

검색의도 파악을 위한 질의어 관계유형에 관한 사례연구

A Case Study on the Types of Queries' Relations for Recognizing User intention

권순진* · 김원일** · 유성준**

Soonjin Kwon, Wonil Kim and Seongjoon Yoo

* 세종대학교 디지털콘텐츠학과

** 세종대학교 컴퓨터공학과

요 약

본 연구는 정보 검색(Information Retrieval)과정에 있어 검색 기술의 적합성을 향상하기 위하여, 질의어 사이의 유용한 관계를 드러내도록 사례를 분석하고, 질의자의 의도를 파악할 수 있게끔 구체화하도록 연구한 것이다. 이를 위하여 먼저, 질의어가 가지는 어휘 의미적 연구 분야와 존재론적 연구 분야의 관련 연구들을 분석하였으며, 국내·외의 어휘 의미론적 네트워크 사례와 정보 검색 기술이 적용된 사이트의 실제 데이터를 분석하여 관계 유형을 추출하고 분석하였다. 다음으로는 일반적으로 검색자가 직면하는 검색 상황에서 자주 발생하는 문제를 중심으로 문제점을 정의하여 해결 방안을 모색하였다. 현행 검색 기술에서 색인어와 질의어를 단순 비교하여 결과를 쏟아주는 검색은 사용자를 혼란하게 하기 때문에 개선이 필요하고, 질의자의 의도에 맞는 질의 결과를 줄 수 있도록 지능적 검색으로 개선할 필요가 있다. 문제점 해결 방안에서, 두 질의어 사이의 관계를 드러냄으로써, 검색자의 의도를 인식하고 식별 및 처리할 수 있는 방안이 필요하였다. 질의어들에 관한 실제 사례를 분석하고 관계 유형을 9가지로 분류함으로써, 관계 유형을 디자인하는 방법을 적시하였으며, 관계 유형의 명칭 부여와 관계 역할의 명칭을 부여할 수 있는 방법과 제한점도 예시하였다.

키워드 : 의도 검색, 검색 질의어, 질의어 관계 유형, 질의어 관계 유형 사례, 관계 유형 식별

Abstract

IR (Information Retrieval) systems have the methods that compare relationships between query and index to identify document that may be fit to the user's query keyword. However, the methods usually ignore the importance of relations that are not expressed in the query. Therefore, in this study, we describe how to refine the queries' relation from keyword and to reveal the hidden intent. A useful relationship between query and keyword in IR with studied and we classified the tion fromrelation. Firstfromall, we did researchmrelated on semantic relationship and ontolhiical researchmin foreign and domestic research, and also analyzed semantic network practices, information retrieval technolhiy, extracted and classified the tion fromrelationships s' relasite's real-world datamin whichminformation retrieval technolhiin fare applied. Next, we souiht to solve the problems occurred frequently i' relasituation that searchers tioically face. I' relacurrent search technolhiy, the mesh searchmresult fare poured by simply comparin ina query with index terms. Therefore, the need for an intelligent search fittn inusers' intent is required. The relationships between two queries to re hiddee and identify relasearcher's intent have to be revealed. By analyzn inthe practical cthes s' queries and classifyn inthem into nine kind fromrelationship tion , we proposed the method to design relation revealn inand role namn i, and we have also illustrated limitations of that methods.

Key Words : Intention driven search, Information retrieval query, Query relation type, Relation type case, Relation type identification

1. 서 론

정보 검색 기술은 통계적 벡터 모델과 자연어 처리를 기

반으로 질의어와 색인 문자열을 비교하여 그에 일치하는 문자열이 포함된 문서를 찾아 보여 줄 수 있도록 발전해왔다. 그러나 현행 검색 기술은 질의어를 색인어와 비교하여 그 결과를 반환하므로, 질의자의 의도와 관계없는 문서를 대량으로 보여 주면서 재검색하는 비효율성 문제가 있다.

이러한 문제를 해결하기 위하여, 정보의 중요성보다는 정보의 필요성 관점에 입각하여 생각해 볼 필요가 있다. 이러한 요구는 정보 검색 전문가 그룹(3rd Search Technology Summit 2010) 130 명의 검색에 대한 요구에서도 알 수 있다. 그러기 위해서 질의자의 의도와 질의어들의 의미적 관

접수일자 : 2010년 10월 18일

완료일자 : 2011년 5월 4일

+ 교신저자 : 김원일(wikim@sejong.ac.kr)

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국 연구재단의 지원을 받아 수행된 기초연구사업(No.2010-0028046)의 성과임

계에 관한 연구가 필요하며, 질의어에 관한 실제 발생 데이터를 분석해 보고, 정보요구자의 의도를 유추할 수 있도록 관계의 유형을 연구하는 것이 필요하다. 시맨틱 기술에 대한 연구는 자원과 시간의 소요가 다대하므로 그 연구 성격 및 평가와 별개로 본 연구를 병행하는 것이 필요하다.

정보 검색(Information Retrieval) 과정에 있어서, 질의자에게 적절한 답변을 주기 위해서는 질의자의 의도를 파악할 수 있어야 한다. 이는 검색창의 질의어(쿼리:query)만으로 의도를 파악하기 위해 질의어들의 관계를 인식할 수 있어야 함을 뜻한다. 두 질의어(two queries)의 관계는 개념 내외의 관계, 그리고 관계의 구성과 관계의 계층 등으로 이루어진다. 질의어들의 관계를 추론하는 것은 매우 중요하지만 식별 및 처리를 위한 인식이 쉽지 않다.

본 연구에서는 질의 의도 관련 연구를 살펴보고 주요기관의 정보 검색 질의어 데이터 사례를 분석해 봄으로써, 질의어 사이의 유용한 관계를 드러낼 수 있게 하고 구체화할 수 있는 요소를 제시하는데 의의를 두고자 한다. 이는 검색자의 의도를 파악하기 위한 방법으로서 질의어의 관계 유형을 정교화 하는 첫 단계라는 의미가 있다. 아울러 관계 유형의 명칭과 역할 부여하는 방법을 예시함으로써 의도 기반 검색 기술이 발전하는데 기여할 수 있다는 의의를 기대해 보고자 한다.

2. 질의 의도 분석 관련 연구

캐롤 쿨싸우(2000)의 연구에 의하면 정보 탐색 과정에서 '사용자의 행동에 관한 예측은 표면적으로 드러나지 않는 관찰이 불가능[1]'하다 하였다. 정보 탐색 과정의 부분으로서 정보 검색 과정도 마찬가지이다. 질의어들의 관계를 추론하는 것은 매우 중요한 작업이지만, 식별 및 처리는 더욱 어렵다. 질의어들의 관계성은 애매하거나 미묘한 어떤 것으로 남아 있어 포착하기 어려운 상태이기 때문이다. Lee & Alan Liu(2005)는 쿼리 변수 추출 및 목표 식별에 대해 '의도 추출 과정에서 주요 과제'라 하였다[2]. Khoo and Na(2007)는 '관계성 유형에 대한 이해와 심층 연구가 필요[3]'하다고 하였다. 관계성 유형에 대한 연구는 검색 사용자의 의도를 명확히 하고 이를 추출하는데 크게 기여할 것으로 보이지만 그 만큼 어려운 연구이다.

검색 기술 발전에 있어서 사용자 의도(그림1)를 포착하기 위한 노력이 계속되고 있다.

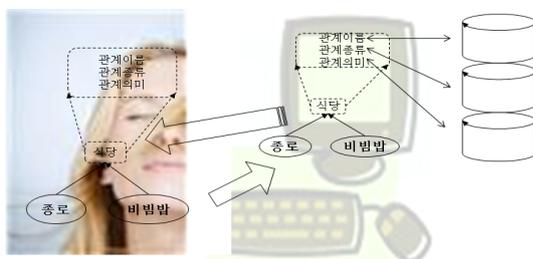


그림 1. 질의어에 드러나지 않는 관계의 인식
Fig. 1. Identifying relation hidden in queries

검색 기술 분야의 노력의 예로, Jansen & Booth & Spink(2007), Adar & Teevan & Dumais(2008) 등은 질의어 데이터 로그로부터 정보 탐색적 위치적 거래적 의도를 파악[4][5]하려고 노력하였다. 그러나 질의어들의 관계가 내포

하는 중층적 구조를 드러내지 못했으며, 사용자 정보(추구) 목표에 따른 질의어 관계 유형을 정립하는 데에는 미흡했다.

2.1 질의어 관계의 특성

관계가 그 자신만으로 존재할 수 없고 두 개념을 연결해야 할 때, 개념 및 엔티티들은 관계들의 통합적 부분이다. 개념 또는 개체들 사이의 결합은 다른 유형들로 범주화 추출 또는 개념화 될 수 있고, 다른 결합들로부터 구분될 수 있으며, 의미를 할당할 수 있다. 그러나 개념의 의미와 그들을 표현하는 언어로부터 분리된 관계를 분석하는 것은 어렵다.

Murphy(2003)가 밝힌 관계의 특성으로는 쉽게 만들어지며, 쌍을 이루고, 의미나 맥락을 다양하게 하며, 다른 단어에 영향성이 있으며, 그리고 개방성, 다양성, 패턴 및 규칙성이 있고, 언어의 차이가 없이 유형이 같으면 동일한 의미 관계를 가지는 특성[6][7]이 있으나, 의도를 유추하기 위한 관계 명칭과 관계 역할을 부여할 수 있는 특성 표현에는 미흡하다.

의미관계는 상위 문서와의 관계, 문서 간의 관계, 문서 내 문장 간의 관계, 문장 내의 질 또는 구 간의 관계, 단어와 단어 간의 관계, 2 개 질의어 간의 관계 등을 들 수 있다. 이러한 관계의 간단한 경우가 단어 간의 관계이기 때문에, 본 연구에서는 가장 기초적인 2 개 단어 간의 의미적 관계를 분석 대상으로 하여 의도의 중층적 구조와 접목할 수 있는 지점을 모색해 보고자 한다.

2.2 의미론적 관계의 논리적 속성

Cruse(2004), Sowa(1984)는 의미론적 관계는 논리적으로 몇 가지 속성이 있음[8][9]을 밝혔다. 즉 단어 자신에게 엔티티를 연관 짓는다면 재귀적이며, 관계의 두 해당자가 양 슬롯을 점유할 수 있다면 그 관계는 대칭적이다. 이는 유사어에서 나타난다. ISA 관계라면 타동적이다. 관계의 양편이 알려져 있고 한 편이 고정되어 있다면 일대일 관계이다. 두 단어 사이의 관계는 의미론적 관계를 더 간단한 관계 또는 두 단어의 의미 측면 사이의 관계 요소로서 메모리에 표현할 수 있다.

Alba & Hasher(1983)는 인간의 기억에 있는 일반적인 지식은 스키마로 불리는 관계의 구조로 조직화되는 것으로서 모델링하였다[10]. Minsky(1974)는 스키마를 구현한 것 중의 하나로 프레임을 정의한 바 있다[11]. 기본적으로 라벨링 슬롯 조합이며, 각각은 프레임에 참여하는 역할을 지시한다.

Rumelhart & Ortony(1977), Schank & Abelson(1977), Schank(1982)는 이벤트 유형에 서브 이벤트의 순서를 나타내는 일시적 요소와 더불어 프레임을 스크립트라 하였다[12][13][14].

Butcher & Kintsch(2003), 그리고 Whitney, Budd, Bramucci & Crane(1995)는 인간의 이해 모델에 있어 프레임, 스크립트 및 스키마 스토리가 중요한 역할을 함을 보여준다[15][16].

2.3 관계의 유형에 관한 선행 연구들

Lumpers 또는 Minimalists[17]는, 철학적 또는 논리적 원칙에 기초한 일반적 관계를 정의하였다[18][8][19].

2~3 가지 관계 유형(어형 변화적 그리고 통합적인 관계) 연구로 Ferdinand de Saussure (1959) 및 다수의 연구자들이 있고, 3 가지 관계 유형(수식, 분류학, 그리고 대기 행렬)[19]으로 분류하였다.

어형 변화적 관계는, 단어 또는 같은 문장 내에 같은 위치에서 발생할 수 있는 문구의 쌍 사이의 관계들이다[20]. ISA(broader-narrower), 부분-전체(part-whole)와 동의어 관계들은, 우리의 의미론적 기억의 일부인 경향으로 나타나며, 그리고 전형적으로 사전(시소러스)에 사용된다.

“Paradigmatic and syntagmatic relations”로 구분하기도 하였다[21]. 선형적인 또는 영구적인 관계로서 전형적인(paradigmatic) 관계와, 후천적인 또는 변하기 쉬운 관계들로서 통합적인(syntagmatic) 관계로 구분하였다. 통합적인 관계들은, 동일한 문장이나 텍스트[20]에서 병발(중중 가까운 통사론의 위치에서)하는 단어 사이의 관계이며, 선형적 또는 우리가 문장을 만들 때 합성 또는 두 어절 사이에 표현되는 순서 관계이다.

이들 외에도 13 가지 이상의 분류 체계로 연구하거나, 어휘적-의미론적 관계의 집합을 개관한 연구들을 발견할 수 있다[7][18][22].

관계의 유형에 관한 선행 연구들은 유형들이 다양하게 나타날 수 있음을 보여준다. 주로 사용되는 곳은 사전이나 어휘 의미망들에 사용될 목적의 기초 연구들로 보인다. 그러나 본 연구에서는 의미 관계 집합들을 기초로 하여 의도 기반을 만들기 위한 관계 유형의 사례 연구에 목적이 있으므로, 관계의 유형 분류 또한 여기에 집중하도록 한다. 예를 들어 맛집을 찾는 경우나 세미나를 찾는 경우 이다(그림2).



그림 2. 질의어로 인식해야할 관계와 의도
Fig. 2. Relations and Intention Categorization

2.4 의도(모형)에 관한 선행 연구

관계 구조를 분해하여 질의자의 의도에 적합한 문서를 식별하기 위해서는 “관계”를 추출하는 방법이 필요하게 된다. 도메인과 검색자가 처한 상황의 장면, 그리고 특정 목적을 추구하는 의도와 목표 계획 등을 인식할 수 있는 방법이 필요하며, 이런 관점에서 선행 연구에서의 의도모형들을 살펴 볼 필요가 있다.

어휘들의 관계인 사전적 의미에 더하여 의도를 모델화하기 위해 추가 되어야 할 것들이 있다. 장기성(2002)의 인지적 관점에 의한 형식화[23] 연구에 의하면, 특정 상황에 필요한 사건들의 집합체인 프레임, 사건 국면의 시간적 인접성을 정보화하는 패턴인 스키마, 그리고 스크립트에다가 플랜을 추가하고 있다. 즉 플랜은 텍스트 참여자들의 의도된 목표를 실현하는 사건들과 행위들로 구성된 지식 형태이며, 스크립트는 이 두 상황의 정보를 저장하고 있다.

김선희(1989)의 연구 중 데이빈슨이나, 또는 Case(2004)의 연구에 의하면 행위가 어떤 목표나 의도를 가질 때 의미가 있고 유용하다[24]. 따라서 기존의 정보검색 질의어를 선형적으로만 보지 않고, 다층적 관계인 의도 모델을 상정한다면, 기존 관계 유형들에서 간과했던 것들이 발견 내지는 복원될 수 있을 것이다.

사용자 목표와 계획 모델에 관한 연구로 Carberry(1988),

Lee & Liu(2005)를 들 수 있다[25][2]. 이들의 연구를 보면, 의도 기반을 갖출 수 있는 질의어 관계와 의도 유추 가능성을 볼 수 있다.

의도 구조 모델 : 지능형 또는 스마트 시스템이 간단한 명령이나 방향에서 사용자의 의도를 이해할 수 있다면, 그것은 시스템을 더 영리하고 더 지능적으로 만들 것이다. 이것은 시스템의 사용을 보다 편리하고 유용하며 효율적이 되도록 할 수 있다.

사용자의 의도는 신념과 의사결정에 필수적으로 함께 등장한다. 질의자의 입력 키워드로부터 미리 정의된 시스템을 가동할 수 있다면 여기로부터 사용자의 의도를 추출할 수 있을 것이다. 의도는 목표가 수행될 경우에 달성될 수 있으며 목표를 달성할 수 있도록 계획이 전제되어야 한다. 목표 모형은 계획과 시작점 그리고 도달점 방향과 경로 등으로 구성할 수 있다. 다만 이 경우에 시스템이 사용자에게 너무 많은 요구를 하게 되면 사용자가 견딜 수 없거나 또는 시스템의 복잡도가 증가하므로 사전에 구조적으로 잘 정의된 의도 모델이 준비되어야 한다. 도메인 온톨로지 영역적 의미가 식별되어야 하며, 사용자 프로파일의 일부 학습이 보장되어야 할 것이다.

2.5 외국의 적용 사례

영어와 일본어에 대해 구축된 가장 광범위한 어휘 의미론적 네트워크는 FrameNet과 일본어 프레임넷 그리고 WordNet과 일본판 워드넷이라 할 수 있는 EDR 등이 있다 [27][28][29][30][31].

프레임넷은 개념 기반의 어휘 분석을 통해 통사론과 의미론의 연결을 형식화하고 전자 사전을 이용하여 자연어 처리가 가능하도록 한 것이다.

WordNet은, 약 150,000개 이상의 영어 명사, 동사, 형용사와 부사들을 포함하는 어휘 데이터베이스이며, synsets라 불리는 동의어적 단어들을 조직화했고, 각각은 어휘 개념을 나타내고 있다. 그것의 디자인은 인간의 어휘의 메모리에 관한 심리-언어학적(psycholinguistic) 이론에 근거한다. 그것의 구축은 어휘집이 어떻게 어휘-의미론적 관계로 구조화하는지 방법에 대한 통찰력을 제공해 준다. 예를 들어 명사들은 주로 ISA 및 부분-전체 관계로 구조화된다. ; 명사는 형용사의 특성 링크와 연결되어 있으며 동사에 기능 링크가 된다. ; 형용사는 주로 반의어(antonymy)에 의해 연결되며; 동사 중 가장 빈번한 관계는 하위 관계(troponymy)이고, 어떤 면밀한 마무리를 표현하는 것이다. WordNet에서 인코딩된 동사들 간의 다른 관계는 어휘적 함의(entailment 즉, 코 고는 것은 잠을 수반)이며 인과적 관계(즉 show/see, feed/eat, have/own)이며 그리고 반의어(antonymy)이다. 관계 명사는 700 여 개다(예:difference, disjunction).

일본판으로 EDR (Electronic Dictionary Research Institute)이 있다(ge.nii.ac.jp , www.nict.go.jp). 이는 일영-영일 기계 번역을 위하여 일본 정부와 8 개의 컴퓨터 관련 업체가 구축하였다. 여기서는 개념 관계를 5 가지의 종류로 나누고 총 36 가지 관계를 규정하였다. 격 관계(Case relations), 사건 관계(Interevent relations), 의미 관계(Semantic relations), 한정 관계(Restriction relations), 의사 관계(Pseudo-relations)등이다.

WordNet의 성공에 이어, EuroWordNet이 구축되었다. 이는 여러 유럽 언어를 망라 할 다국어 어휘 데이터베이스

이다[32]. EuroWordNet은, WordNet을 본떴지만 어휘 의 미론적 관계의 풍부한 집합을 사용한다. 예를 들어, 인과적 관계들은 비사실 서술적 원인적 관계들(non-factive causal relations) 즉 하나의 이벤트는 또 다른 사건을 유발 하는 것 같으나 필연적으로 그렇지 않은, 즉 search→ find } 및 사실 서술적(factive) 원인적 관계(원인적 관계는 필연적으로 지킨다:the causal relation necessarily holds, 즉 죽이다kill→죽다die)로 나누어진다. 원인적 관계도 결과(즉 찾다(search)→발견하다(find))를 유발하는 부주의한 원인적 관계들로부터 그것을 구별하기 위한 의도의 속성으로 표시될 수 있다. 대체적으로 동의어 반의어(antonymy) 및 부분-전체 관계의 5 가지 유형도 EuroWordNet에 사용된다. 더욱이 관계의 집합은 개념들 (예를 들어 비행기는 프로펠러 제트분출 그리고 여러 부분품-날개,코,후미 그리고 문짝-의 결합을 가지는 등의 개념집합 들) 중의 관계성을 지시하는 결합과 분리의 속성으로 레이블 될 수 있다. 세계적으로 다양한 언어에 대한 FrameNet과 WordNet들이 구축 되어가고 있는 중이다.

2.6 국내의 적용 사례

한국어 프레임넷 구현 사례[33]는 아직 평가가 이르다. 워드넷 프로젝트들은 글로벌 워드넷 협회 웹 사이트 (<http://www.globalwordnet.org/>)에 나열된다. 여기에는 국내의 부산대학교에서 수행하는 한국어 워드넷도 포함되어 있다.

한국의 21 세기 세종계획(<http://www.sejong.or.kr/>)의 사례 중에서 방성원 호정은 김종인(2001)의 연구는 의존명사의 정보 표상 구조[34]에 관한 언급이 있다. 의존 명사, 대명사, 수사 범주에 속하는 언어 단위의 사전 기술 구조에 정보 값을 체계적으로 명시할 수 있는 정보 항목과 표상 구조를 설정한 예를 발견할 수 있다. 명사 수식 또는 제약에 관한 목록을 총 148 개로 나열하였고 표상 방식으로는 조사 제약 등 9 가지를 예를 들어 설명하고 있다.

21 세기 세종 계획[35]의 목적은 한국어로 세계의 모든 문화 정보를 처리할 수 있는 고도의 기술을 개발하고자 1998 년 부터 추진된 중장기 사업임에도 불구하고, 실질적인 고도화 단계가 필요할 것으로 본다. 고도화 단계 및 활용 단계에서는 본 연구가 기여할 수 있을 것이다. 다양한 프레임넷과 워드넷들을 기초로 어휘 의미들에 대한 데이터 베이스 들이 구축되어 가고 있지만, 이들을 활용하는 단계에 있어서 필요하리라 예상되는 의도 기반의 모형을 구축하는 연구가 수행되어야 할 것이다.

의도 기반의 정보 검색 질의어 관계 유형 연구는 새로운 분야이다. 외국의 경우에 비해 국내의 대용량 질의어 데이터를 분석해 보려는 연구 결과는 많지 않으며, 질의어 관계 유형에 대한 분석도 드물다. 국내 네이버의 경우는 그나마 2008 년 부터 질의어를 분석하여 해마다 질의어 백서로 일 반에 공개(네이버 트렌드 연감 2008, 2009)하고 있지만, 관계 분석은 공표가 이루어지지 않고 있다. 국회 도서관과 한국 과학기술 정보 연구원 에서는 질의어를 중요 자원으로 인식 및 관리하고 있고, 사용자 의도를 반영할 수 있도록 노력하고 있으나 매체 유형별로 고정된 상태로 표현되어 있고, 질의 의도를 표현하거나 질의어에 실시간으로 연동되는 수준에는 미치지 못하고 있다. 이것은 실제 사례 분석에서 다를 것이다.

3.질의 의도를 인식하는 문제

두 질의어로 질의자의 의도를 유추할 수 있도록 ‘한정하는 관계’를 식별하고 추출하는 방안과 의도의 기준을 설정할 수 있는 요소를 제시하는 것이 본 연구의 주된 문제이다. <표1>의 문제 사례 경우처럼 “종로” 또는 “미국”에 관한 색인어 집합보다는 두 질의어에서 탈락된 의도가 중요하다. “종로” 지역에 있는 “비빔밥”을 잘하는 식당의 관련성은 “식당”이라는 관계 속성이 매개체가 될 수 있으며, “미국” 지역의 “세미나”의 관련성은 학습 모임이 매개체가 될 수 있다. 종로 비빔밥의 경우를 더욱 상세히 보자면,

- ① 종로 지역에는 종각이 있으며 세종로로 불리는 등의 다양한 의미를 포함하고 있기 때문에 종로를 색인어로 포함하는 문서는 다수 존재한다.
- ② 비빔밥 또한 요리의 한 종류이기도 하며, 비빔밥만으로도 돌솥/콩나물/전주/얼무... 등 많은 종류가 나오고, 구글 검색 결과로는 14,200,000 여 개가 나오지만, 모두가 “비빔밥”이라는 낱말이 들어 있는 문서를 가져와서 쏟아 놓은 결과를 보여 주게 된다.

표 1. 문제 사례
Table 1. Problem case

질의 사례	현행 검색결과	사람이 할 경우	의도
종로 비빔밥	① “종로”와 “비빔밥”이 들어 있는 문서를 찾아 모두 나열하여, ②그 목록을 보고 사람이 선택하도록 함	종로에 있는 비빔밥 잘하는 식당을 찾고, 그중 전주 비빔밥을 선호할 경우, 해당하는 식당에 관한 관련 정보를 획득함	원하는 답은 “종로”와 “비빔밥”이 아니고, 비빔밥 잘 하는 “식당”이 답 이며, 그 외는 “관련” 정보에 해당함
미국 세미나	대부분 종교 행사나 지나간 일정을 보여줌	미국 지역에서 향후 개최되는 세미나 일정을 탐색	참석 일정, 참석 가치 판단

<표1>예에서 보는 바와 같이, 정보검색 사용자의 의도에 맞는 답(종로 비빔밥의 경우는 “잘하는 식당 추천 정보”)을 찾아주어야 함에도 불구하고, “종로” 또는 “비빔밥”이라는 색인어가 포함된 수 많은 문서들을 추출하여 보여 줄 가능성이 많다. 앞의 질의어가 뒤의 질의어에 대해서, 의미적 포함 관계로서 범위를 한정하는 “관계”와 그 관계의 속성을 가진 문서를 추출하여 검색 질의자의 의도에 맞게 제공하여야 함을 알 수 있다.

따라서 “종로”와 “비빔밥”이 들어있는 문서를 나열하기 보다는, 종로와 비빔밥의 관계성을 유추하고 그 관계구조를 분해하여 질의자의 의도에 적합한 문서를 식별해야 한다. “관계”를 추출하며 적절한 기준에 의해 의도를 선택하여 질의 적합성을 높일 수 있는 방법이 필요하다. 그러기 위하여 맛집 도메인과 식사가 필요한 검색자가 처한 상황의 장면, 식사를 하고자하는 목적 추구 의도와, 종로에 있는 비빔밥 잘하는 식당을 찾는 목표 계획 등을 인식할 수 있는 방법이 필요하다.

의도기반 질의어 관계 인식 : 시스템 구성요소와 처리방법으로는 목표의 내용과 방향 계획에, 관계의 식별, 목표모델

의 탐색, 사용자 의도 표현, 후보 목표 모델 선택의 단계가 필요할 것이다(그림3).



그림 3. 의도 모델 선택
Fig 3. Selecting Intention model

질의어들의 관계를 식별하고 유추된 목표 모델을 의도 기반의 목표 모델 중에서 탐색하여 사용자 의도를 포함한 질의어로 재생성하는 과정이 필수적이다. 이 과정을 거친 후에 기존 정보 검색 서비스 에이전트의 활용이 이루어진다면 정보 검색의 사용자 요구 적합성은 향상될 것이다.

4. 질의어 관계 실제 사례 분석

실제 정보 검색 기술이 활용되는 상황에서 어떻게 질의어 관계를 인식할 수 있을 것인지를 알아볼 필요가 있다. 사례 데이터 분석을 위하여 관련 전문 기관 4곳에 데이터를 요청하였으며, 그중에서 한국 과학기술 정보 연구원과 국회도서관의 질의어 데이터 현황은 다음과 같다.

4.1. 한국 과학기술 정보 연구원의 데이터 분석 사례

20xx 년 10월의 NDSL(<http://www.ndsl.kr/>) 총 검색횟수 411,254 건 중 검색 질의어가 2 개인 것은 10,787 건이다. 이들 중에서 검색 반환 값을 클릭으로 반응한 것에서 중복 데이터를 제거한 것으로 총 3,203 건을 분석 대상으로 하였다.

현행 시스템에 고정된 의도 표현으로 논문을 찾은 것이 1,184 건이며, 동향 분석은 823 건이었다. 산업표준 16건 연구 보고서 269건 특허 350건 학위 논문 561건 등이었다<표 2>.

표 2. NDSL 검색의도
Table 2. Intentions of Queries in NDSL

논문	1184	37.0
동향분석	823	25.7
산업표준	16	0.5
연구보고서	269	8.4
특허	350	10.9
학위논문	561	17.5
합계	3203 건	100.0 %

2 개 단어 질의어 및 의도별 질의 반환 값 : 분석 대상 데이터 중 질의어 관계와 의도는 <표3>과 같다. <표3>의 첫 예에서 볼 수 있는 것처럼 “ ‘1인’ ‘창업’ ”의 논문을 찾는 의도로 검색을 요구한 질의는, ‘1인’은 ‘창업’을 수식하는 관계이다. 이런 관계는 문법적인 관계일 수도 있지만, 정보 검색의 적합성을 향상시킬 수 있는 제약 관계 또는 의존적 관계로 보아도 될 것이다.

<표3>에서와 같이, 관계의 유형은 다양하지만 주로 4가지 유형에 주목하였다.

첫째는 앞의 질의어가 뒤의 질의어를 수식하거나 의미적으로 제한하는 관계 유형이며,

표 3. 과학기술 정보 사이트의 질의어 관계
Table 3. Types of Queries' Relation, on NDSL Site

2단어 검색어	DB이름	관계 유형
1인; 창업	논문	1.제한/수식하는 관계
가상현실; 시장	동향분석	1.제한/수식하는 관계
노즐; 유량측정	산업표준	1.제한/수식하는 관계
2차전지; 전기자동차	연구보고	1.제한/수식하는 관계
2액형 폴리우레탄	특허	1.제한/수식하는 관계
3D XML	학위논문	1.제한/수식하는 관계
1차냉각수응력부식균열	논문	2.원인결과 관계
환경호르몬; 출산	동향분석	2.원인결과 관계
1-MCP;1-METHYLCYCLOPROPENE	논문	3.약어보완
2액형;2액형폴리우레탄	논문	4.구체화하여 보완
나노; 나노구조	산업표준	4.구체화하여 보완
핵융합;상온핵융합	연구보고	4.구체화하여 보완
간농양; 김두	논문	5.특정 논문/저자
해상풍력; 정지영	연구보고	5.특정 논문/저자
강문수;위생개폐장치	특허	5.특정 논문/저자
가시오갈피;한중수	학위논문	5.특정 논문/저자
자전거; BICYCLE	산업표준	6.한글 영문 병용
감성공학; 감성	논문	7.개념포괄 탐색
고분배; 고분배	논문	8.에러수정
황일순; 허섭	연구보고	9.저자 포괄

둘째는 앞의 질의어가 뒤의 질의어의 원인이 되는 관계 유형이고,

셋째는 구체화하여 보완하는 관계 유형, 넷째는 약어를 상세히 보완하는 관계 유형이었다.

대부분의 사례가 앞의 질의어가 뒤의 질의어를 수식하거나 제한하는 관계였다. 둘째와 셋째 넷째 사례의 경우에서도, 앞의 질의어가 뒤의 질의어를 제한하는 관계로 볼 때 검색 적합성을 향상할 수 있을 것으로 보인다. 한국 과학기술 정보 연구원에서는 사용자의 검색 의도를 6 가지로 반환 받도록 되어 있다. 이는 KISTI의 기관 특성에 의해 미리 기준을 설정한 것으로 보이나 정보 검색의 목표나 계획은 보이지 않는다.

4.2. 국회도서관 데이터 분석 사례

국회 도서관(National Assembly Library)의 보유 데이터는 분류 체계가 저장 위주로 구축되었으며, 분류 체계 또한 이를 만족하기 위해 2 계층의 구조로 구성되어 있다<표 4>. 20xx년 10월 한 달 간 질의한 질의어는 총 3,185,187 건이었다. 이 중에서 질의어 2개로 이루어진 총 10,387 개의 데이터 중에서 분석의 편의를 위해 숫자나 특수 문자 영어 한자로 된 질의어는 제외하였으며, 중복 데이터 또한 제외하였다. 가나다 항 순으로 각 10 개 내외씩의 데이터를 발췌 후 분석하였다. 분석 결과 중 KISTI와 마찬가지로 매체 유형별로 7가지를 볼 수 있지만 의도 구조는 보이지 않음을 알 수 있다.

표 4. 국회도서관 검색의도

Table 4. Intentions of Queries in NAL

1.단행본	도서
2.학위 논문 3.학술지	비도서 자료 세미나 자료 학위 논문
	국내 국외
	인터넷 자료

2개 단어 질의어와 관계 유형 : 한국 과학기술 정보 연구원의 분석 예와 같은 기준으로 국회 도서관의 질의어 관계유형을 분석하였다<표5>.

표 5. 국회도서관 사이트의 질의어 관계 샘플

Table 5. Types of Queries' Relation, on NAL Site

2단어 검색어	관계 유형
가격차별화 전략	1.제한/수식하는 관계
가계조사 국민건강보험	2.원인/결과관계
가곡 교원	7.개념포괄탐색
가공식품 학교급식	2.원인/결과관계
가구디자인 재활용	1.제한/수식하는 관계
가구디자인 친환경	1.제한/수식하는 관계
가구소득 진로	2.원인결과 관계
나노 신소재	1.제한/수식하는 관계
나노입자 포장	7.개념포괄탐색
나라살림 교육예산	7.개념포괄탐색
나목 국어	7.개념포괄탐색
나무 나뭇가지	4.구체화하여 보완
나뭇잎 문양	7.개념포괄탐색
다구치 직교표	5.특정 논문/저자
다극체제 국제질서	7.개념포괄탐색
다리 다리프로젝트	4.구체화하여 보완
다문화 가정	7.개념포괄탐색

4.3.분석 결과

두 기관의 임의의 데이터 89 건과 140 건을 분석한 결과, 9 가지의 의미 관계 유형이 나왔으며, 표본을 늘릴 경우 더욱 다양한 관계 유형이 나올 수 있고, 수작업으로는 많은 시간이 소요되어 89 건과 140 건으로 제한하였다.(89건은 46건이 수식 관계이며, 이는 전체의 51.7%에 해당함-표6 참조)

색인어를 단순히 비교하는 검색 기술로는 질의자의 검색의도 파악이 어려우며, 국회 도서관과 같이 영역 지정이 미리 되어 있는 사이트의 데이터를 분석해 본 경우에도 탐색하는 의도별 목표 문서 종류가 6가지로 나타났다(표4). 검색자의 의도 종류는 보이지 않음을 확인할 수 있었다.

검색 기술의 정확도를 향상할 수 있는 방법으로는 가장 중요한 것으로 의도 기반을 어떻게 구축해 나가는가에 달려 있으며, 둘째로는 목적이나 목표에 관한 색인을 문서에서 만들고, 검색 창에도 별도로 목적이나 목표 등 의도에 관련된 키워드 색인 비교 창을 신설하도록 표준화해 나갈 필요

가 있다. 따라서 현재의 웹에서 의미망을 확장해 나가는 것과 더불어 목적이나 목표 등의 의도에 관련된 키워드 색인을 만들어 나가는 작업도 중요한 작업으로 병행해 나가야 할 것이다.

그럼으로써, 문서 색인 또는 메타데이터를 생성할 때 목적과 목표 또는 의도와 의도를 달성하기 위한 계획 등을 별도로 색인할 수 있는 방법을 표준화하고 기술을 개발하여 정보 검색 적합성 향상에 기여하게 될 것이다.

표 6. 관계 유형

Table 6. Relation Types

1).A가 B를 제한 또는 수식하는 관계	46
2).A가 B의 원인이 되는 관계	3
3).A는 약어이며 B는 Full-Spell로서 보완	1
4).A를 더 구체화하여 보완하는 B	9
5).특정 논문 주제와 저자	5
6).한글 영문 병용	11
7).개념포괄 탐색	9
8).에러수정	1
9).저자 포괄	4

4.4.관계 식별 방법

의도 기반을 형성함에 있어 중요한 요소로 관계를 식별하는 것이 출발점이다. 질의어 사이에는 복합적인 관계가 있으나, 관계는 검색창에는 입력이 되지 않으므로, 사람이 인식하는 관계를 미리 온톨로지로 정의하여 기계학습 시켜야 한다. 이럴 경우에, ①관계 유형과 구조를 미리 정의하여 구축해 놓고, 질의어 입력 시에 관계를 동적으로 시각화하여 표시함으로써, 입력자가 관계를 선택하게 할 수 있게 하면 미리 정의된 관계 온톨로지를 이용하여 다양한 의도에 대처할 수 있을 것이다. ②검색자가 관계를 입력하게 하는 방법은, 검색사용자에게 불편을 초래하여 기피하거나, 역효과가 예상되므로, 추가 입력을 최소화할 필요가 있다.

따라서 관계 온톨로지를 구축하여 관계를 식별할 수 있도록 하는 방안의 경우에 필요 작업은, ①질의어의 개념을 추출 ②명사와 명사의 관계(예 : 포함 관계) 추출, 관계 종류/명칭/유형/범위에 대한 온톨로지를 구성 ③두 질의어의 “관계 추출” 프로세스(규칙의 구성 및 실제 표현) 가동 ④선택 규칙 기술 (“ ‘종로’ ‘비빔밥’ ”으로 “ 비빔밥 잘하는 ‘식당’ ”을 선택하는 규칙을 기술) 등이다. 해당되지 않는 결과는 의도 범주를 설정하여 별도 루틴으로 변경한다. 즉 현행 검색 프로세스를 거치도록 한다.

4.5. 관계 유형 디자인 방법 예시

관계 유형은 국어사전류에서 추출한 용어 리스트를 별도로 준비하여 찾아 낼 수 있게 된다. 질의어 개념과 관계를 표현하기 위해서는 여러 형태의 관계를 도출하고 관계 유형으로 정의하여야 한다. 예를 들어 서울 지역 아래에 종로 광진 두 지역이 있고, 비빔밥은 돌솥 비빔밥과 전주 비빔밥의 두 유형을 미리 지정하며, 관계 유형도 미리 정의하여야 한다. : ① 서울-종로 관계는 하향 관계, 서울-광진 관계는 하향 관계, ② 종로-전주 비빔밥 관계는 하향 관계, 종로-돌솥 비빔밥의 관계는 상향 관계, ③ 전주비빔밥-식당이름 “맛난식당” 연결 관계 등이다.

관계 유형의 명칭 부여 : 관계 유형의 명칭은 명사를 사용한다. 동사형으로 명명하게 되면, 관계 유형이 방향성을 갖게 되므로 에러의 원인이 될 가능성이 크므로 피한다.

관계 역할의 명칭 부여 : 역할 명칭도 명사로 부여하되, 해당 역할이 인스턴스가 되는 클래스를 나타내도록 한다. 서울 종로 등이다. 관계 역할의 유형도 기술해야 한다. 서울은 종로의 지리적 상위 클래스, 종로는 서울의 지리적 하위 클래스 등으로 부여한다. 역할의 개수는 복잡성을 피하기 위해 하나만 기술한다. 여러 개일 경우는 한 개로 변환하는 것이 바람직할 것이다.

위의 방법으로 관계의 표상을 구성하고, 검색 질의어로 종로와 비빔밥이 입력될 경우를 가정하여, 처리 과정을 나타내면 다음과 같다.

“종로” “비빔밥”이 검색창에 입력되면, ① 지역관계인 서울-종로 관계를 확인한다. 종로와 광진이 둘 다 지역관계에 해당되며, 비빔밥은 해당되지 않는다. 여기서는 지역 관계를 확인하게 된다. 다음은, ② 상하 관계인 종로-전주 비빔밥을 확인한다. 전주만 해당되며, 돌솥은 해당되지 않는다. 여기서는 전주 비빔밥을 확인하게 된다. ③ 연결 관계인 전주 비빔밥-“맛난식당”을 연결해준다. ④ “종로 비빔밥”으로 질의 시에, “맛난 식당”을 표시해주도록 한다. 이후는 후속 검색 프로세스가 현행과 동일하다.

5.의도 인식에 필요한 요소와 개념적 구성

검색질의어로부터 사용자의 의도를 인식하기 위해서, 대체로 다음 3 단계가 필요하며(그림4), 프로세스와 요소항목은 아래와 같다.



그림 4. 질의 의도 파악에 필요한 개념 요소
Fig. 4. Concepts for Recognizing intention

의도 인식에 필요한 요소 항목과 프로세스

두 질의어를 입력하여 정보를 검색하는 사용자의 의도는 입력 시에 탈락된다. 탈락되는 질의자의 의도를 모델링하여 미리 시스템에서 저장하고, 검색 적합성을 향상할 수 있는 중요한 질의어 관계들을 복원할 수 있을 것이다. 처리 과정과 요소들을 나열하면 다음과 같이 요약될 수 있다.

첫째, 입력된 질의어의 전처리 및 관계 식별
둘째, 해당 영역 결정
셋째, 목표 구조 참조 및 생성
넷째, 목표 모델로 사용자 의도 참조 및 생성
다섯째, 선택된 목표와 질의어 관계 유형의 비교 후 매칭
여섯째, 의도 질의어를 재생성 후 검색

6. 결론 및 향후 연구

본 연구에서는 정보 검색(Information Retrieval) 과정에 검색 기술의 적합성을 향상하기 위하여 질의어 사이의 유용한 관계를 드러내는 사례를 분석하고 질의자의 의도를 파악할 수 있게 구체화하는 방법을 연구하였다.

이를 위하여 먼저 질의어가 가지는 어휘 의미적 연구 분야와 존재론적 연구 분야의 관련 연구들을 분석하였으며, 국내·외의 어휘 의미론적 어휘 네트워크 사례와 정보 검색 기술이 적용된 사이트의 실제 데이터를 분석하여 관계 유형을 분석하였다. 다음으로는 일반적으로 검색자가 직면하는 검색 상황에서 자주 발생하는 문제를 선정하여 해결 방법을 모색하였다. “종로” “비빔밥”을 질의할 때, 수 많은 색인어 집합에 관련된 문서를 쏟아주는 비효율성 문제를 지적하였고, 문제점 해결 방법에 있어서는, 두 질의어 사이의 관계를 드러냄으로써, 검색자의 의도를 인식하고 식별하며 처리해야 할 필요성과 요소를 적시하였다.

사용자의 일상적 또는 전문적 정보 추구 행태별로 효과적인 상호 작용적 또는 계층 구조적 상황 설정은 후속 연구에서 다루게 될 것이다.

본 연구 결과는 일반적으로 구축되는 시맨틱 기술에 의도 기반을 접목하여 활용할 수 있으므로 정보 검색 기술이 진화하는데 기여할 것으로 기대한다.

연구용 질의어 데이터를 협조해주신 한국 과학기술 정보 연구원의 김재수 단장과 국회 도서관의 마을순 서기관, 이병련 서기관께 감사드립니다.

참 고 문 헌

- [1] 캐롤쿨싸우 저 김효정·이병기 공역, 정보탐색과정론, p145~150~, 한국디지털도서관포럼, 2000
- [2] C-H.L. Lee and A. Liu, “Modeling the Query Intention with Goals,” *Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA'05)*, vol. 2, pp. 535-540, 2005
- [3] C.S.G. Khoo and J.C. Na, “Semantic Relations in Information Science,” *Annual Review of Information Science and Technology*, vol. 40, pp. 157-228, 2007.
- [4] B.J. Jansen, D.L. Booth and A. Spink, “Determining the informational, navigational, and transactional intent of Web queries,” *Information Processing and Management*, vol. 44. no. 3, pp. 1251-1266, 2008
- [5] E. Adar, J. Teevan, and S.T. Dumais, “Large Scale Analysis of Web Revisitation Patterns,” *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pp. 1197-1206, 2008, Available: dl.acm.org, 2011.3.1
- [6] M.L. Murphy, *Semantic relations and the lexicon: Antonym, synonymy, and other paradigms*, Cambridge: Cambridge University Press, 2003
- [7] M.L. Murphy 저, 임지룡 윤희수 옮김, 의미관계와 어휘사전 : 반의관계, 동의관계, 기타 계열들, 박이정, 2008.

- [8] J.F. Sowa, *Conceptual structures: Information processing in mind and machine*, Reading, MA: Addison-Wesley, 1984.
- [9] A. Cruse, *Meaning in language: An introduction to semantics and pragmatics*, Oxford: Oxford University Press, 2004
- [10] J.W. Alba and L. Hasher, "Is memory schematic?" *Psychological Bulletin*, vol. 93, no. 2, pp. 203-231, 1983, Available: www.psych.utoronto.ca, 2011.3.1.
- [11] M. Minsky, "A framework for representing knowledge," *AI Memo*, no. 306, MIT A.I.Lab., 1974
- [12] D.E. Rumelhart and A. Ortony, "The representation of knowledge in memory," pp. 99-135, 1976, Available: www.cs.northwestern.edu, 2011.3.1.
- [13] R.C. Schank and R.P. Abelson, *Scripts, plans, goals, and understanding*, Lawrence Erlbaum Associates, 1977.
- [14] R.C. Schank, *Dynamic memory: A theory of reminding and learning in computers and people*, Cambridge University Press, 1982.
- [15] K.R. Butcher and W. Kintsch, "Text comprehension and discourse processing," In I.B. Weiner (Ed.) *Handbook of psychology*, vol. 4, pp. 575-595, 2003, Available: books.google.co.kr, 2011.3.1
- [16] P. Whitney, D. Budd, R.S. Bramucci and R.S. Crane, "On babies, bathwater, and schemata: A reconsideration of top-down processes in comprehension," *Discourse Processes*, vol. 20, no. 2, pp. 135-166, 1995
- [17] M.W. Evens, *Relational models of the lexicon: Representing knowledge in semantic networks*, Cambridge University Press, 1988
- [18] J.F. Sowa, *Knowledge representation: Logical, philosophical, and computational foundations*, Pacific Grove, CA: Brooks/Cole, c2000
- [19] O. Werner, "How to teach a network: Minimal design features for a cultural acquisition device or C-KAD," In M.W. Evens (Ed.), *Relational models of the lexicon: Representing knowledge in semantic networks* (chap. 6, pp. 141-166). Cambridge University Press, 1988
- [20] R.E. Asher, (Ed.), *The encyclopedia of language and linguistics*, Oxford: Pergamon Press, 1994
- [21] F.W. Lancaster, *Vocabulary control for information retrieval*, Arlington, VA: Information Resources Press, 1986
- [22] 이희자 우재숙, "국어사전의 '관련어' 연구," pp. 171-172 '관련어 정의의 포함관계', *한국사전학* 제7호, 2006.4.
- [23] 장기성, "정구조의 형식과 의미 관계" *언어과학연구*, 제22권, 2002
- [24] 김선희, "데이빌슨의 의도(Intention)이론", *철학*, 제31권, 1989 (및 D.O. Case의 *정보추구행태론*(2004)의 pp. 59~60의 의도와 행위 참조)
- [25] S. Carberry, "Modelling the User's Plans and Goals," *Computational Linguistics*, vol. 14, no. 3, 1988.
- [27] 미국 Berkeley 대학교, "FrameNet project", Available: framenet.icsi.berkeley.edu, 2011.3.1
- [28] 일본 게이오 대학교, "Japanese FrameNet", Available: jfn.st.hc.keio.ac.jp, 2011.3.1
- [29] C. Fellbaum, (Ed.), *WordNet: An electronic lexical database*, Cambridge, MA: MIT Press, 1998
- [30] G.A. Miller and C. Fellbaum, "Semantic networks of English," *Cognition*, vol. 41, no. 1-3, pp. 197-229, 1991
- [31] G.A. Miller, "WordNet: A lexical database," *Communication of the ACM*, vol. 38, no. 11, pp. 39-41, 1995
- [32] A. Alonge, N. Calzolari, P. Vossen, L. Bloksma, I. Castellon, M.A. Marti, and W. Peters, "The linguistic design of the EuroWordNet Database," *Computers and the Humanities*, vol. 32, no. 2-3, pp. 91-115, 1998
- [33] 김태석, "병렬 코퍼스를 기반으로 하는 한국어 프레임넷 구현에 관한 연구," 동의대 석사학위 논문, 2006
- [34] 방성원 호정은 김종인, "세종 의존명사/대명사/수사 전자사전의 정보표상 구조," *2001년도 제13회 한글 및 한국어 정보처리 학술대회 (2001.10)*, *한국정보과학회 언어공학연구회 학술발표 논문집*, pp. 341-347
- [35] 국립국어원, "21세기 세종계획, 전자사전 개발," Available: www.sejong.or.kr/, 2011.3.1

저 자 소 개



권순진(Soonjin Kwon)

1985년: 영남대학교 경영학사
 2008년: 중앙대학교 정보경영 석사
 2010년: 세종대학교 디지털콘텐츠학과 박사
 사수료

관심분야: 의도 기반 정보검색, 지능형 시스템, 질의어관계
 Phone: 010-8008-3219
 E-mail: sjkwon@sju.ac.kr



김원일(Wonil Kim)

1988년: Southern Illinois 대학교 컴퓨터 공학과 학사
 1990년: Southern Illinois 대학교 석사
 2000년: Syracuse 대학교 컴퓨터 정보학과 박사
 2002~2003년: 아주대학교 BK교수
 2003년~현재: 세종대학교 디지털콘텐츠학과 교수

관심분야 : 인공지능, 정보보안, 멀티미디어 콘텐츠
 Phone : 02-3408-3795
 E-mail : wikim@sejong.ac.kr



유성준(Seongjoon Yoo)

1982년 : 고려대 전자공학과(학사)
1990년 : 고려대 전자공학과(석사)
1996년 : 시라큐스대 전산학과(박사)
2002년 ~ 현재 : 세종대학교 컴퓨터 공학과
교수

관심분야 : 자연어처리, 정보검색, 데이터마이닝, 소셜프로
그래밍, 스마트폰 앱

Phone : 02-3408-3755

E-mail : sjyoo@sejong.ac.kr