

인용 구문 분석을 통한 피인용 논문 정보 서비스 설계

정한민*, 김태홍**, 최희석*

*한국과학기술정보연구원 융합서비스센터

**한국한의학연구원 지능화추진팀

jhm@kisti.re.kr, thkim@kiom.re.kr, choish@kisti.re.kr

Design of an Information Service for Cited Articles through Citation Analysis

Hanmin Jung*, Taehong Kim**, Heeseok Choi*

*Convergence Service Center, KISTI

**Intellectual Information Team, KIOM

요 약

인용 분석은 정보 검색 위주의 단순 학술정보 서비스를 고도화시킬 기회를 제공한다. 그렇지만, 대부분의 연구가 커뮤니티, 연구자, 논문 간 인용 지수 중심의 분석에 초점을 맞추고 있어, 인용 특성 파악을 통한 피인용 논문 정보 서비스를 제공하기에는 어려움이 있다. 본 연구는 논문 내 인용 문구 분석을 통해 피인용 논문의 인용 특성을 구조적으로 제시할 수 있는, 피인용 논문을 위한 정보 서비스 가능성을 파악하는 것을 목표로 한다. 구글 스킴라를 통해 확보한 인용-피인용 논문 원문들을 대상으로 인용 유형 비율 분석과 인용 확산 분석을 수행하는 방식으로 피인용 논문 특성을 파악하였으며, 정보 서비스에 반영할 수 있는 가능성을 확인하였다.

1. 서론

인용은 논문에 반드시 포함되어야 하는, 학문 발전 과정에서 매우 중요한 요소이다. 이러한 중요성으로 인해, 다양한 관점에서의 인용 분석 연구가 수행되어 왔다. 학회지 전체를 대상으로 인용 지수 기반으로 분석하는 연구[1], 특히나 논문의 인용 관계를 네트워크화하고 인용 강도를 분석하는 연구[2], 전체 또는 특정 커뮤니티 연구자들의 피인용 특성 및 결과를 분석하는 연구[3-5], 인용과 연구 분야와의 연관 관계 분석을 시도한 연구[6], 논문 내 참고 문헌과 매칭되는 피인용 논문을 인식하는 연구[7] 등이 그 대표적인 예이다.

국내 대표 과학기술 학술정보 서비스인 ScienceON에서도 인용/피인용 관계를 특정 논문 중심으로 제공하는 ‘논문 타임라인’ 서비스를 제공하고 있다((그림 1) 참조)¹⁾. 이 서비스는 해당 논문을 설명할 수 있는 태그 클라우드, 해당 논문의 인용/피인용 논문들, 그리고 그들의 주요 키워드들을 시각적으로 제공함으로써 기존 정보서비스들의 단순 참고문헌 제공을 한 단계 넘어서는 정보 서비스라고 할 수 있다.

본 연구는 여기에서 한 발 더 나아가 논문 내 인용 문구 분석을 통해 피인용 논문의 인용 특성(예. 인용 부분, 인용 이유, 인용 유형 등)을 구조적으로 제시할 수 있는 피인용 논문을 위한 정보 서비스 설계 가능성을 파악하고자 한다. 즉, 커뮤니티 수준보다 한 단계 더 깊이 들어간 논문 수준의 서비스가 ‘논문 타임라인’이라면, 본 연구는 원문 내 구문(Phrase) 수준으로 한 단계 더 들어가서 인용 특성을 분석하는 데 초점을 맞추고 있다.



(그림 1) ScienceON 논문 타임라인 서비스 예시

2. 인용 구문 및 인용 특성 분석

본 연구는 새로운 피인용 논문 정보 서비스의 가

1) <https://scienceon.kisti.re.kr/aiq/mlls/selectAIQMllsList.do>

능성과 실효성을 파악하는 데 목적을 두고, 구글 스칼라에서 “딥러닝”으로 검색한 결과²⁾ 중 10회 이상 피인용된, 피인용 논문 4편과 이들을 인용한 인용 논문 45편을³⁾ 대상으로 소규모 인용 구문 분석을 수행하였다[8-11]. 해당 피인용 논문들은 2014~2017년까지 매년 1편씩 포함하고 있어 연도별로 치우치지 않으며, 모두 다른 학회에 게재되어 특정 학회(커뮤니티) 특성에 영향 받지 않아 적절한 샘플링이라 판단하였다. 인용 구문 분석은 (표 1)과 같은 인용 유형을 기준으로 한다.

(표 1) 인용 유형

인용 유형	인용 목적
용어	인용 문구 내 용어 자체 참조
수식	수식 또는 그 구성 요소 참조
정의	인용 문구 내 용어 정의 참조
특징	알고리즘이나 구조에 대한 설명 참조
사실:연구	관련 연구나 사례 참조(연구에 초점)
사실:방법론	관련 방법론이나 적용 분야 참조(방법론에 초점)

다음은 인용 유형 기반 인용 구문 분석 결과를 통해 얻을 수 있는 분석 서비스의 예시이다.

(1) 인용 유형 비율 분석: 각 피인용 논문에 대해 인용 유형의 비율을 분석하면, 조사(Survey) 연구, 특정 분야에 기존 알고리즘을 적용한 연구, 실증 및 사례 연구(Empirical Study) 등으로 분류하여 연구 성향 파악에 도움을 줄 수 있다.

(2) 인용 확산 분석: 매년 지속적으로 인용이 일어나는지, 매년 새로운 학회에서 추가로 인용되는 지 등의 인용 확산 양상을 통해 해당 논문의 인용 현황을 파악하는 데 도움을 줄 수 있다.

3. 결론

본 연구는 논문 내 인용 문구 분석을 통해 피인용 논문의 인용 특성을 구조적으로 제시하고 단순 메타데이터를 넘어서는 분석 기반 정보 서비스를 설계할 수 있는 기반을 제공할 수 있다는 데 의미가 있다.

향후 연구에서 대규모 인용 논문 집합을 대상으로 자동 인용 문구 추출 및 분석 가능성을 확인하고, 커뮤니티 확산 지수 등 새로운 정보 서비스 지표 개발을 시도할 예정이다.

2) <https://scholar.google.com/> (2020.9.14. 검색 기준)
 3) 전체 48편은 중복 포함이며, 이 중 두 편은 서비스가 중지되었으며, 한 편은 동일 논문을 인용하지만 제목이 달라 다른 논문인 것처럼 검색된 경우로, 이들은 분석에서 제외함.

참고문헌

[1] 이건남, 이종범, “한국실과교육학회지의 인용 분석”, 한국실과교육학회지 25(1), 2012.
 [2] 이민정, 김용대, 장원철, “특허 인용 네트워크 분석”, 응용통계연구 29(4), 2016.
 [3] 김미진, “생명과학 분야 연구자들의 공동연구 유형별 연구 생산성과 인용 성과 분석”, 한국문헌정보학회지 52(3), 2018.
 [4] 서선경, 최호남, 김병규, 최선희, 김정환, “Cited-by Linking 데이터 기반 자연과학 및 공학 분야 학술논문 인용 패턴 분석”, 정보관리학회지 33(2), 2016.
 [5] 이재운, “2008-2014 KCI 인용지표 변화 분석”, 한국정보관리학회 학술대회 논문집, 2016.
 [6] 광철완, “빅데이터 연구 논문의 주제 분야 연관 관계 분석: 동시 인용 관계를 적용하여”, 정보관리학회지 35(1), 2017.
 [7] 김재훈, 김순영, 임석중, 황혜경, “학술논문과 참고문헌의 자동매핑 사례 분석”, 한국콘텐츠학회논문지 19(11), 2019.
 [8] 안성만, “딥러닝의 모형과 응용사례”, 지능정보연구 22(2), 2016.
 [9] 이창기, 김준석, 김정희, 김현기, “딥 러닝을 이용한 개체명 인식”, 한국정보과학회 동계학술발표논문집, 2014.
 [10] 김지원, 표현아, 하정우, 이찬규, 김정희, “다양한 딥러닝 알고리즘과 활용”, 한국정보과학회지 33(8), 2015.
 [11] 이우식, “딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가지수 방향성 예측”, 한국데이터정보과학회지 28(2), 2017.
 [12] 김종우, 윤기현, 허진혁, 전광일, “합성곱 신경망을 이용한 언어장애인용 문장 인식”, 한국정보과학회 한국소프트웨어융합학술대회논문집, 2019.
 [13] 김종립, 신강욱, 홍성택, 이호현, 조현식, “LSTM을 활용한 수도권데이터 정제 알고리즘에 관한 연구”, 한국통신학회 하계융합학술발표회논문집, 2018.