: February 27, 2025
Published: March 30, 2025

Funding: The author received manuscript fees for this article from Korea Institute of Intellectual Property.

Conflict of interest: No potential conflict of interest relevant to this article was reported.

원저

상표 검색을 위한 호칭 유사성 판단에 관한 연구: 한글 영문 문자 상표 검색 정확성 향상을 중심으로*

황솔빈[1], 심우철[2,3], 고봉수[4], 노한성[5,6]

[1]한국특허정보원 지능정보전략실 사원
[2]한국특허정보원 지능정보전략실 대리
[3]충남대학교 컴퓨터공학과 석사과정
[4]한국특허정보원 지능정보전략실 과장
[5]한국특허정보원 지능정보전략실장
[6]충남대학교 지식재산융합학과 박사과정

교신저자: 노한성 (neodream@kipi.or.kr, neodream@o.cnu.ac.kr)

차례

국문초록

상표는 본인의 상품과 타인의 상품을 구별하기 위해 사용하는 표장(기호, 문자, 도형, 소리 등)이다. 상표권은 이러한 상표를 보호하고 신뢰성을 유지함으로써 소비자의 이익을 보호하고 산업 발전에 기여하는 지식재산권의 일종이다. 상표권을 등록하려면 선출원 상표와의 동일성 또는 유사성을 판단해 식별 가능성을 검토해야 하며, 특히 상표의 호칭 유사 여부는 중요한 요소로 작용한다. 상표권 제도가 유지되기 위해서 상표의 호칭 유사 여부를 판단하는 문자 상표 호칭 검색 시스템의 중요성이 강조되어 왔다. 그러나 상표는 한글과 영문으로 구성되어 있으며 상표 호칭 검색 시스템의 유사성 판단을 위해선 한 언어의 발음을 다른 언어로 변환하는 데 어려움이 따른다.

본 연구는 이러한 문제를 해결하고자 문자 상표, 특히 한글과 영문 상표 중심 호칭 유사성 판단을 위한 검색 방법론을 제안한다. 한글 또는 영문으로 이루어진 상표에서 인공지능 기반의 전자(轉字) 모델과 음성 전사(轉寫) 모델을 활용하여 한국어와 영어를 단일 발음으로 통일함으로써 문자 상표의 호칭 유사 기준에 근접하게 변환하는 방법론을 제안한다. 또한, 실제 상표 호칭 유사 사례를 통해 본 방법론의 성능을 검증하여 우수성을 확인하였다.

주제어

문자 상표, 호칭 유사성 판단, 전자(轉字), 음성 전사(轉寫), 인공지능

1. 서론

상표는 상품이나 서비스의 출처를 식별하기 위한 핵심적인 수단으로, 현대의 복잡한 시장 환경에서 그 중요성이 더욱 강조되고 있다. 상표법에서 상표는 “상품을 생산·가공·증명 또는 판매하는 것을 업으로 영위하는 자가 자기의 업무에 관련된 상품을 타인의 상품과 식별하기 위해 사용하는 표장(기호, 문자, 도형, 소리 등)”[1]으로 정의하고 있다.

상표권은 지식재산권의 일종으로, 이런 상표를 보호하고 신용 유지를 도모함으로써 수요자의 이익을 보호하거나 산업 발전에 이바지하는 목적이 있다[2]. 상표권을 등록받기 위해서는 선출원 상표와의 동일·유사성 판단(표장의 호칭, 외관, 관념·의미)을 통해 타 상표와의 식별성을 확인해야 하며, 이는 상표법 “선출원(先出願)에 의한 타인의 등록상표(등록된 지리적 표시 단체표장은 제외한다)와 동일·유사한 상표로서 그 지정상품과 동일·유사한 상품에 사용하는 상표는 등록을 받을 수 없다”[3]에 근거하고 있다.

상표권 등록에서 상표의 유사 여부는 상품 또는 서비스의 출처에 대한 소비자의 혼동을 방지하기 위해 매우 중요하다. 특히 표장의 호칭, 외관, 관념·의미 중 상표의 호칭 유사성은 상표 유사성 판단에서 핵심적인 요소로 작용한다.

이러한 상표권 제도가 유지되기 위해서 상표의 호칭 유사 여부를 판단하는 문자 상표 호칭 검색 시스템의 중요성이 강조되어 왔다. 그러나 이러한 시스템은 상표의 언어적 다양성과 발음의 복잡성으로 인해 어려움을 겪고 있다. 특히 국내에 출원되는 상표는 한글뿐만 아니라 영문으로도 구성되어 있으며, 이는 문자 상표 호칭 검색 시스템의 유사성 판단에서 추가적인 복잡성을 야기한다. 보통 한글, 영문으로 구성된 문자 상표의 경우, 영문으로 표기된 상표는 달리 호칭할 특별한 사정이 없는 한 원칙적으로 영어식 발음을 따른다.[4] 하지만 이는 문자 상표 호칭 검색 시스템에서 문제를 일으킨다.

언어학에서 전자(轉字)는 한 문자 체계에서 다른 문자 체계로 체계적으로 대체하는 방법이며, 음성 전사(轉寫)는 소리에 맞게 발음을 문자로 옮기는 방법이다. 자세한 내용은 배경 연구에서 설명한다.

<표1 영문 상표의 전자(轉字) 알고리즘 변환 결과 비교>

영문 상표	Soundex	Metaphone	NYSIIS
QT	Q300	KT	QT
cutie	C300	KT	CATY

영어 전자(轉字) 알고리즘의 경우 한글 상표의 특성 및 영문 철자와 실제 발음 간 불일치를 정확하게 반영하지 못하며, 상표 호칭 유사성 판단에 어려운 상황이 발생한다. 예를 들어 <표1>에서 QT 라는 영문 상표와 선출원된 cutie라는 상표를 비교하면 실제 발음은 모두 큐티 로 동일하지만, 전자(轉字) 알고리즘으로 변환한 결과는 서로 상이하다. 또한, 경우에 따라 유사성이 높게 측정되더라도 유사한 상표가 너무 많아서 상표 간의 세밀한 비교가 어려워진다. 즉, 너무 비슷한 상표가 많아 검색 결과가 지나치게 포괄적으로 나오게 되며, 이러한 상황은 상표 심사 과정

* 상표의 호칭 유사성 판단을 위해 상표 한글과 영문 간 전자(轉字) 방법을 비교하고, 인공지능 기반 전사(轉寫) 모델을 활용한 상표심사기준 지침서 기반 호칭 유사성 비교 접근법 제안.

1) 상표법 제2조 1항.
2) 상표법 제1조.
3) 상표법 제34조 제7항.
4) 대법원 2005. 11. 10. 선고 2004후2093 판결.

에서 효율적인 결정을 내리기 어렵게 만든다. 결과적으로 상표 간 미세한 차이를 구별하지 못하여 소비자의 혼동을 초래할 가능성이 높아지고, 상표 심사의 정확성과 효율성이 저하되는 문제를 야기한다.

본 연구의 주요 목표는 이러한 문제를 해결하기 위해 영어 전자(轉字) 알고리즘을 활용하는 상표 호칭 유사성 판단 방법의 한계를 극복하고, 한국어 전자(轉字) 모델을 활용해 보다 효율적이고 정확한 상표 호칭 유사성 판단 방법을 제안하는 것이다. 나아가 전자(轉字) 결과에 한국어의 음운론적 특성을 반영하여 더욱 정밀한 음성 전사(轉寫)를 수행한다. 또한, 변환된 음성 전사(轉寫) 결과를 자모 분리 후 유사성 판단 알고리즘을 활용해 상표 간 호칭 유사성을 더 정밀하게 측정한다. 본 연구는 제안한 방법들을 통해 심사기준을 준용하여 한글과 영문 상표의 호칭 검색 정확성을 향상시켜 결과적으로 상표 검색 시 혼동을 유발하는 유사 상표는 하위에 랭크되게 하고, 거절 인용 상표가 상위 검색 결과가 상위에 랭크되게 하는 것을 목표로 한다.

본 연구를 통해 상표 호칭 유사성 판단에서의 언어적, 발음적 한계를 극복하고, 상표심사기준 지침서[5]를 기반으로 상표 검색 과정에서 정확성과 효율성을 높일 수 있는 방법을 제안한다. 이는 상표권 등록 및 분쟁 해결 과정에서 신뢰할 수 있는 유사성 판단을 제공하여 상표 심사의 정확성과 효율성을 향상할 수 있음을 시사한다.

2. 배경 연구

2.1. 전사(轉寫)와 전자(轉字)

전사(轉寫, transcription)와 전자(轉字, transliteration)는 언어학에서 서로 다른 방식으로 소리와 문자를 대응시키는 방법으로 정의된다. 전사(轉寫)는 소리에서 문자로 옮기는 과정을 말하며, 발음에 따라 문자를 표기하는 방식을 포함한다. 반면, 전자(轉字)는 문자에서 문자로 거의 1대1 대응하여 옮기는 방식으로, 발음보다는 원본 문자와의 시각적 일치를 중시한다.

전사(轉寫)는 철자 전사(orthographic transcription)와 음성 전사(phonetic transcription)로 나뉜다. 철자 전사(轉寫)는 발음이 아닌 문자 체계에 맞추어 표기하는 방식으로, 동일 언어 내에서 유사성을 중시한다. 철자 전사(轉寫)는 외국어 고유 명사나 지명을 다른 언어로 옮길 때 주로 사용되며, 철자 체계에 따라 통일된 표기를 제공한다. 음성 전사(轉寫)는 발음에 맞게 표기하는 방식으로, 국제음성기호(IPA), SAMPA와 같은 표준 체계를 통해 특정 발음의 세부적인 차이를 기록한다.

반면, 전자(轉字)는 문자 체계 간의 변환에서 사용되며, 원본 문자의 시각적 일관성을 유지하려는 목적으로 이루어진다. 전자(轉字)는 발음과 무관하게 문자 자체를 변환하기 때문에 발음보다 문자 체계의 일관성을 중시하는 경우, 특히 다른 문자 체계 간의 변환에서 유용하게 쓰인다.

전사(轉寫)와 전자(轉字)는 종종 혼용되기도 한다. 실생활에서 전사(轉寫)는 발음에 따라 문자를 옮기는 방식으로 인식되지만, 실제 전사(轉寫)의 많은 경우 전자(轉字)의 요소가 포함되어 발음과 문자 간 혼동을 야기할 수 있다. 예를 들어, 한글 상표를 로마자로 표기할 때 발음을 기준으로 전사(轉寫)한다고 해도, 로마자 표기법이 따로 있기 때문에 학술적, 실용적 목적에 따라 전사(轉寫)와 전자(轉字)가 결합한 형태로 나타난다.

5) 특허청 상표디자인심사국 상표심사정책과, 「상표심사기준」, 2024. 5. 1. 개정, 특허청, 2024, 260-273면.

<표2 전자(轉字), 음성 전사(轉寫) 예시>

원본	전자(轉字)	음성 전사(轉寫)
SINLA	신라	실라

따라서 본 연구에서는 전자(轉字)는 철자에 맞게 발음을 문자로 옮기는 방법으로 <표2>와 같이 SINLA를 신라로 변환하는 방법으로 정의한다. 또한, 음성 전사(轉寫)는 소리에 맞게 발음을 문자로 옮기는 방법으로 <표2>와 같이 신라를 실라로 변환하는 방법으로 정의한다.

2.1.1. 영어 전자(轉字) 알고리즘

일반적으로 영어 단어는 하나의 문자에 다양한 한국어 발음이 존재하며, 단어와 발음 간의 불일치가 빈번하게 발생한다.[6] 따라서 하나의 발음 규칙으로 변환하여 유사한 발음을 가진 문자를 식별하는 방법이 필요하다. 이를 위해 영어 전자(轉字) 알고리즘이 활용되며, 이 방식은 복잡한 철자와 발음 간의 불일치를 간소화하여 발음 유사성을 체계적으로 판단할 수 있게 한다.

이와 관련한 알고리즘으로 Soundex가 있다. 고숙현, 이재성(2007)[7]은 Soundex는 “영어 단어의 유사성 비교에 사용되는 알고리즘으로, 첫 음절을 제외한 알파벳의 모음을 모두 무시하고 발음이 유사한 자음에 동일한 코드 번호를 부여하여 동일하게 생성된 Soundex 코드에 대해 유사어로 인식 한다”고 한다. 그러나 이 방식은 , “모음과 연자음을 제거하는 조건이 전혀 다른 단어에 동일한 코드가 생성될 가능성을 높여주는 반면, 정확도를 낮추는 요인이 될 수 있다”는 한계가 있다고 설명한다. 즉, Soundex 알고리즘은 영어 철자와 발음의 복잡한 불일치를 충분히 반영하지 못하여 유사한 발음을 가진 이름들을 정확하게 식별하는 데 어려움을 겪는다.

이러한 한계를 극복하기 위해 Metaphone 알고리즘이 개발되었다. 고숙현과 이재성(2007)은 Metaphone은 “Soundex의 단점을 보완한 알고리즘으로, 단어의 첫 음절을 제외한 모음(A, E, I, O, U)을 제거하는 것은 Soundex와 동일하지만, 동일한 자음이 연속되는 경우 모음이 뒤에 나타나지 않는 한 제거하지 않는다. 또한 Soundex와 달리 일률적인 치환이 아니라 별도의 조건에 따른 규칙에 의해 처리한다.”고 설명한다. 그러나 Metaphone도 여전히 영어 발음의 복잡성을 완전히 반영하지는 못하며, 다양한 발음 변이에 대응하는 데 한계가 있다. 이는 Metaphone이 Soundex 방식을 발전시킨 것으로, 기본적으로 Soundex의 특성을 계승하고 있기 때문이다. 따라서 영어 철자와 발음의 불일치 문제를 완전히 해결하지 못하며, 상표명에서 빈번하게 나타나는 창의적인 철자 변형이나 조어에 효과적으로 대응하기 어렵다.

영어 전자(轉字) 알고리즘의 성능을 더욱 향상하기 위해 NYSIIS(NewYorkState Identification and Intelligence System) 알고리즘[8]이 1970년에 제안되었다. NYSIIS는 Soundex와 Metaphone 알고리즘에 비해 더 높은 정확도를 제공하며, 특히 이름의 첫 글자와 마지막 글자를 중심으로 특정 철자 변환 규칙을 적용하는 점에서 기존 알고리즘보다 개선된 성능을 보인다. 예를 들어, 이름의 첫 글자에서 K 를 C 로 변환하거나, PH 를 FF 로 바꾸는 규칙을 사용하여 전자(轉字)의 정확도를 높이고자 한다. 또한, 이름의 마지막에서 IE 를 Y 로 변환하는 등의 규칙을

6) Jae Sung Lee & Key-Sun Choi, “English to Korean statistical transliteration for information retrieval”, *Computational Linguistics*, Vol.12 No.1(1998), pp. 17-37.

7) 고숙현·이재성, “문맥을 고려한 유사 외래어 검출 알고리즘의 성능 향상”, 제19회 한글 및 한국어 정보처리 학술대회, 2007, 114-121면.

8) Petar Rajkovic & Dragan Jankovic, “Adaptation and Application of Daitch-Mokotoff Soundex Algorithm on Serbian Names”, XVII Conference on Applied Mathematics, Novi Sad, Serbia, 2007, pp. 2-3.

적용해 철자와 발음 간의 불일치를 줄이려는 특징을 갖고 있다.

그러나 영어 발음의 복잡한 불규칙성, 즉 동일한 철자가 여러 가지 발음으로 읽히거나 다양한 철자들이 동일한 발음을 가지는 경우를 완벽하게 반영하지 못한다는 한계를 가지고 있다. 예를 들어, 발음상의 차이를 미세하게 반영하지 못하거나, 발음이 유사한 다른 철자 조합을 충분히 구분하지 못하는 문제가 여전히 존재한다.

2.1.2. 한국어 전자(轉字) 및 음성 전사(轉寫)

영어를 한국어 발음으로 변환하는 전자(轉字) 기법은 정확성만 보장된다면 영어 단어와 발음의 불일치를 보다 체계적으로 처리할 수 있는 장점이 있으며, 특히 한국어 발음 비교에 있어 한국어의 음운론적 특성을 반영함으로써 더 정밀한 비교가 가능하다.

영어를 한국어로 변환할 때, 딥러닝 기반의 전자(轉字) 모델인 Smart-G2P[9]가 대표적이다. 김남수 등(2022)이 개발한 Smart-G2P는 기존의 규칙 기반 방식과 달리 딥러닝 기술을 활용하여 영어의 복잡한 발음 패턴을 학습하고, 이를 한글 문자로 변환한다. Smart-G2P는 대규모 데이터를 학습하여 예외적인 발음이나 변이에 대한 대응력이 뛰어나며, 영어의 한국어 전자(轉字)에 있어 매우 유용한 도구로 평가된다.

또한 한국어 음성 전사(轉寫) 기법에서는 한국어의 음운 변동 규칙을 적용하여 텍스트를 실제 한국어 발음에 가깝게 변환하는 기법이 있다. 문성민 등(2021)은 "두음 법칙, 자음 동화, 구개음화 등 한국어의 음운론적 특성을 고려하여 정확한 발음열을 생성하기 위해서는 형태론, 음운론, 통사론에 대한 언어학적 지식과 이를 전산화하는 기술의 융합이 필요하다."[10]고 설명한다. 이러한 필요성에 따라 개발된 음성 전사(轉寫) 모델인 g2pk[11]는 한글을 국립국어원의 표준 발음법을 기반으로 한국어의 복잡한 음운 변동 규칙을 효과적으로 처리하여 한국인의 실제 발음에 가깝게 변환한다. Smart-G2P에서 제공하는 g2pk 모델은 이러한 장점을 바탕으로 한국어 음성 전사(轉寫)의 정확도를 높이며 유용하게 활용될 수 있다.

2.2. 문자열 유사성 판단 기법

전자(轉字) 및 음성 전사(轉寫) 이후에 문자열의 유사성을 정밀하게 측정하기 위해 다양한 방법이 필요하다. 이는 상표의 호칭 유사성을 정확하게 판단하여 소비자의 혼동을 방지하기 위함이다.

유사성 판단방식에는 두 문자열의 같은 위치에 있는 문자를 비교하여 다른 경우마다 거리를 증가시키는 방법이 있다. 이는 문자열의 길이가 동일할 때 위치별 차이를 측정하여 유사성을 계산하는 데 사용된다. 이러한 기능을 수행하는 알고리즘으로 해밍 거리(Hamming Distance)가 있다.

정상원과 정기창(2020)은[12] 해밍 거리의 한계에 대해 다음과 같이 설명하였다. "Hamming Distance는 두 개의 문자열 A와 B를 낱말 별로 1대 1로 비교하는 방법으로, 만약 같은 위치에 있는 낱말이 다르다면 패널티를 준다. 문자열의 길이가 다르면 비교가 어렵고, 삽입이나 삭제에

9) 김남수, "소량 데이터만을 이용한 고품질 종단형(End-to-End) 기반의 딥러닝 다화자 운율 및 감정 복제 기술 개발", 서울대학교, 2022, 22-23면.
10) 문성민 외 2인, "한국어 발음 변환기(G2P)의 현황과 성능 향상에 대한 언어학적 제안", 「언어와 정보」, 제26권(2022), 27-46면.
11) 김남수, 앞의 보고서, 22-23면.
12) 정상원·정기창, "문자열 유사성 알고리즘을 이용한 공종명 인식의 자연어처리 연구 - 공종명 문자열 유사성 알고리즘의 비교", 「한국건설관리학회 논문집」, 제21권 제6호(2020), 125-134면.

따른 변화는 고려하지 않는다."

이는 해밍 거리가 문자열의 길이가 동일해야 하며, 삽입이나 삭제에 따른 변화를 고려하지 않는다는 것을 의미한다. 따라서 상표명에서 자음이나 모음의 추가 또는 누락이 발생하는 경우 유사성을 제대로 반영하지 못한다. 또한, 해밍 거리는 위치별 차이만을 측정하기 때문에 발음상의 유사성이나 어두 부분의 중요성을 고려하지 않는다.

또 다른 유사성 판단 방식에는 한 문자열을 다른 문자열로 변환하기 위해 필요한 최소한의 편집(삽입, 삭제, 대체) 횟수를 측정하는 방법이 있다. 편집 거리가 작을수록 두 문자열의 유사성이 높다고 판단하며, 이러한 기능을 수행하는 알고리즘으로 레벤슈타인 거리(Levenshtein Distance)가 있다.

정상원과 정기창(2020)은 레벤슈타인 거리에 대해 "모든 편집 연산에 동일한 가중치를 부여하며, 문자 위치나 발음상의 중요도를 반영하지 않는다"라고 설명하였다. 따라서 상표 검색에서 중요한 어두 부분의 일치 여부나 발음상의 유사성을 충분히 고려하지 못하며, 레벤슈타인 거리 기반 유사성 판단 방식은 한국어의 음운 변동에 따른 발음 차이를 반영하기 어렵다.

또 다른 방법으로는 문자열을 N개의 연속된 문자 단위로 분할하여 비교하는 방법인 N-gram 기법이 있다. 이는 문자열의 부분적인 유사성을 파악하고, 유사한 패턴이나 구조를 감지하는 데 유용하다. 예를 들어, 프로스펙스를 2-gram으로 분할하면 프로, 로스, 스펙, 펙스 로 나뉜다. 그러나 N-gram 기법은 상표의 호칭 유사성을 평가하는 데 한계가 있다. 문자 단위의 부분 일치에 초점을 맞추기 때문에, 전체적인 발음이나 음운 변동을 고려하지 못한다.

특히 한국어의 경우 음운 규칙에 따라 발음이 변하는데, N-gram 기법은 이러한 음운 변화를 반영하지 못하여 상표의 실제 호칭 유사성을 정확하게 측정하기 어렵다. 또한 상표명 또한 짧은 음절로 구성된 경우가 많아 N-gram 기법에 적합하지 않다. 정상원과 정기창(2020)은 N-gram 알고리즘의 한계에 대해 다음과 같이 언급하였다. "N-GRAM(N=2) 대비 N-Gram (N=3)이 가장 좋지 못한 결과를 보이는 데, 이는 공종명에 사용되는 문자들의 최소단위가 대부분 낱말 세 개 이하이기 때문이라고 해석할 수 있다." 이는 N값이 커질수록 한글의 짧은 음절 구조로 인해 유사성 측정에 어려움이 발생함을 나타낸다.

따라서 N-gram 기법은 상표 호칭 유사성을 평가하는 데 효과적이지 않다. 더욱이, N값을 작게 설정할 경우 가능한 N-gram의 조합 수가 기하급수적으로 증가하게 된다. 이는 연산 속도를 저하할 뿐만 아니라, 검색 시 발생하는 연산량도 크게 증가시킨다. 예를 들어, N=2에서 가능한 모든 2-gram을 생성하고 비교하는 과정은 N=3보다 훨씬 더 많은 연산을 필요로 하며, 이는 실시간 검색이나 대규모 데이터 처리 시 비효율성을 초래할 수 있다. 따라서 N-gram 기법을 사용할 때는 적절한 N값을 선택하는 것이 중요하며, 너무 작은 N값은 오히려 시스템의 성능 저하를 가져올 수 있다.

위와 같은 문제를 해결하기 위해 문자열 내 문자들의 일치와 위치를 고려하여 유사성을 측정하는 방법이 필요하다. 이는 단순한 편집 거리보다 문자열의 구조적 유사성을 반영한다. 이러한 기능을 수행하기 위해 Jaro Distance 알고리즘이 사용된다.[13] Jaro Distance 알고리즘은 문자간 일치 여부와 그 위치를 모두 고려하여 문자열의 유사성을 평가한다.

13) William E. Winkler, "Advanced Methods For Record Linkage", Proceedings of the Section on Survey Research Methods, American Statistical Association, 1994, pp. 2-12.

<그림1 Jaro Distance 산식>

$$d_J = \begin{cases} 0, & \text{if } m = 0 \\ \frac{1}{3}\left(\frac{m}{|s|} + \frac{m}{|t|} + \frac{m-t}{m}\right), & \text{otherwise} \end{cases}$$

Jaro Distance 알고리즘의 문자열 거리 계산 방식은 <그림1>과 같다. s와 t는 비교하려는 두 문자열, m은 일치하는 문자수, t는 전위된 문자 수 절반이다. 이는 문자열 사이에서 일치하는 문자와 전위된 문자(문자 순서가 뒤바뀐 경우)의 수를 기반으로 유사성을 계산하는 방식이다. 특히 단순한 편집 거리(삽입, 삭제, 치환 등을 고려하는 방식)와 달리 문자열의 전반적인 구조를 고려하여 보다 정교한 유사성 측정을 가능하게 한다.

그러나 Jaro Distance 알고리즘에는 한 가지 한계가 있다. 문자열의 시작 부분에 특별한 가중치를 부여하지 않는다는 점이다. 상표 심사와 같이 특정 분야에서는 단어의 어두(초기 부분)에 소비자가 더 민감하게 반응하는 경향이 있다. 이는 상표의 발음이나 인상이 보통 처음 몇 글자에서 형성되기 때문인데, 자로 거리는 이를 충분히 반영하지 못한다.

상표심사기준 지침서에 따르면 “상표는 특히 여러 음절로 이루어진 단어에서는 어두 부분이 강하게 발음되고 인식되므로”[14] 상표의 특성에 맞는 유사성 판단방식으로는 문자열의 시작 부분이 일치할수록 더 높은 유사성을 부여하여 초기 문자의 중요성을 반영하는 방법이 필요하다. 이는 상표의 호칭 유사성 판단에서 소비자가 단어의 초기 부분에 더 민감하게 반응하는 특성을 고려한 것이다. 이러한 기능을 수행하는 알고리즘으로 Jaro Winkler 알고리즘이 있다. Jaro Winkler 알고리즘은 Jaro Distance의 개념을 따르지만, 문자열의 시작 부분이 일치할수록 더 높은 유사성을 부여하여 어두 부분 유사성을 강조한다.

<그림2 Jaro Winkler 산식>

$$Djw = Dj + (l \times p \times (1 - Dj))$$

Jaro Winkler 알고리즘의 계산 방식은 <그림2>와 같다. D_j는 Jaro Distance의 값이며 l은 문자열의 시작부터 연속적으로 일치하는 문자 수(최대 4개), p는 스케일링 팩터(scaling factor)로 일반적으로 0.1로 설정되는 값이다. 이 알고리즘은 산식과 같이 어두 부분의 문자열 유사성을 강조한다. 즉 상표 호칭 유사성 판단에서 어두 부분의 유사성을 중점적으로 비교하는 데 적합하다.

정상원·정기창(2020)[15]은 Jaro Winkler 알고리즘의 특성에 대해 다음과 같이 설명하고 있다. “Jaro Winkler 알고리즘도 레벤슈타인 거리와 마찬가지로 편집 거리의 개념을 사용한다. 하지만, 위의 경우와 다르게 두 가지의 경우에 더욱 높은 관련도 점수를 할당하는데, 이는 첫 번째로 같은 낱말이 문자열 내 특정한 거리 내에 있을 때이다. 즉, ‘부수기’와 ‘깨기’를 비교하였을

14) 특허청 상표디자인심사국 상표심사정책과,「상표심사기준」, 2024. 5. 1. 개정, 특허청, 2024, 260-273면.
15) 정상원·정기창, “문자열 유사성 알고리즘을 이용한 공종명 인식의 자연어처리 연구 - 공종명 문자열 유사성 알고리즘의 비교”,「한국건설관리학회 논문집」, 제21권 제6호(2020), 125-134면.

때, 낱말 '기'는 각각 문자열의 3번째와 2번째에 자리 잡고 있지만, 그 거리가 가까우므로 같은 일치하는 것으로 계산하는 것이다. 두 번째로는 문자열의 시작점부터 A와 B의 문자열이 일치하기 시작하는 것에 높은 점수를 주게 된다. 이는 알고리즘에 방향성을 부여하게 된다. 방향성은 문자열이 일치하는 방향이 같은 것을 중요시하는 것이다." 이처럼 Jaro Winkler 알고리즘은 문자열의 시작 부분 일치에 가중치를 부여하여 어두 부분의 중요성을 강조한다.

또한 정상원·정기창(2020)은 여러 문자열 유사성 판단 알고리즘을 비교한 결과 Jaro-Winkler 알고리즘이 가장 높은 정확도를 보였음을 확인하였다. "결론적으로는 다섯 가지의 알고리즘 중 Jaro Winkler 알고리즘이 가장 좋은 성능을 보여주었다. 이는 아마도 문자열의 방향성을 고려하는 알고리즘의 특색 때문으로 추정하는데, 공종명에서 자주 발견되는 언어적 특징이(수식어 + 명사 + 명사형 동사) 위 알고리즘에서 좋은 결과를 이끌어내는 것으로 보인다." 라고 설명한다.

또한 Barton Beebe & Jeanne C. Fromer(2018)는 Jaro Winkler 측정법은 단순한 동일 문자 매칭을 넘어, '혼동 가능성이 있는 유사성'을 정량적으로 평가하여 상표 비교를 객관적으로 분석하는 데 유용하다고 평가한다.[16] 문자 전위 반영, 초기 문자 가중치, 정규화된 유사도 점수는 기존 문자열 편집거리 대비 상표 유사성 판단에 효과적이라 주장한다. 또한 TTAB(Trademark Trial and Appeal Board)의 의견 데이터를 활용한 분석 결과, 0.875 임계값이 상표 거절 사례에 보수적인 임계값임을 확인하고, 이 임계값 분석을 통해 Jaro Winkler 측정법이 상표 유사성 평가에서 실무적으로도 성능이 우수함을 주장한다.

또한 S. C. Cahyono (2019)[17] 는 Jaro Winkler 방법이 문서 검색 및 유사성 측정 분야에서도 효과적으로 활용될 수 있음을 입증하였다. Jaro Winkler 알고리즘이 문자열의 시작 부분, 즉 'prefix'를 강조하여 비교하는 특성을 활용함으로써, 문서 내에서 중요한 어조나 구조적 특징을 빠르게 포착할 수 있다는 점에 주목하였다. 특히, 논문은 Jaro Winkler가 단순한 동일 문자 매칭을 넘어서, 문서의 초반부에서의 유사성을 정량적으로 평가하는 데 강점을 보인다고 설명한다. 또한, Doc2Vec(Paragraph Vector)와의 비교 실험을 통해, 최신의 딥러닝 기반 방법론과 견줄만한 효율성을 보이면서도, 간단한 알고리즘 구조로 인해 구현 및 해석이 용이하다는 점을 강조하였다.

또한 P. Pitchandi & M. Balakrishnan (2023)[18] 는 적응형 Jaro-Winkler 알고리즘을 기반으로 한 문서 클러스터링 기법의 효과성을 상세히 분석하였다. 이 연구에서는 전처리, 특징 추출, 특징 지식 구축, 그리고 문서 클러스터링의 네 단계로 구성된 프로세스를 통해, RD-TFD 기법과 Chimp Optimization Algorithm (COA)을 활용하여 핵심 특징을 선별한 후, Adaptive Jaro-Winkler와 Jellyfish Search Clustering 기법을 결합함으로써 문서 간의 유사성을 정량적으로 평가하고 클러스터링의 정확도와 효율성을 크게 향상시켰다. 연구 결과, 제안된 방법은 전통적인 클러스터링 기법들(예: k-means, Krill Herd 알고리즘, Moth Flame Optimization 알고리즘)과 비교하여 우수한 성능을 보였으며, 특히 대용량 데이터셋에 적용 시 실시간 문서 검색 및 주제 탐지에 효과적임을 입증하였다.

16) Barton Beebe & Jeanne C. Fromer, "Are We Running Out of Trademarks? An Empirical Study of Trademark Depletion and Congestion", *Harvard Law Review*, Vol.131 No.4(2018), pp. 45-50.

17) S C Cahyono, "Comparison of document similarity measurements in scientific writing using Jaro-Winkler Distance method and Paragraph Vector method", *IOP Conference Series: Materials Science and Engineering*, Vol.662(2019), 052016, pp. 1-9.

18) Perumal Pitchandi & Mathivanan Balakrishnan, "Document clustering analysis with aid of adaptive Jaro-Winkler with jellyfish search clustering algorithm", *Advances in Engineering Software*, Vol.175(2023), 103322.

3. 관련 연구

유사상표 검색 시스템을 구축하기 위해 상표 특성을 고려한 근사매칭에 관련된 연구가 진행되었다. 서창덕·김회율(2000)[19]은 문자기반 상표를 대상으로 질의상표와 유사한 상표들을 정확하게 검색하기 위한 방법을 제안했다. 이 연구에서 사용된 방법론은 상표를 N-gram 기반(2-gram)으로 분할하여 다수의 질의어로 확장하는 방식이다. 그러나 N-gram 방식으로 분리된 단어가 부분적으로 매칭되면서 불필요한 노이즈가 발생하였으며, 이는 검색 성능 저하의 원인이 되었다.

이 문제를 해결하기 위해 서창덕·김회율(2000)은 고빈도 리스트(요부가 아닌 일반적 단어)를 통해 각 단어에 서로 다른 가중치를 부여하는 방법을 채택하여 N-gram 검색 성능을 개선하였다. 이러한 방식으로 검색의 정확도가 15.3% 상승했으며, 특히 상표의 요부, 즉 중요한 요소나 주지 저명한 내용을 의도적으로 배제함으로써 검색 성능을 극대화하는 결과를 얻었다. 이 연구는 상표 검색에서 단순한 문자 일치보다 상표의 핵심 요소를 파악하고, 노이즈를 줄이기 위한 새로운 접근법을 제시함으로써 상표 호칭 유사성 판단의 정확성과 효율성을 높였다. 즉 상표 검색에서 요부 판단의 중요성을 재확인하는 연구였다.

특허 및 상표 검색의 효율성 향상을 위해 질의 로그 데이터를 분석하는 연구가 진행되었다. 이지연·백우진(2006)은[20] 한국특허정보원의 특허기술정보서비스를 이용한 17,559명의 사용자가 작성한 100,016개의 질의 로그 데이터를 분석하여 검색 개선을 위한 방안을 제시하였다. 이 연구에서는 다음과 같은 방법론을 사용하였다.

먼저, 개별적인 질의 로그 분석과 함께 2,202개의 복수 질의문으로 구성된 탐색 세션을 분석하여 사용자의 검색 패턴을 파악하였다. 그 결과, 사용자들은 질의문을 재작성할 때 부연하기(Paraphrasing), 특정화하기(Specialization), 일반화하기(Generalization), 교체하기(Alternation), 중단하기(Interruption)의 다섯 가지 방법을 사용하는 것으로 나타났다. 또한, 특허 및 상표 검색에서 사용자들은 일반 웹 검색보다 불리언(Boolean) 연산자를 더 많이 사용하는 경향이 있었다.

이러한 분석을 기반으로, 연구진은 검색 시스템의 효율성을 높이기 위한 다양한 개선안을 제시하였다. 예를 들어, 상위 빈도 사용자에게 전용 서버를 할당하여 검색 속도를 향상하고, 질의어의 수에 따라 검색 요청을 분산 처리하며, 특정 필드 검색을 최적화하는 방안을 제안하였다. 또한, 동일 세션 내에서 중복된 질의에 대한 결과를 캐싱하여 검색 효율성을 높일 수 있음을 언급하였다.

이 연구는 질의 로그 분석을 통해 특허 및 상표 검색 시스템의 성능을 향상할 수 있는 중요한 통찰력을 제공하며, 사용자들의 검색 행태를 이해함으로써 보다 효과적인 검색 시스템 개발에 기여할 수 있음을 보여준다.

최근에는 딥러닝 기술을 활용하여 상표 호칭 유사성을 판단하는 연구가 활발히 진행되고 있다. 이대호 등(2018)[21]은 CNN(Convolutional Neural Network)을 활용하여 상표의 발음 특성을 특징 벡터로 변환하고, 이를 통해 상표 간의 유사성을 판단하는 방법을 제안하였다. 이 연구에서는 다음과 같은 방법론을 사용하였다.

---

19) 서창덕·김회율, “문자기반 유사상표 검색을 위한 가중치 부여 근사매칭”, 「전자공학회논문지-CI」, 제37권 제1호(2000), 43-54면.

20) 이지연·백우진, “질의로그 데이터에 기반한 특허 및 상표검색에 관한 연구”, 「정보관리학회지」, 제23권 제2호(2006), 61-79면.

21) 이대호 외 2인, “딥러닝 기반 텍스트 상표 검색 서비스”, 세진마인드, 2018, 3-6면.

다국어 상표를 국제 음성 기호(IPA) 또는 로마자 표기법으로 변환하여 상표의 발음 정보를 추출하였다. 그 후 변환된 발음 기호를 N-gram(2-gram) 단위로 분할하여 연속된 음운적 특징을 추출하였다. 최종적으로 음운적 특징 벡터 생성, 추출된 2-gram을 n차원 공간에서 발음 순서에 따라 직선으로 연결하여 2차원 음운적 특징 벡터(Phonetic Feature)를 생성하였다. 이 과정에서 각 발음 기호에 대한 매핑 값을 사전(Dictionary)으로 구성하여 음운적 유사성을 반영하였다.

생성된 음운적 특징 벡터를 CNN의 입력으로 사용하여 상표 간의 유사성을 분류하였다. 약 12,553쌍의 유사 상표와 34,020쌍의 비유사 상표 데이터를 사용하여 모델을 학습하였으며, 실험 결과 약 92%의 정확도를 달성하였다. 이는 기존의 코사인 유사성을 사용한 방법의 정확도인 74.2%보다 약 17.8% 향상된 결과이다.

이 연구는 상표의 호칭 유사성 판단에서 음운적 특징을 효과적으로 활용함으로써, 딥러닝 모델의 성능을 향상할 수 있음을 보여준다. 특히, 2-gram 기반의 음운적 특징 벡터 생성 방법은 상표의 발음 정보를 풍부하게 표현하여 CNN 모델이 상표 간의 미세한 발음 차이를 학습할 수 있도록 하였다.

전종수(2019)의 연구에서도 상표 호칭 유사성 판단을 위한 다양한 알고리즘 비교가 이루어졌다.[22] 이 연구에서는 N-gram 알고리즘, multi N-gram 알고리즘, Fasttext 알고리즘, 그리고 초성 기반 N-gram 알고리즘을 활용하여 상표명을 분석하였다. 특히, 초성 기반 N-gram 알고리즘은 상표 발음을 초성, 중성, 종성으로 분해한 후 초성 정보를 먼저 반영하여 상표 호칭 유사성을 측정하는 방식으로, 기존 방식보다 개선된 성능을 보였다. 전종수(2020)의 연구는 다양한 알고리즘의 성능을 비교하여 상표 호칭 유사성 판단에 적합한 방법론을 제시하였고, 초성 기반 유사성 판단 알고리즘이 상표 발음의 호칭 유사성을 평가하는 데 효과적임을 확인하였다.

4. 실험 설계

4.1. 실험 목표

본 연구의 실험 목표는 한글과 영문으로 구성된 상표의 호칭 유사성 판단에 있어서, 한국어를 중심으로 한 인공지능 기반 단일 발음 통일 방법론의 효과성을 확인하는 것이다. 또한 상표 호칭 유사 검색 실험을 통해 본 방법론의 성능을 검증하여 효과성을 입증하고자 하였다.

4.2. 실험 데이터

본 연구는 실험 데이터 범위를 문자 상표로 한정하였다. 문자 상표 중 한글, 영문, 복합 문자 상표만을 대상으로 실험을 진행하였다. 선출원에 의한 거절 사유는 상표의 호칭, 외관, 관념·의미 중 상표의 호칭 유사성에 의해 거절된 데이터로 한정하였다.

<표3 문자 상표 종류>

분류	한글	영문
한글 상표	신라	
영문 상표		QT
복합 문자 상표	나키	Naki

22) 전종수, “자연어 처리 기법을 활용한 문자열 발음검색 모형에 관한 연구”, 광운대학교 대학원, 석사, 2019 13-43면.

문자 상표는 <표3>과 같이 상표는 한글로만 구성된 한글 상표, 영문으로만 구성된 영문 상표, 한글과 영문 둘 다 존재하는 복합 문자 상표로 이루어져 있다.

<표4 출원 거절 인용 상표 쌍>

구분	출원 상표			구분	거절 인용 상표	
	한글	영문			한글	영문
출원 상표 1	신라		↔	거절 인용 상표 1		Sinla
출원 상표 2		QT	↔	거절 인용 상표 2	큐티	
출원 상표 3	나키	Naki	↔	거절 인용 상표 3	내키	Nakyy
•				•		
•				•		
•				•		
출원 상표 4,030	국물		↔	거절 인용 상표 4,030	궁물	

연구에서는 상표 호칭 유사성 판단 성능을 평가하기 위해 KIPRIS^{Plus}를 통해 입수한 거절결정서를 활용하여 4,030건의 상표 호칭 거절 쌍 데이터를 추출하였다. 이 데이터는 출원 상표와 거절 인용 상표로 구성되어 있으며 각각 4,030건 총 8,060건의 상표가 존재한다. <표4> 와 같이 출원 상표 1과 거절 인용 상표 1은 상표법 “선출원(先出願)에 의한 타인의 등록상표(등록된 지리적 표시 단체표장은 제외한다)와 동일·유사한 상표로서 그 지정상품과 동일·유사한 상품에 사용하는 상표는 등록을 받을 수 없다”에 근거하여 호칭 유사에 의해 각각 출원과 거절 인용된 상표다.

이 데이터는 실제 심사관들이 발음의 유사성, 문자의 유사성 등을 고려하여 판단한 결과로 구성되어 있다. 심사관들은 상표의 호칭 유사성을 심도 있게 분석하여 거절 여부를 결정하였으므로, 해당 데이터는 상표 호칭 유사성 판단에 관한 실제 사례를 정확하게 반영하고 있다. 따라서 데이터의 정확성과 신뢰성이 높아 본 연구에서 제안하는 방법론의 효과를 검증하는 데 매우 적합하다.

<표5 상표 호칭 거절 쌍 데이터 비율>

분류	KIPRIS^{Plus}	건수	비율
출원 상표	상표(한글)	1,369	34.0
	상표(영문)	1,884	46.7
	복합 문자 상표(한글, 영문)	777	19.3
	총합	4,030	100.0
거절 인용 상표	상표(한글)	1,443	35.8
	상표(영문)	1,920	47.6
	복합 문자 상표(한글, 영문)	667	16.6
	총합	4,030	100.0

또한, 4,030건의 상표 호칭 거절 쌍 데이터는 한글, 영문, 복합 문자 상표로 이루어져 있으며 비율은 <표5>와 같다.

4.3. 실험 범위

본 연구는 상표심사기준 지침서[23]를 기반으로 국내 심사관들이 상표 호칭 유사성 판단을 진행하는 방식과 동일하게 연구 범위를 지정하고자 한다.

문자 상표의 경우 한글과 영문으로 구성되어 있어 이를 하나의 발음 규칙으로 변환하기 위해서는 여러 단계의 처리가 필요하다. 첫째로, 영문 상표를 한국어 발음으로 정확하게 변환하는 과정에서 어려움이 발생한다. 이는 영문 상표의 철자와 실제 발음 사이의 불일치로 인해 노이즈가 발생할 확률이 높기 때문이다. 예를 들어, 같은 철자를 가진 영문 상표도 다양한 발음을 가질 수 있어, 이러한 변이는 상표 호칭 유사성 판단에서 오류를 유발할 수 있다.

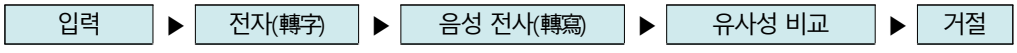
이러한 문제를 해결하기 위해, 영문 상표를 한국어 발음으로 변환하는 기존 방법을 활용하여 본 연구 방법론에 적용한 후 비교 연구하였다. 이는 특허청 상표심사기준 지침서에서 제시한 바와 같이, “외국문자 상표에 관한 호칭 유사 여부의 판단은 내국인 관례상의 호칭은 물론 해당 외국인의 대표적인 호칭도 함께 고려하여야 한다”는 원칙을 따르는 것이다. 이를 통해 영문 상표의 철자와 발음의 불일치로 인한 노이즈를 최소화하고 정확한 발음 규칙을 적용할 수 있다. 이를 위해 다양한 전자(轉字) 알고리즘 및 모델을 비교 실험하여 가장 적합한 방식을 선택하였다.

한글 상표의 발음을 한국인의 실제 발음에 가깝게 변환하는 단계가 필요하다. 이는 특허청 상표심사기준 지침서에서 제시한 바와 같이 “한글의 두음 법칙이나 자음 동화 현상 등 음운 변동을 반영하여 상표의 발음열이 실제 발음을 정확히 나타내도록 하는 것이 중요하다.”에 근거한 것이다. 이를 위해 음성 전사(轉寫) 알고리즘을 검토하고 적용하였다. 이러한 변환은 상표의 호칭을 정확하게 파악하고 유사성 비교 정확도를 높이는 데 필수적이다.

상표의 호칭 유사성을 비교하기 위해서는 문자열 유사성 판단 알고리즘이 필요하다. 특허청 상표심사기준 지침서에서 제시한 바와 같이 “상표는 특히 여러 음절로 이루어진 단어에서는 어두 부분이 강하게 발음되고 인식되므로” 문자열의 시작 부분이 일치할 때 더 높은 가중치를 부여하는 알고리즘을 고려한다. 이는 소비자에게 강한 인상을 주는 상표의 초기 발음을 중심으로 유사성을 평가하기 위함이다.

또한, 한국어 음운론적 특성을 고려하여 자모 분리 후 비교하는 방법을 적용하였다. 한국어의 음절은 초성, 중성, 종성의 자모 조합으로 이루어져 있어, 자모 단위로 분해하여 비교하면 발음의 미세한 차이까지도 반영할 수 있다. 이는 특허청 상표심사기준 지침서에서 제시한 바와 같이 “자음 접변 현상 등 음운 변동이 상표의 호칭 유사성에 직접적인 영향”을 미치기 때문이다.

<그림3 상표심사기준 지침서를 고려한 상표 호칭 유사판단 절차>



따라서 상표심사기준 지침서를 기반으로 특허청 심사관들은 <그림3> 과 같은 절차에 따라 상표 호칭 유사성 판단을 진행한다.

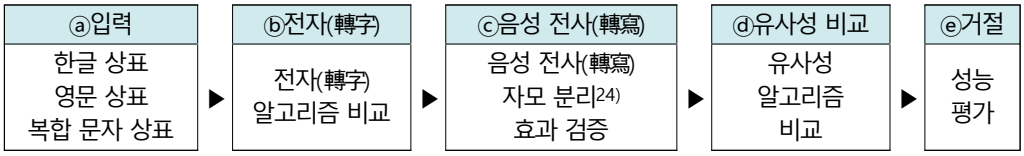
5. 방법론

본 연구에서 제안한 방법론은 실제 상표심사기준 지침서를 기반으로 특허청 심사관들이 상표 호칭 유사성 판단을 진행하는 방식을 최대한 고려하여 설계하였다. 한글 영문 상표를 하나의

23) 특허청 상표디자인심사국 상표심사정책과, 「상표심사기준」, 2024. 5. 1. 개정, 특허청, 2024, 260-273면.

발음 규칙으로 변환하고 유사성을 비교하는 접근은 심사관용 검색 시스템을 참고한 것이다. 이러한 방식으로 고도화하는 것은 상표 호칭 유사성 판단의 정확성과 효율성을 높이는 데 유리하며 특히 심사 과정에서의 일련의 절차들을 본 연구의 상표 호칭 유사 판단 방식에 적용함으로써, 보다 체계적이고 신뢰성 있는 연구 수행이 가능해진다.

<그림4 유사판단 단계별 연구 >



본 연구에서는 <그림4> 와 같이 상표심사기준 지침서 상표 호칭 유사판단 절차를 고려해 상표 호칭 검색 고도화의 각 연구범위를 지정하고 최신 자연어 처리 기술을 적용 및 비교 선정하여 상표 검색의 정확도와 신뢰성을 향상하고자 한다. 이를 위해, 실제 유사판단 절차를 따라 입력, 전자(轉字), 음성 전사(轉寫), 유사성 비교, 선택의 다섯 가지 단계로 연구 범위를 지정하여 단계별로 연구를 진행하였다.

먼저 유사성 판단 알고리즘을 하나로 고정하여 다른 변수들의 영향을 통제하였다. 이는 실험 과정에서 변수의 수를 줄여 각 단계별로 적용되는 알고리즘이나 모델의 효과를 명확하게 파악하기 위함이다. 만약 유사성 판단 알고리즘까지 동시에 변경하면, 각 단계에서의 성능 변화가 어떤 요인에 의한 것인지 정확하게 알기 어렵다.

본 연구에서는 문자열 유사성 측정 알고리즘인 Hamming, Levenshtein, N-gram (2-gram), Jaro Winkler 중 Jaro Winkler 알고리즘을 초기 유사성 판단 도구로 선정하였다.

상표심사기준 지침서에서 제시한 바와 같이 “상표는 특히 여러 음절로 이루어진 단어에서는 어두 부분이 강하게 발음되고 인식되므로” 문자열의 시작 부분이 일치할 때 더 높은 가중치를 부여하는 Jaro Winkler 알고리즘을 선택했으며 상표 호칭 유사성 판단에 있어 초기 알고리즘으로 적합하다고 판단하였다. 초기에 유사성 판단 알고리즘을 고정함으로써, 이후 단계인 ⓑ전자(轉字), ⓒ음성 전사(轉寫), 단계별 적용된 방법이 상표호칭 유사성 판단에 미치는 효과를 명확하게 평가할 수 있다.

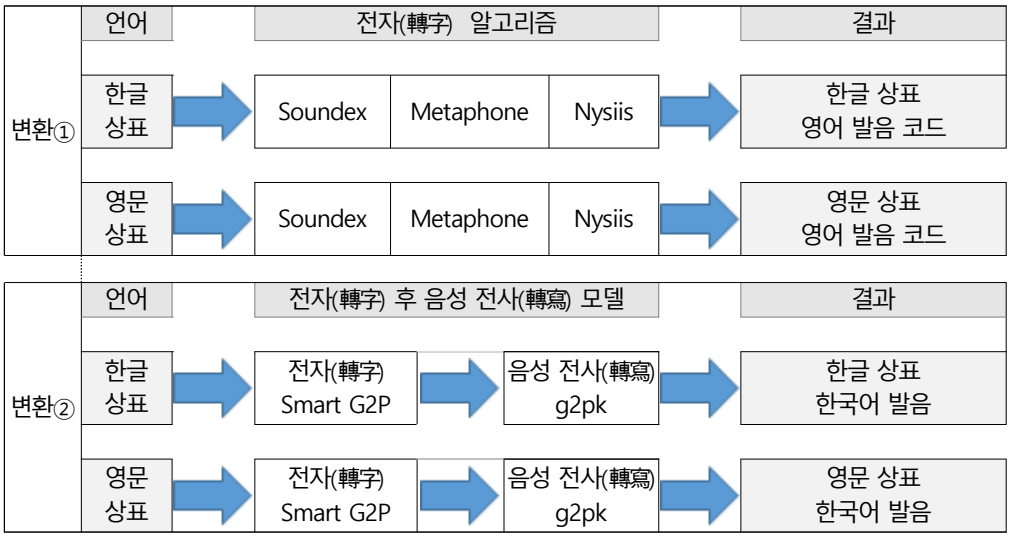
<그림4> ⓑ전자(轉字) 실험 단계에서 고정된 유사성 판단 알고리즘을 사용해 다양한 전자(轉字) 알고리즘 및 모델의 성능을 비교하였다. 비교 대상 알고리즘은 기존의 규칙 기반 알고리즘인 Soundex, Metaphone, NYSIIS와 인공지능 기반의 한국어 전자(轉字) 모델인 SmartG2P이다. 각 알고리즘 및 모델을 활용하여 전자(轉字) 후, 상표 호칭 유사성 판단 결과를 분석한다. 즉 알고리즘 및 모델로 생성된 전자(轉字) 문자를 기반으로 초기 고정된 유사성 판단 알고리즘 Jaro Winkler로 유사성을 측정하여 상표 간 호칭 유사성을 평가한다. 영어 전자(轉字) 알고리즘과 한국어 전자(轉字) 모델 중 상표 호칭 유사성 판단에 효과적인 방법을 결과를 통해 검증하였다.

<그림4> ⓒ음성 전사(轉寫) 단계에서 전자(轉字) 모델 효과 검증 후, 한글 상표를 실제 한국인의 발음과 가깝게 변환하기 위해 음성 전사(轉寫) 모델인 g2pk를 적용한다. 음성 전사(轉寫) 모델인 g2pk는 국립국어원의 표준 발음법을 기반으로 개발된 모델로, 한글의 두음 법칙, 자음 동화 현상 등 음운 변동을 정확하게 반영한다. 이를 통해 상표의 발음이 실제 한국인의 발음에 더

24) 자모 분리란 한글의 초성, 중성, 종성을 분리하는 것.

욱 가까워지며, 상표 호칭 유사성 비교 시 정확도를 높일 수 있다. 음성 전사(轉寫) 전후의 상표 문자를 고정된 유사성 판단 알고리즘 Jaro Winkler를 통해 성능 비교하여, 음성 전사(轉寫)가 유사성 판단의 정확도에 미치는 영향을 분석한다. 이를 통해 음성 전사(轉寫) 알고리즘의 적용이 상표 호칭 유사성 판단의 성능을 향상함을 검증하였다.

<그림5 발음 변환 과정>



실험을 통해 상표 호칭 유사성 판단에서 <그림5> 변환① 한글의 영문 전자(轉字)보다 <그림5> 변환② 영문의 한글 전자(轉字) 및 음성 전사(轉寫)가 더 효과적임을 입증하였다. 영문 상표를 한국어 발음으로 변환하는 방식은 한국어의 음운론적 특성을 보다 잘 반영해 정확한 발음 비교를 가능하게 한다. 특히, 국내에서 출원된 상표인 만큼 한국어의 특성을 고려한 변환이 상표 간의 미세한 발음 차이를 더 정확하게 반영하며, 이를 통해 상표 심사 과정에서 발생할 수 있는 혼동 가능성을 줄이는 데 유리한 점을 실험을 통해 확인하였다.

<표6 자모 분리 예시_신라>

음절	초성	중성	종성
실	ㅅ	ㅣ	ㄹ
라	ㄹ	ㅏ	-

상표 데이터의 전자(轉字), 음성 전사(轉寫) 후에도 상표 호칭의 미세한 음운 변동을 더욱 정확하게 반영하기 위해 <표6>과 같이 자모 분리를 적용하였다. 자모 분리는 한글 음절을 초성, 중성, 종성 단위로 분해하여 비교함으로써 발음상의 미세한 차이까지도 고려할 수 있게 한다. 자모 분리 변환 전후의 상표 문자를 고정된 유사성 판단 알고리즘 Jaro Winkler를 통해 성능 비교하여, 자모 분리가 유사성 판단의 정확도에 미치는 영향을 분석하였다.

마지막으로 <그림1> ⓓ유사성 비교 실험을 진행하였다. 전자(轉字), 음성 전사(轉寫) 적용 및 자모 분리 변환으로 상표 호칭의 비교가 실제 발음 비교에 가까워짐에 따라, 유사성 판단 알고리즘의 성능을 다시 평가하였다. 기존에 고정했던 Jaro Winkler 알고리즘 외에도 Levenshtein Distance, N-gram(2-gram), Hamming Distance 등의 알고리즘을 적용하여 비교 분석한

다. 초기 고정한 Jaro Winkler 알고리즘이 상표 호칭 유사성 판단에 가장 적합한 문자열 비교 방식인지 실험을 통해 검증하였다.

6. 평가

6.1. 평가 기준

평가 방법은 다음과 같다. 상표 호칭 거절 쌍 데이터를 활용해 각 출원 상표를 입력하여 해당 상표와 유사한 거절 인용 상표를 찾는 총 4,030번의 유사성 판단 검색을 수행하였다.

<표7 상표 호칭 유사성 판단 실험 방식>

구분		출원 상표	
		한글	영문
실험 1	출원 상표 1	신라	
실험 2	출원 상표 2		QT
실험 3	출원 상표 3	나키	Naki
.			
.			
.			
실험 4,030	출원 상표 4030	국물	

구분	거절 인용 상표	
	한글	영문
거절 인용 상표 1		Sinla
거절 인용 상표 2	큐티	
거절 인용 상표 3	내키	Nakyy
.		
.		
.		
거절 인용 상표 4,030	궁물	

<표7>과 같이 검색 대상은 4,030건의 거절 인용 상표로 한정하였는데, 각각 검색 실험 마다 동일한 거절 인용 상표를 대상으로 검색을 진행함으로써 모든 출원 상표에 대해 동일한 조건에서 평가를 수행하여 단계별 방법론의 성능을 비교하기 위함이다.

<표8 kukka 검색 결과 예시>

출원 상표
kukka

Rank	결과	
1	coocaa	
2	COUCOU	
3	Quokka	
4	KOOAN	
5	KHUKRI	
6	Qookka Games	거절 인용 상표
7	쿠기커피	
8	CO CAR ON	
9	쿼카맥주	
10	코코코~ 알루	

<표8>은 실험에 사용되는 4,030건 거절 인용 상표를 대상으로 검색을 수행한 결과 예시이다. 출원 상표인 kukka를 입력했을 때 검색 결과를 출력하면 Rank 6에 거절 인용 상표 Qookka Games가 위치하게 된다.

검색 성능 지표에는 Accuracy, Recall, Precision, F1 Score 등이 존재하는데, 본 연구에서는 Precision과 MRR 지표를 사용한다. 본 연구는 제안한 방법들을 통해 한글과 영문 상표의 발음 비교 정확성을 향상시켜, 상표 심사 기준에 따라 검색 결과 상단에 거절 인용 상표가 최대한 검색되도록 하는 것을 목표로 하기 때문이다.

Accuracy는 전체 데이터에서 정답을 맞춘 비율을 평가하지만, 검색 결과의 순위를 고려하지 않기 때문에 본 연구 목적에 적절한 지표가 아니다. 예를 들어, 올바른 유사 상표가 데이터셋에 50개 존재하고 검색 시스템이 48개를 정확히 반환했다면 Accuracy는 96%로 높게 나올 수 있다. 그러나 심사관이 검토하는 상위 10개 결과 중 올바른 유사 상표가 2개뿐이라면, 실제 업무에서는 성능이 낮은 것으로 평가된다.

F1 Score 역시 Precision과 Recall의 조화 평균으로 검색 결과의 정확성을 평가하지만, 특정 상표가 검색 결과에서 얼마나 빨리 등장하는지는 반영하지 못한다. 유사 상표 10개 중 8개가 검색 결과에 포함되었다면 Recall이 80%로 높게 측정될 수 있지만, 실제로 중요한 상표가 검색 결과 하위에 위치한다면 실무적으로 검색 성능 개선을 체감하기 어렵다.

이러한 문제를 해결하기 위해 본 연구에서는 Precision@Rank와 MRR을 활용하였다. Precision@Rank는 상위 N위 안에서 거절 인용 상표가 얼마나 포함되는지를 평가하며, MRR은 올바른 유사 상표가 검색 결과에서 처음 등장하는 순위를 반영한다. 이는 심사관이 방대한 상표 데이터에서 유사상표 검색을 효율적인 수행해야 하는 실무 환경과 부합하며, 본 연구의 목표인 검색 결과 상단에 거절 인용 상표 노출 여부를 평가하는 데 효과적이다.

<그림6 Precision@Rank N 산식>

$$\text{Precision@Rank N} = \frac{\text{상위 N위 내 거절 인용 상표}}{\text{총 실험 건수}}$$

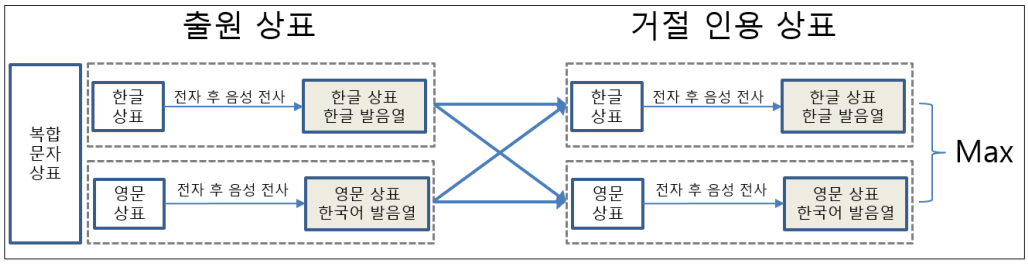
성능 평가는 Precision@Rank N 방식으로 이루어 졌으며, 이는 <그림6>과 같이 검색 결과의 상위 N위 내에 거절 인용 상표가 위치한 비율과 건수를 나타낸다. 검색 결과의 상위 N위까지에서 관련성이 있는 거절 상표를 얼마나 정확하게 찾아냈는지를 평가하였다.

<그림7 MRR 산식>

$$\mathbf{MRR} = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{\mathbf{rank}_i}$$

또한, <그림7> MRR(Mean Reciprocal Rank)을 지표로 사용하였다. MRR은 검색 시스템의 전체적인 성능을 평가하기 위한 지표로, 각 입력에 대해 첫 번째로 거절 상표가 등장한 순위의 역수를 평균한 값이다. MRR 값이 높을수록 관련 상표가 검색 결과 상위에 위치한다는 것을 의미하며, 사용자가 원하는 결과를 빠르게 얻을 수 있음을 나타낸다.

<그림8 복합 문자 상표 유사도 판단 방식>



복합 문자 상표의 경우 유사도 판단을 <그림8>과 같이 한글-한글, 영문-영문뿐만 아니라 한글-영문, 영문-한글까지 총 네 가지 경우의 수를 모두 고려해 유사성 판단을 진행하고, 최댓값을 대표 유사도 값으로 선정하였다. 이는 복합 문자 상표의 발음 유사성을 한글 영문 교차검증을 통해 정확하게 판단하기 위함이다.

6.2. 평가 결과

본 연구에서 제시한 단계별 방법론 평가 결과이다. 단계별 방법론 적용 및 검색 성능을 Precision@Rank N 과 MRR을 통해 검색 성능 지표로 평가하였다. 평가 결과를 통해 단계별 방법론의 효용성을 입증하였다.

<표9 전자(轉字) 별 MRR>

유사도 알고리즘	Jaro Winkler			
	Soundex	Metaphone	Nysiis	Smart G2P
MRR	0.3568	0.4509	0.4352	0.6080

<표10 전자(轉字) 별 Precision@Rank N>

유사도 알고리즘	Jaro Winkler			
Rank	Soundex	Metaphone	Nysiis	Smart G2P
1	22.98% (926)	35.96% (1449)	35.26% (1421)	54.09% (2180)
2	35.48% (1430)	44.76% (1804)	42.98% (1732)	61.64% (2484)
3	42.93% (1730)	49.60% (1999)	47.32% (1907)	65.31% (2632)
5	52.03% (2097)	55.31% (2229)	52.68% (2123)	68.78% (2772)
10	60.84% (2452)	62.93% (2536)	59.28% (2389)	72.63% (2927)
20	66.15% (2666)	69.48% (2800)	66.60% (2684)	75.68% (3050)
30	68.93% (2778)	72.66% (2928)	69.78% (2812)	77.47% (3122)
40	70.84% (2855)	74.64% (3008)	71.91% (2898)	78.29% (3155)
50	72.41% (2918)	76.55% (3085)	73.67% (2969)	79.21% (3192)

다양한 전자(轉字) 알고리즘 및 모델로 생성된 결과를 기반으로 고정된 유사성 판단 알고리즘 Jaro Winkler로 유사성을 측정하여 상표 간 호칭 유사성을 평한 결과이다.

실험 결과, SmartG2P 모델은 기존 알고리즘들에 비해 압도적으로 높은 성능을 보였다. <표10>을 보면, Rank 1에서 SmartG2P의 정확도는 54.09%로, Soundex의 22.98%, Metaphone

의 35.96%, NYSIIS의 35.26%에 비해 높은 성능을 나타냈다. <표9>를 보면 MRR 또한 Smart G2P는 0.6080으로, 다른 알고리즘들 대비 0.17 ~ 0.25 높게 평가되었다.

선정된 전자(轉字) 모델(Smart G2P)을 적용한 후, 상표를 한국인의 실제 발음과 가깝게 변환하기 위해 음성 전사(轉寫) 모델 g2pk를 적용 Jaro Winkler로 유사성을 측정한 결과이다.

<표11 음성 전사(轉寫) MRR>

유사도 알고리즘	Jaro Winkler	
	Smart G2P	Smart G2P → g2pk
MRR	0.6080	0.6197

<표12 음성 전사(轉寫) Precision@Rank N>

유사도 알고리즘	Jaro Winkler	
Rank	Smart G2P	Smart G2P → g2pk
1	54.09% (2180)	55.36% (2231)
2	61.64% (2484)	63.00% (2539)
3	65.31% (2632)	66.40% (2676)
5	68.78% (2772)	69.73% (2810)
10	72.63% (2927)	73.60% (2966)
20	75.68% (3050)	76.50% (3083)
30	77.47% (3122)	78.24% (3153)
40	78.29% (3155)	78.96% (3182)
50	79.21% (3192)	79.73% (3213)

<표11> 에서 알 수 있듯이, g2pk 모델을 적용함으로써 MRR 값이 0.6080에서 0.6197로 향상되었다. 또한, Rank 1에서의 정확도가 54.09%에서 55.36%로 증가하였으며, 상위 순위 전반에 걸쳐 성능이 향상되었다. 이는 음성 전사(轉寫)를 통해 한글 상표의 실제 발음을 더욱 정확하게 반영함으로써 상표 호칭 유사성 판단의 정확도가 개선되었음을 의미한다.

전자(轉字), 음성 전사(轉寫) 후 상표 호칭의 미세한 음운 변동을 더욱 정확하게 반영하기 위해 자모 분리 적용후 Jaro Winkler로 유사성을 측정한 실험 결과이다.

<표13 자모 분리 MRR>

유사도 알고리즘	Jaro Winkler	
	Smart G2P → g2pk	Smart G2P → g2pk → 자모 분리
MRR	0.6197	0.6546

<표14 자모 분리 Precision@Rank N>

유사도 알고리즘	Jaro Winkler	
Rank	Smart G2P → g2pk	Smart G2P → g2pk → 자모 분리
1	55.36% (2231)	57.52% (2318)
2	63.00% (2539)	67.02% (2701)
3	66.40% (2676)	71.44% (2879)
5	69.73% (2810)	75.19% (3030)
10	73.60% (2966)	78.96% (3182)
20	76.50% (3083)	81.94% (3302)
30	78.24% (3153)	83.13% (3350)
40	78.96% (3182)	84.04% (3387)
50	79.73% (3213)	84.91% (3422)

<표13> 에서 볼 수 있듯이, 자모 분리를 적용한 후 MRR 값이 0.6546으로 더욱 향상되었다. Rank 1에서의 정확도 또한 57.52%로 증가하여, 자모 분리가 상표 호칭 유사성 판단에 긍정적인 영향을 미침을 확인할 수 있다. 이는 자모 분리를 통해 한글 상표의 자음접변[25] 특성을 반영함으로써, 발음상의 미세한 차이까지 고려할 수 있게 되었기 때문이다.

전자(轉字), 음성 전사(轉寫) 적용 및 자모 분리로 상표 호칭의 비교에서 발음 비교의 정확도가 상승함에 따라, 유사성 판단 알고리즘의 성능을 다시 평가하였다. 기존에 고정했던 Jaro Winkler 알고리즘 외에도 Levenshtein Distance, N-gram(2-gram), Hamming Distance 등의 알고리즘을 적용하여 비교 분석한 결과이다.

<표15 유사성 판단 알고리즘 MRR>

변환	Smart G2P → g2pk → 자모 분리			
	Hamming	Levenshtein	N-gram(2-gram)	JaroWinkler
MRR	0.5422	0.5936	0.6415	0.6546

<표15>는 각 유사성 판단 알고리즘에 대한 Mean Reciprocal Rank(MRR) 결과를 나타낸다. 결과에 따르면, Jaro Winkler 알고리즘이 MRR 0.6546으로 가장 높은 성능을 보였으며, 그 뒤를 이어 N-gram(2-gram)이 0.6415의 MRR 값을 기록하였다. Levenshtein Distance와 Hamming Distance는 상대적으로 낮은 성능을 보였다.

25) 두 말 사이에서 윗말의 종성(終聲)과 아랫말의 초성(初聲)이 합칠 때에 발음상(發蔭上)으로 자음의 소리값(音價)이 변하는 현상.

<표16 유사성 판단 알고리즘 Precision@Rank N>

변환	Smart G2P → g2pk → 자모 분리			
Rank	Hamming	Levenshtein	N-gram(2-gram)	JaroWinkler
1	48.86% (1969)	53.33% (2149)	56.40% (2273)	57.52% (2318)
2	54.86% (2211)	59.80% (2410)	65.21% (2628)	67.02% (2701)
3	57.15% (2303)	62.61% (2523)	69.06% (2783)	71.44% (2879)
5	59.73% (2407)	65.63% (2645)	72.83% (2935)	75.19% (3030)
10	63.82% (2572)	70.22% (2830)	78.14% (3149)	78.96% (3182)
20	67.27% (2711)	74.34% (2996)	82.13% (3310)	81.94% (3302)
30	69.08% (2784)	76.33% (3076)	84.24% (3395)	83.13% (3350)
40	70.74% (2851)	78.21% (3152)	85.56% (3448)	84.04% (3387)
50	71.84% (2895)	79.85% (3218)	86.50% (3486)	84.91% (3422)

<표16>의 Precision@Rank N 결과를 분석한 결과, Rank 1~10까지의 상위 순위에서는 Jaro Winkler 알고리즘이 가장 높은 정확도를 보였다. 반면, Rank 20 이상에서는 N-gram(2-gram) 알고리즘의 예측률이 상대적으로 더 높은 것으로 나타났으나, 본 연구에서는 상위 순위인 Rank 10위 이내에서의 검색 성능 개선을 우선 고려하였다. 결과적으로, 전체적인 평균 예측률보다는 유사한 상표가 상위 랭킹 내에서 얼마나 검색되는가를 핵심적인 평가 요소로 판단했으며, 이를 반영하기 위해 MRR(Mean Reciprocal Rank) 지표를 활용하였다. 상위 랭킹에서 더 높은 성능을 보이는 알고리즘이 MRR 값에서도 우위를 점하게 됨에 따라 Jaro Winkler 알고리즘이 가장 높은 MRR 값을 기록한 것으로 분석되었다. 따라서, <표16> 실험결과에서 N-gram(2-gram)이 Rank 20 이상에서 Jaro Winkler에 비해 상대적으로 높은 예측률을 보이고 있으나, MRR 및 Rank 10까지의 상위 모두에서 Jaro Winkler 방법이 우수함을 확인할 수 있다.

<표17 유사성 판단 알고리즘 연산 소요 시간>

방법	총 소요 시간 (초)	총 처리 건수	건당 소요 시간 (초)
Jaro Winkler	6.08	4,030	0.00151
Levenshtein	8.21		0.00204
Jaro	6.18		0.00153
N-gram	31.84		0.00789
Hamming	16.67		0.00414

또한 <표17> 볼 수 있듯이, N-gram 알고리즘은 연산 시간이 가장 길었으며, 건당 소요 시간도 0.00789초로 다른 알고리즘 대비 현저히 높았다. 반면, Jaro Winkler 알고리즘은 6.08초로 비교적 빠른 연산 시간을 보이며, 상위 검색 성능과 연산 성능을 동시에 고려했을 때 가장 효율적인 방법으로 나타났다.

<표18 상표 호칭 검색 성능 종합 비교 – MRR>

유사도	Jaro Winkler			
	Metaphone	Smart G2P	→ g2pk	→ 자모 분리
MRR	0.4509	0.6080	0.6197	0.6546

<표19 상표 호칭 검색 성능 종합 비교 - Precision@Rank N>

유사도	Jaro Winkler			
Rank	Metaphone	Smart G2P	→ g2pk	→ 자모 분리
1	35.96% (1449)	54.09% (2180)	55.36% (2231)	57.52% (2318)
2	44.76% (1804)	61.64% (2484)	63.00% (2539)	67.02% (2701)
3	49.60% (1999)	65.31% (2632)	66.40% (2676)	71.44% (2879)
5	55.31% (2229)	68.78% (2772)	69.73% (2810)	75.19% (3030)
10	62.93% (2536)	72.63% (2927)	73.60% (2966)	78.96% (3182)
20	69.48% (2800)	75.68% (3050)	76.50% (3083)	81.94% (3302)
30	72.66% (2928)	77.47% (3122)	78.24% (3153)	83.13% (3350)
40	74.64% (3008)	78.29% (3155)	78.96% (3182)	84.04% (3387)
50	76.55% (3085)	79.21% (3192)	79.73% (3213)	84.91% (3422)

위의 <표18>에서 알 수 있듯이, 본 연구에서 제안한 방법들을 단계적으로 적용함에 따라 MRR 값이 지속적으로 향상되었음을 확인할 수 있다. 본 연구에서 비교한 규칙기반 영문 전자(轉字) 알고리즘 중 가장 우수한 성능을 보였던 Metaphone의 MRR 값은 0.4509에 불과하였다. 그러나 인공지능으로 학습한 영문을 한글로 전자(轉字) 하는 SmartG2P를 적용함으로써 MRR이 0.6080으로 크게 상승하였으며, g2pk를 통한 음성 전사(轉寫)를 추가하여 0.6197로 향상되었다. 이후 자모 분리를 적용하여 MRR은 0.6546로 증가하였다. 이는 규칙 기반 알고리즘 중 성능이 가장 우수한 Metaphone 대비 약 1.45배의 성능 향상을 달성한 것으로, 본 연구에서 제안한 방법론의 효과를 명확히 보여준다. 본 연구의 제안방법론이 한글과 영문 문자 상표의 호칭 유사성 판단에 있어 가장 우수한 성능을 보임을 실험적으로 입증하였다. 결국, 상표 호칭 유사성 판단에서 한국어의 음운론적 특성을 고려하고, 영문 상표를 한국어 발음으로 변환하는 방법이 한글 상표를 영어 발음으로 변환하는 방법보다 효과적임을 확인하였다.

7. 결론

본 연구에서는 상표 호칭 유사성 판단의 정확성과 효율성을 높이기 위한 상표심사기준 지침서 기반 방법론을 제안하고, 이를 실험적으로 검증했다. 상표를 영문으로 전자(轉字) 하는 상표 호칭 유사성 판단 방식이 가진 한계, 즉 영문 상표의 철자와 발음 불일치, 한국어 음운론적 특성을 충분히 반영하지 못하는 점, 복합 문자 상표 간의 비교 어려움 등을 해결하고자 다양한 전자(轉字) 및 음성 전사(轉寫) 방식을 도입했다. 그 결과, 한국어로 전자(轉字) 및 음성 전사(轉寫)하여 상표를 비교하는 방식이 상표 호칭 유사성 판단에 있어 효과적인 접근임을 확인하였다.

그러므로 본 연구에서 얻은 주요 결과는 다음과 같다. 한글을 영문으로 전자(轉字) 하는 기존 알고리즘에 비해 영문을 한글로 전자(轉字) 및 음성 전사(轉寫)하는 방식을 활용하여 상표 호칭 유사성 판단의 정확성과 효율성이 크게 향상했다. 이는 국내에서 출원되는 상표임을 고려하여 한국어의 음운론적 특성을 그대로 전자(轉字)에 반영함으로써 보다 유사한 상표를 효과적으로 찾아낼 수 있었기 때문이다. 또한 상표심사기준 지침서에 기반 하여 상표 유사 판단에 있어 심사관이 수행하는 모든 절차를 최대한 고려함으로써 한국인이 판단하는 발음의 유사성과 가장 유사한 결과를 도출하였다. 이는 상표 심사 과정에서의 혼동 가능성을 줄이고 심사의 신뢰성을 높이는 데 기여할 것으로 기대된다. 마지막으로 인공지능 기술을 활용하여 상표 호칭 유사성 판단 분야에서 새로운 연구 방향을 제시하였다.

그러나 본 연구에는 몇 가지 한계가 존재한다. 첫째, SmartG2P 모델은 상표 데이터를 기반

으로 학습된 모델이 아니기 때문에, 상표 특유의 발음 변형이나 예외적인 발음을 완벽하게 반영하지 못할 수 있다. 이는 상표에서 흔히 나타나는 창의적인 철자 변형, 신조어, 의도적인 발음 왜곡 등이 모델에 충분히 반영되지 않을 수 있음을 의미한다.

또한, 상표에는 '요부(要部)'가 존재하며, 이는 상표심사기준 지침서에서도 중요한 요소로 언급되고 있다. 심사관은 상표의 요부를 추출하여 유사성을 판단하는데, 본 연구에서는 이러한 요부를 고려한 실험을 수행하지 못하였다. 이는 상표의 주요 부분이 소비자에게 미치는 영향력을 충분히 반영하지 못하였음을 의미한다.

마지막으로 본 연구는 상표의 호칭 유사성에 초점을 맞추었으며, 시각적 요소(외관)나 의미적 유사성(관념)은 고려하지 않았다. 이는 상표 전체의 유사성 판단에서 중요한 요소들을 일부 배제한 것으로 볼 수 있다.

이러한 한계를 극복하기 위해, 향후 연구에서는 다음과 같은 방향으로 확장하고자 한다. 상표 호칭 거절 쌍 데이터 등 상표 심사 과정에서 사용되는 데이터를 모델 학습에 활용하여, 상표 특유의 발음 변형과 예외적인 사례들을 더욱 정확하게 반영할 계획이다. 이를 통해 모델의 정확성을 높이고, 상표 호칭 유사성 판단의 정확도를 한층 더 향상할 수 있을 것이다. 특히 한글, 영어, 복합 문자 상표를 세분화해 각 특성에 맞는 모델을 생성 및 실험, 상표 사례별 검색 결과를 분석 및 연구할 예정이다.

또한 상표의 요부를 자동으로 식별하고 분석할 수 있는 방법론을 연구할 것이다. 이를 위해 상표를 공백이나 특수문자 등을 기준으로 분리하여 각 구성 요소를 개별적으로 분석하고, N-gram 방식 등 다양한 방법을 활용하여 세밀하게 비교하는 방식을 도입할 수 있다. 심사관이 실제로 상표를 주요 부분으로 분리하거나 중요하다고 생각하는 상표 요소를 재검색하여 호칭 유사성을 판단하는 요부 검색 방식과 동일한 접근을 적용할 수 있다. 이에 따라 늘어나는 비교 연산과 복잡한 구조에서 오는 한계를 보완하기 위해 Rozinek와 Mares(2024)[26]가 제안한 가우시안 가중치 기반 컨볼루션 방식을 적용한 Convolutional Jaro 및 Convolutional Jaro-Winkler 접근법(ConvJW)에 대한 추가적인 연구가 필요하다. Baloi et al. (2024)[27]와 같이 GPU 기반 유사도 측정 및 머신러닝 기법을 결합함으로써, 복합 문자 및 음성 전사의 다양한 변형을 더욱 정밀하게 반영하고자 한다. 이들 최신 기술은 대규모 데이터셋에서도 빠르고 정확한 유사도 평가를 가능하게 하여, 상표 심사 과정에서의 혼동 가능성을 크게 줄이는 데 기여할 것으로 기대된다.

또한 딥러닝 기반의 멀티모달[28] 상표 유사성 판단 방법으로 확장하여, 상표의 시각적 요소(외관)와 의미적 요소(관념)까지 통합적으로 고려하는 것이 필요하다. 이는 상표의 전체적인 유사성을 평가하는 데 필수적이며, 상표에 관한 실제 인식을 보다 정확하게 반영할 수 있다. 예를 들어, 시각적으로 유사한 로고나, 문자는 다르지만 통념적으로 유사한 상표는 사람들에게 혼동을 줄 수 있으므로, 이러한 요소들을 함께 검색 및 분석하는 것이 중요하다.

결론적으로, 본 연구는 상표 호칭 유사성 판단 분야에서 인공지능 기술의 활용 가능성을 입증하였으며, 상표 호칭 검색의 정확성과 효율성을 높이는 데 기여하였고, 향후에도 관련 연구를 지속하여 정확하고 효율적인 상표 검색 기법을 연구 개발할 예정이다.

---

26) Ondřej Rozinek & Jan Mareš, "Fast and Precise Convolutional Jaro and Jaro-Winkler Similariy", 2024 35th Conference of Open Innovations Association (FRUCT), IEEE, 2024, pp. 604-613.

27) Aurel Baloi et al., "GPU-based similarity metrics computation and machine learning approaches for string similarity evaluation in large datasets", *Application of Soft Computing*, Vol.28(2024), pp. 3465-3477.

28) 멀티모달은 텍스트, 이미지, 음성, 비디오 등 서로 다른 유형의 데이터를 동시에 사용하여 보다 정교하고 깊이 있는 정보 처리를 가능하게 하는 기술.

참고문헌

단행본(국내 및 동양)

특허청 상표디자인심사국 상표심사정책과, 「상표심사기준」, 2024. 5. 1. 개정, 특허청, 2024.

학술지(국내 및 동양)

문성민 외 2인, "한국어 발음 변환기(G2P)의 현황과 성능 향상에 대한 언어학적 제안", 「언어와 정보」, 제26권(2022).

서창덕·김회율, "문자기반 유사상표 검색을 위한 가중치 부여 근사매칭", 「전자공학회논문지-CI」, 제37권 제1호(2000).

이지연·백우진, "질의로그 데이터에 기반한 특허 및 상표검색에 관한 연구", 「정보관리학회지」, 제23권 제2호(2006).

정상원·정기창, "문자열 유사성 알고리즘을 이용한 공종명 인식의 자연어처리 연구 – 공종명 문자열 유사성 알고리즘의 비교", 「한국건설관리학회 논문집」, 제21권 제6호(2020).

학술지(서양)

Aurel Baloi et al., "GPU-based similarity metrics computation and machine learning approaches for string similarity evaluation in large datasets", *Application of Soft Computing*, Vol.28(2024).

Barton Beebe & Jeanne C. Fromer, "Are We Running Out of Trademarks? An Empirical Study of Trademark Depletion and Congestion", *Harvard Law Review*, Vol.131 No.4(2018).

Jae Sung Lee & Key-Sun Choi, "English to Korean statistical transliteration for information retrieval", *Computational Linguistics*, Vol.12 No.1(1998).

Perumal Pitchandi & Mathivanan Balakrishnan, "Document clustering analysis with aid of adaptive Jaro-Winkler with jellyfish search clustering algorithm", *Advances in Engineering Software*, Vol.175(2023).

S C Cahyono, "Comparison of document similarity measurements in scientific writing using Jaro-Winkler Distance method and Paragraph Vector method", *IOP Conference Series: Materials Science and Engineering*, Vol.662(2019).

학위논문(국내 및 동양)

전종수, "자연어 처리 기법을 활용한 문자열 발음검색 모형에 관한 연구", 광운대학교 대학원, 석사, 2019.

판례

대법원 2005. 11. 10. 선고 2004후2093 판결.

연구보고서

김남수, "소량 데이터만을 이용한 고품질 종단형(End-to-End) 기반의 딥러닝 다화자 운율 및 감정 복제 기술 개발", 서울대학교, 2022.

이대호 외 2인, "딥러닝 기반 텍스트 상표 검색 서비스", 세진마인드, 2018.

기타 자료

고숙현·이재성, "문맥을 고려한 유사 외래어 검출 알고리즘의 성능 향상", 제19회 한글 및 한국어 정보처리 학술대회, 2007.

Ondřej Rozinek & Jan Mareš, "Fast and Precise Convolutional Jaro and Jaro-Winkler Similariy", 2024 35th Conference of Open Innovations Association (FRUCT), IEEE, 2024.

Petar Rajkovic & Dragan Jankovic, "Adaptation and Application of Daitch-Mokotoff Soundex Algorithm on Serbian Names", XVII Conference on Applied Mathematics, Novi Sad, Serbia, 2007.

William E. Winkler, "Advanced Methods For Record Linkage", Proceedings of the Section on Survey Research Methods, American Statistical Association, 1994.