

## 한국어 정보검색을 위한 질의어 생성에 관한 연구

이덕남<sup>1</sup>, 박인철<sup>2\*</sup>

### A Study on Generation of Query for Korean Information Retrieval

Deok-Nam Lee<sup>1</sup> and In-Chol Park<sup>2\*</sup>

**요 약** 현재 인터넷의 급속한 개발과 함께 인터넷을 통해 사용자들의 질의 성향을 올바르게 파악하여 사용자들에게 보다 질 좋은 정보를 제공해 줄 수 있어야 한다고 말해도 과언이 아니다. 따라서 한국어 질의어에 대해 형태소 분석과 구문 분석을 통한 키워드 사이에 의미 연관성을 생성하는 방법을 제안한다. 이 접근은 단순한 키워드나 키워드 사이에 단순한 조합 보다 의미 연관성을 적용하였다.

그 결과로서 사용자들의 질의 성향을 더욱 정확하게 반영할 수 있는 질의를 생성하고 현존하는 질의 형태에 대한 몇 가지 문제를 해결 수 있는 효율적인 한국어 정보검색을 위한 사용자 의도가 포함된 계층적 질의어를 생성하는 방법을 제안하고자 한다.

**키워드:** 한국어, 한국어 질의, 의미 분석, 구문분석, 정보검색

**Abstract** At present age, great many informations are no exaggeration to say that supply information of better quality to users depend on that grasp correctly user's query intention through internet along with fast development of internet. Therefore, this thesis suggest that generating meaning relation between keywords with that result by passing through morpheme analysis and syntactic analysis about Natural Language Query.

This approach is implied more meaning relation than query by simple keyword or simple combination between keywords. Therefore, it is going to permit much more efficient information retrieval because of solving problem about existent query form, and generating query that user's query intention is reflected more correctly.

**Key words:** Natural Language, Natural Language Query, Meaning Analysis, Syntax Analysis, Information Retrieval

## 서 론

사용자 질의에 대해서 검색되는 정보의 양이 많다고 해서 항상 양질의 정보가 제공되는 것은 아니다. 이는 오히려 검색의 질을 낮추어 사용자들의 판단을 방해하고 있으며, 이러한 문제들은 대부분 현재의 검색 엔진들이 사용하는 방법론의 한계점 때문에 나타나는 현상들이다. 따라서 현재의 정보검색 시스템들이 내세우는 목표는 얼마나 많은 데이터를 보유하고, 사용자에게 더 많은 검색 결과를 제공하기 보다는 사용자의 요구에 적합하고, 정확

한 정보를 신속하게 원하는 사용자에게 제공하는가에 있다[1].

검색 엔진에 의한 검색 결과들이 불확실한 또 다른 이유는 사용자가 사용하는 용어와 검색엔진이 인식하는 용어의 의미적 차이 때문이다. 그래서 이상적인 정보검색 시스템은 각 문서와 질의의 내용을 완전하게 이해하는 것이지만 실제로 이것은 거의 불가능하므로 대부분의 정보검색 시스템들은 문서들과 질의의 내용에 근접하는 구조화된 방법을 사용해야 한다.[2]

정보 검색 시스템의 효율성을 평가 분석하는 기준으로 재현율(Recall)과 정확률(Precision)을 주로 이용하는데, 재현율은 실제로 관련된 문서 중 검색 시스템에 의하여 검색된 문서의 비율을 의미하고, 정확률은 검색 시스템에 의하여 검색된 문서 중 관련된 문서의 비율을 의미한다. 이상적인 정보 검색 시스템에서는 높은 재현율과 높은

본 논문은 2005학년도 호원대학교 교내학술연구조성비에 의해 연구되었음.

<sup>1</sup>전주대학교 정보기술컴퓨터공학부

<sup>2</sup>호원대학교 컴퓨터게임학부

\*교신저자: 박인철(cpark@sunny.howon.ac.kr)

정확률을 동시에 유지하는 것이 바람직 하지만, 실제의 경우 대부분 재현율과 정확률은 반비례 현상을 보인다 [1][3].

메타 형태가 아닌 이들 개별적인 검색 엔진들의 경우 비교적 낮은 커버율 때문에 질의에 대한 재현율 개선을 위해 다양한 질의 선택 방법을 이용하고 있다. 메타 검색 엔진들을 제안하는 대표적인 사이트로는 국내의 "zagia", "seek", 국외의 "MetaCrawler", "MetaSearch", "Savvy-Search" 등을 들 수 있다. 이와 같은 메타 검색 엔진들은 부분적으로 재현율 문제에 대한 접근 방법을 제안하고 있으나, 여전히 정확률 문제로 고심하고 있다[4-6].

이를 근거로 일반적인 메타 검색 엔진들에 있어서 문제가 되고 있는 정확률 문제를 개선하고, 한국어 처리 기법을 이용하여 사용자의 질의 의도에 적합한 한국어 질의어 패턴 생성에 의한 접근 방법을 본 논문에서 제안하고자 한다. 본 논문은 사용자들이 구체적으로 그들의 검색 의도를 구조화된 트리로 나타내는 기법을 제시하는 Scime와 Kerschberg[4][7-9]가 제안한 아이디어를 기반으로 한다. 이는 한국어 질의 형태를 사용하고 있지 않으나 질의를 사용자가 수작업에 의해 구조화함으로써 사용자의 검색 의도를 질의에 좀 더 구체적으로 반영하고 있다.

본 논문에서는 한국어 질의 유형을 분석하고, 이를 통하여 한국어 질의에 대한 문장 패턴 구조를 생성하여 각 노드에 가중치를 부여하고, 메타 검색 엔진을 통하여 여러 검색 엔진(야후, 알타비스타, 다음, 심마니, 엠파스, 네이버, 라이코스 등)에서 수집해온 정보를 문서의 빈도수와 문서 내 용어 빈도수, 그리고 질의의 각 노드에 부여된 가중치에 의해 재순위화 하여 사용자가 원하는 정보가 상위에 랭크되도록 하고자 한다.

본 논문의 구성은 다음과 같다.

1장은 서론이며 2장에서는 국내 메타검색 현황, 한국어 질의 문장 패턴 유형에 대해 살펴보고 3장에서는 본 논문에서 제안하는 한국어 질의 패턴 기반 메타검색 에이전트에 대해 설명하고 4장에서는 실험 및 평가를 제시하며 5장에서는 결론 및 향후 연구과제로 구성된다.

## 2. 관련연구

### 1. 국내 메타검색 현황

웹상에 존재하는 수많은 정보들 중에서 사용자가 원하는 정보를 정확히 검색하기 위한 많은 연구가 진행되어 왔다[10-13]. 시간이 흐를수록 웹상의 정보량은 기하급수적으로 증가하고 있어서, 이들 정보들 중에서 사용자의 요구에 적합한 정보만을 검색하는 일은 더욱 어려워지고

있다. 이 때문에 현재 정보 검색 시스템에서 사용자의 질의에 의해 검색된 결과에 대한 재현율과 정확률을 개선하기 위한 방안으로 사용자의 질의에 대한 처리에 초점이 맞춰지고 있다[1][13][14][15].

현재 국내에는 메타검색 서비스를 하고 있는 사이트도 적을 뿐만 아니라 성능 면에서도 사용자의 검색 요구에 부합하는 메타검색 엔진은 거의 없는 실정이다. 또한 한국어 질의가 지원되는 메타검색 엔진은 사라지고 있는 실정이다.

## 2 질의 문장 패턴 유형

### 2.1 기본문의 개념

한국어 기본문은 기본적인 문장 패턴에 대한 규정으로서 필수 논항과 서술어로 구성되어 있다[16]. 기본문의 구성체인 논항과 서술어는 일정한 순서를 지니고 있으며, 예외적인 경우 단서조항을 둔다. 문장은 최소한 하나의 주어와 서술어를 지녀야 한다. 다만, 문장이 문장 안에 내포된 경우나 연결어미를 이용하여 문이 접속될 경우 논항의 일부가 생략될 수 있다. 모든 서술어는 일정수의 논항을 요구하며, 동사, 형용사, 혹은 '명사, 대명사, 수사+이다'로 구성된다. 서술어는 부사어, 부사구, 부사절에 의해 수식을 받는다.

### 2.2 기본문의 분류

기본문은 서술어가 요구하는 논항의 수에 따라 1 항술어, 2 항술어, 3 항술어로 나뉜다. 각 술어에 의해 요구된 논항은 그 위치가 각각 정해져 있으며, NP1은 주어 자리, NP2는 목적어 혹은 보어 자리, NP3는 필수 부사어 자리이다.

논항1 : NP1(주어)	-NP+주격조사
논항2 : NP2(목적어/술어)	-NP+목적격조사/보격조사
논항3 : NP3(필수부사어)	-NP+부사격조사

#### 2.2.1 1항술어

(문형1) S → NP1 VP	-1항술어
컴퓨터가 확인된다.	(1)
컴퓨터는 편리하다.	(2)
컴퓨터는 기계이다.	(3)
*확인한다.	(4)
*확인하다. 컴퓨터가	(5)

예문 (1)은 동사를 서술어로 하며, 예문 (2)는 형용사, 예문(3)은 명사에 지정사 '이'가 결합되어 서술어 역할을 하는 예이다. 그리고 예문 (4)는 필수격이 생략되어 비문

이며, 예문 (5)는 고정순서 위반으로 비문이다.

### 2.2.2 2항술어

(문형2) S → NP1 NP2 VP -2항술어

(문형3) S → NP1 NP3 VP -2항술어

xml은 문서구조를 나타낸다.(NP2) (1)

xml이 희망이 되었다.(NP2) (2)

\*유니코드를 xml은 채택한다. (3)

xml은 html과 비슷하다.(NP3) (4)

html과 xml은 비슷하다. (5)

예문 (3)은 고정순서 위반으로 비문이다. 그리고 예문 (4)가 2항술어인 반면, 예문 (5)는 'html과 xml은' NP1이 되는 1 항술어로 본다.

### 2.2.3 3항술어

(문형4) S → NP1 NP2 NP3 VP -3항술어

(문형5) S → NP1 NP3 NP2 VP -3항술어

한국어처리는 언어학을 모태로 여긴 다. (1)

한국어처리는 검색에 편리성을 준다. (수여동사) (2)

한국어처리는 언어학을 전산학적으로 말한다. (발화동사) (3)

\*한국어처리는 언어학으로 모태를 여긴다. (4)

전술한 바와 같이 VP가 수여동사거나 발화동사일 경우 문형 4,5 모두 적용될 수 있다. 그러나 예문 (4)와 같이 수여동사, 발화동사 이외의 동사가 문형 (5)로 쓰였을 경우 비문으로 간주한다.

## 2.3 논항 확대

NP1, NP2, NP3는 앞에 관형어를 선행시키거나 다른 NP와 결합하여 논항을 확대할 수 있다. 관형어는 관형사, NP+(관형격 조사) 그리고 관형절을 포함한다. 관형절에 의한 논항의 확대는 절에 의한 확대이므로 문의 내포에서 다르다.

또 다른 확대의 방법은 접속 조사를 이용한 병렬구성, 접속 부사를 이용한 나열식 구성 그리고 병렬구성으로 확대된 논항이 접속 부사로 결합된 복합 구성이 있다. 결국 각각의 논항 1, 논항 2, 논항 3은 아래와 같은 방법으로 확대될 수 있다.

새 버전 (1)

새 버전의 개발자 (2)

마이크로소프트와 IBM이 공동으로 개발하였다. (3)

basic, C, COBOL 그리고 Pascal은 프로그래머이다. (4)

\*나는 IBM 한국지사 서비스 사업본부에 문의하였다. (5)

위의 예문에서 (1)은 관형사에 의한 논항 확대이고, (2)은 관형격 조사, (3)와 (4)는 각각 접속 조사와 접속 부사를 이용한 논항 확대의 예제이다. (5)의 경우는 다음의 NP 구성 규칙에 의해 비문으로 처리된다.

(규칙 1) NP → (nc|nn)<sup>+≤3</sup>|nb|np

(규칙 2) NP' → NP

규칙 1은 NP 구성 규칙으로써 3 개 이하로 제약된 복합명사를 이루는 연속된 nc(일반명사)나 nn(수사), 또는 nb(의존명사), np(대명사)의 생성을 표현한다. 이후 NP는 규칙 1의 정의를 따른다. 아래의 예문에서 (1)-(4)는 정문이 되며 (5)는 비문으로 간주한다.

넷스케이프 버전 (1)

인터넷 익스플로러 설치 (2)

칠천 팔백 오십라인 (3)

애플릿 하나 (4)

\*인터넷 익스플로러 설치 방법 (5)

규칙 2는 NP를 논항 확대된 NP'로 간주함을 나타내며, 이후 NP'는 논항확대를 표시한다.

이러한 한국어 기본문을 바탕으로 한국어 질의 문장 패턴의 대한 정보를 형태소 분석과 구문분석을 통하여 추출하고자 한다.

## 3. 한국어 질의 패턴 기반 메타 정보검색 에이전트

### 1. 한국어 질의 패턴 기반 메타 정보검색 시스템 구조

[그림 1]은 한국어 질의 패턴 기반 메타검색 에이전트의 전체 구조와 구성요소를 보여주는 중요한 정보의 흐름들을 묘사하고 있다. 이 에이전트는 8개의 서브시스템과 4개의 중요한 정보 저장소로 구성되어 있다.

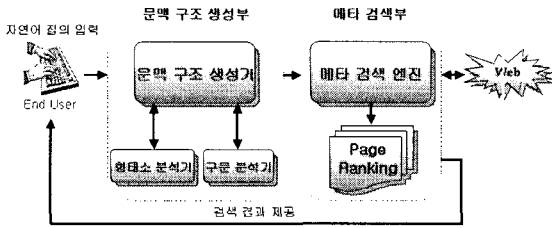


그림 1. 한국어 질의 패턴 기반 메타검색 에이전트

예를 들어, “사무용기기 중에서 사무용가구로 의자와 책상을 찾아줘” 라는 한국어 질의는 질의 패턴 구조를 생성하기 위한 첫 번째 단계로 형태소 분석 과정을 거치게 된다. 형태소 분석은 사용자가 입력한 한국어 질의어 대해 최소한의 의미를 갖는 형태소로 분리하고, 분리된 형태소가 갖는 품사 정보의 추출 및 그 원형을 복원하는 과정이다. 형태소 분석 과정에서 형태소 분석기에 의해 한국어 질의는 “사무용기기/명사 + 중/명사 + 에서/조사 + 사무용가구/명사 + 로/조사 + 의자/명사 + 책상/명사 + 을/조사 + 찾아줘/동사”와 같이 분석된다. 두 번째 단계로 구문 분석 과정을 거치게 되는데 구문 분석은 질의어 대하여 질의를 구성하는 구성 성분의 문법적 역할과 수식 관계를 찾아내어 문장 내에서 수행하는 기능을 분석하는 과정이다. 구문 분석 과정에서 구문 분석기에 의해 자연어 질의는 “사무용기기/관형어 + 중에서/ 부사어(처소-출발점) + 사무용가구/부사어(자격) + 의자/목적어의 나열 + 책상/목적어 + 찾아줘/서술어”와 같이 분석된다. 이 과정에서 각 어절에 대한 수식 관계에 의해 깊이(depth) 정보가 주어지게 된다.

분석되어 생성된 구문 구조는 질의 패턴 생성기를 통하여 구문 분석 결과를 분석하고, 각 노드를 생성하여 노드마다 가중치를 부여하게 된다. 이는 곧 메타 검색부의 입력으로 전달되어 각 검색 엔진에 적합한 불리언 연산 구조에 의해 불리언 질의로 변환된 후 검색 엔진들의 질의 입력부로 전달되어 검색된 결과를 가져오게 된다. 검색되어 온 결과 문서들은 문서 내 단어 빈도수와 질의 패턴 생성부에서 부여된 가중치에 의해 검색된 결과를 재순회화 하여 사용자의 질의 의도가 좀 더 명확히 반영된 검색 결과를 제공하게 된다.

## 2. 질의 패턴 생성 알고리즘

[표 1]은 정보검색을 위한 한국어 질의에 대해 문법 자질 정보를 이용하여 한국어 질의 패턴 구조를 생성하는 알고리즘이다.

표 1. 한국어 질의 구문분석 알고리즘

```

If(SearchA(출발점을 나타내는 부사어)){
  If(SearchB(부사어를 수식하는 관형어)){
    head = CreateRootNode(관형어의 원형)
    If(SearchA(자격을 나타내는 부사어)){
      CL1=CreateChildNode(자격을 나타내는 부사어의 원형,head)
      // 자격을 나타내는 부사어의 원형으로 Create Level-1 노드
      If(SearchA(목적어)){
        CL2=CreateChildNode(목적어의 원형,CL1)
        // 목적어의 원형으로 Create Level-2 노드
        If(SearchB(대등적으로 연결된 목적어))
          AddChildNode(대등적으로 연결된 목적어의 원형, CL2)
          // 대등적으로 연결된 목적어의 원형을 Add Level-2 노드
      }
    }
  }
}
else{
  If(SearchA(목적어)){
    CL1=CreateChildNode(목적어의 원형,head)
    // 목적어의 원형으로 Create Level-1 노드
    If(SearchB(대등적으로 연결된 목적어))
      AddChildNode(대등적으로 연결된 목적어의 원형, CL1)
      // 대등적으로 연결된 목적어의 원형을 Add Level-1 노드
  }
}
}
else{
  If(SearchA(자격을 나타내는 부사어)){
    head = CreateRootNode(자격을 나타내는 부사어의 원형)
    If(SearchA(목적어)){
      CL1 = CreateChildNode(목적어의 원형, head)
      // 목적어의 원형으로 Create Level-1 노드
      While(SearchB(대등적으로 연결된 목적어))
        AddChildNode(대등적으로 연결된 목적어의 원형, CL1)
        // 대등적으로 연결된 목적어의 원형을 Add Level-1
    }
  }
}
}
}

```

[표 2]의 한국어 질의 출발점을 나타내는 부사격 조사 “에서”와 자격을 나타내는 부사격 조사 “로”, 그리고 대등적 연결어미 나열을 의미하는 “(,콤마)”가 한 문장 안에 존재하는 형태이다. 이는 구문 분석 결과를 바탕으로 앞의 두 과정에서 보인 바와 같이 부사격 조사 “에서”가 나타나는 어절을 수식하고 있는 “사무용기기”가 자격을 나타내는 부사격 조사 “로”를 갖는 어절인 “사무용가구”를 의미적으로 포함하게 되고, 또한 목적어의 나열형 어절인 “의자”와 “책상”은 의미적으로 동등한 위치에서 “사무용가구”에 의미적으로 포함되게 된다.

표 2. 한국어 질의 구문 분석 결과

[질의] 사무용기기 중에서 사무용가구로 의자, 책상을 찾아줘
[찾아줘/찾(24) ASERV (0)]
+ -- ROLE [사무용가구로/사무용가구(1)]
+ -- OBJ [책상을/책상(1)]
+ -- NP-SET [의자/의자(2)]
+ -- SRC [중에서/중(1)]
+ -- NP-MOD [사무용기기/사무용기기(2)]

표 4. 질의 10에 대한 검색 결과

키워드 \ 사용자	1	2	3	4	5	6	7	8	9	10	가중치
사무용기기	0	0	0	-1	2.01	5.01	0	2.01	0	0	8
사무용가구	-1.08	3.99	5.98	6.65	8.23	5.63	0	8.23	9.88	7.30	9
의자	-1.54	5.77	6.65	6.51	6.41	4.81	-1.60	8.41	9.75	8.07	10
책상	5.60	6.05	7.22	8.51	6.39	-1.81	-1.60	8.39	5.40	8.55	10
근로	5.60	-1.43	3.91	5	-1.02	0	-1.60	-1.02	5.57	-1.66	10
사무용품	6.03	3.99	6.70	0	-1.52	5.63	6.63	4.52	5.02	0	9
문어	5.60	-1.43	3.91	0	0	4.81	-1.60	0	0	0	10
변	7.71	0	7.22	0	0	4.81	8.48	0	0	0	10
편류티	0	-1.43	5.00	5	7.08	6.26	0	7.06	5.21	7.26	10
가중치합	40.08	36.08	48.68	33.67	42.64	41.77	33.51	42.64	36.83	35.84	
SCW	+1	+9	+5	+8	+13	-1	0	+11	+8	+6	
WCW	-5	-4	-7	-5	+14	-1	-11	+14	+1	0	
사용자요약	5	11	8	12	18	2	7	19	17	16	
가중치요약	10	15	1	17	-1	6	18	5	13	16	

## 4. 실험 및 평가

### 1. 실험에 사용된 한국어 질의

실험에 사용된 한국어 질의들은 질의에 대한 패턴 구조가 생성 가능한 의미적으로 포함 관계를 지닌 질의 문장 중 [표 3]과 같이 10개의 질의를 이용하였다.

표 3. 실험을 위한 한국어 질의 문장

[질의1] 지리학에 관련하여 지도제작법에 관한 자료를 찾아줘
[질의2] 자연과학 분야 중에서 식물에 관한 자료로 야생화를 찾아줘
[질의3] 사회과학 분야에서 경제용어로 맵스주의에 대해 알려줘
[질의4] 천문학 분야 중에서 별자리로 이철자리를 알려줘
[질의5] 의학 분야에서 민간요법으로 단식요법을 알려줘
[질의6] 포유류 중에서 식육목에 포함되는 데스만에 관한 자료를 찾아줘
[질의7] 중국음식과 관련하여 북경요리에 속하는 수안양로우에 대해 알려줘
[질의8] 무용가 중에서 북한무용가인 최희희에 관한 자료를 찾아줘
[질의9] 동양철학 분야 중에서 인도 고대철학으로 구르파에 대해 알려줘
[질의10] 사무용기기 중에서 사무용가구로 책상, 의자를 찾아줘

현재 국내에 서비스 중인 메타 검색 엔진은 한국어 질의를 사용자 의도에 맞게 지원하지 못하고 있기 때문에 실험을 위하여 한국어 질의에서 구문 분석기에 의해 생성되는 키워드들을 상용 메타검색 엔진의 입력으로 제공하여 검색된 결과와 비교하여 실험을 진행하였다. 각 질의들에 대한 검색된 문서중 상위 30개 이내의 문서들에 대한 순위와 사용자 순위, 그리고 본 논문에서 제안하는 방법에 의한 순위를 비교 실험 하였다.

### 2. 실험 결과 및 평가

아래의 [표 4]는 질의 10에 대한 상용 메타검색 엔진의 검색 결과에 대한 단어 빈도수와 가중치, 그리고 질의 가중치를 반영한 검색 결과 순위 및 순위 변화폭을 보여주고 있다.

본 논문에서 제안하는 한국어 질의 패턴 구조에 의한 재순위화 결과는 WRC(weighted ranking change width)의 절대값의 합은 116, 오차범위 문서 수 7개로 상용 메타검색 엔진에 의한 검색 결과보다 사용자 순위에 더 근접한 결과를 보이고 있다. 전체 실험 질의들에 대한 실험 결과는 [표 5]에 보이고 있다.

표 5. 전체 실험 질의들에 대한 순위 변화 결과

질의	총 검색 문서 수	사용자 순위에 대한 변화폭 합		오차범위 내의 문서수		전체 문서에 대한 오차범위 내 문서의 비율	
		상용	질의패턴	상용	질의패턴	상용	질의패턴
질의1	27	128	136	13	12	48.19%	44.44%
질의2	26	166	168	12	8	46.15%	30.77%
질의3	20	63	36	9	15	45.00%	75.00%
질의4	24	132	142	8	10	33.33%	41.67%
질의5	23	158	94	6	9	26.09%	39.13%
질의6	18	55	80	8	6	44.44%	33.33%
질의7	19	76	56	7	10	36.84%	52.63%
질의8	25	212	120	4	10	16.00%	40.00%
질의9	23	153	108	7	8	30.44%	34.78%
질의10	22	140	116	3	7	13.64%	31.82%

[표 5]의 결과를 살펴보면, 전체 10개의 질의에 대한 실험 결과, 6개 질의에 대하여 변화폭의 합이 감소하는 것을 알 수 있다. 이는 6개 질의에 대해서 사용자 순위와의 거리가 가까워짐을 의미하게 되며, 상용에 비하여 사용자 순위와 거리가 먼 나머지 4개의 질의 중 3개에 대해서는 상용 메타검색 엔진과 비교하여 10 이내의 거리를 보이고 있어 상용 메타검색 엔진의 결과와 큰 차이를 보이지 않는다.

오차범위 내 문서수도 7개 질의에 대해서 상용 메타검색 엔진에 비하여 좋은 결과를 보이고 있으며, 나머지 3개 중 1개 질의에 대해서도 차이가 1문서 차이를 보이고 있다. 이는 질의마다 다소 차이가 있으나 향상되지 않은 질의들도 향상되는 비율에 비해 소폭의 차이를 보이므로 전체적으로 상용 메타검색 결과에 비해 유사한 성능이거나 그 이상임을 알 수 있다.

실험 결과 드러난 특징은, 질의가 길수록 검색 성능이 향상되는 결과를 보이고 있다는 것이다. 이는 질의가 짧을수록 단어 빈도수가 미치는 영향이 커지는 현상에서 기인한 것으로 보인다.

## 5. 결론 및 향후 연구과제

본 논문에서는 사용자 질의 의도가 좀 더 명확히 반영된 검색 결과를 얻기 위하여 두 가지를 제안하였다.

첫째, 한국어 질의에 대한 패턴 구조를 생성하여 질의에 대한 가중치를 부여하는 것이다.

둘째, 부여된 가중치를 이용하여 검색 결과를 재순위화 하여 사용자의 요구에 적합한 문서들이 검색 결과의 상위에 랭크되도록 하는 것이다.

실험 결과, 본 논문에서 제안한 방법에 의한 검색 결과는 사용자 순위에 좀 더 근접한 결과를 보였다.

향후 연구 과제로는 더욱 다양한 한국어 질의 유형에 대한 패턴 구조를 파악해야 할 것이고, 한국어 질의를 분석하여 추출되는 키워드들에 대한 유의어나 다의어에 대한 이용 방법에 대해서도 연구가 더 진행 되어야 할 것이다. 이는 유의어 사전을 이용하여 질의의 키워드를 확장하고, 또한 질의에 있어서 다의어에 대한 제약을 가할 수 있어 사용자의 의도가 더욱 명확하게 반영된 질의를 생성 가능하게 할 수 있을 것이다.

## 참고 문헌

- [1] ETRI IT정보센터, “웹기반 검색엔진의 한글 색인 및 질의 처리 기술 동향”, 2000.12.13
- [2] 박미화, 원형석, 이원일, 이근배, “구문분석에 기반한 한국어 질의로부터의 불리언 질의 생성”, 제10회 한글 및 한국어 정보처리 학술대회, 1998, pp.73-80
- [3] Berthier Ribeiro-Neto, Ricardo Baeza-Yates, "Modern Information Retrieval", Addison-Wesley, 1999
- [4] Lawrence, S. and Giles, C. L., "Accessibility of Information on the Web," Nature, vol. 400, 1999, pp.107-109.
- [5] Wooju Kim, Larry Kerschberg, Anthony Scime, "Learning for automatic personalization in a semantic taxonomy-based meta-search agent", Electronic Commerce Research and Applications, vol.1, 2002, pp.150-173
- [6] W. Kim, L. Kerschberg, and A. Scime, Personalization in a Semantic Taxonomy-Based Meta-Search Agent, presented at International Conference on Electronic Commerce 2001 (ICEC 2001), Vienna, Austria, 2001.
- [7] Scime, A. and L. Kerschberg, "WebSifter: An Ontology-Based Personalizable Search Agent for the Web," International Conference on Digital Libraries: Research and Practice, Kyoto Japan, 2000, pp.493-446.
- [8] Scime, A. and L. Kerschberg, "WebSifter: An Ontological Web-Mining Agent for E-Business," Proceedings of the 9th IFIP 2.6 Working Conference on Database Semantics (DS-9): Semantic Issues in E-Commerce Systems, Hong Kong, 2001.
- [9] Kerschberg, L., et al., "A Semantic Taxonomy-Based Personalizable Meta-Search Agent," Forthcoming in Lecture Notes in Artificial Intelligence, 2002.
- [10] Markus Mittendorfer, Werner Winiwarter, "Exploiting syntactic analysis of queries for information retrieval", Data & Knowledge Engineering, vol.42, 2002, pp.315-325
- [11] Xiaoying Tai, Fuji Ren, Kenji Kita, "An information retrieval model based on vector space method by supervised learning", Information Processing and Management, vol.38, 2002, pp.749-764
- [12] Jose Perez-Carallo, Tomek Strzlkowski, "Natural language information retrieval: progress report", Information Processing and Management, vol.36, 2000, pp.155-178
- [13] Ashwin Rao, Allan Lu, Ed Meier, Salahuddin Ahmed, Daniel Pliske, "Query processing in TREC-6", Information Processing and Management, vol.36, 2000, pp.179-186
- [14] 주형식, “한국어를 이용한 대화형 질의 처리 에이전트의 설계 및 구현”, 전북대학교 석사학위논문, 2000
- [15] 이재윤, 최보영, 정영미, “문헌 자동분류에서 용어가 중치 기법에 대한 연구”, 제7회 한국정보관리학회 학술대회 논문집, 2000, pp.41-44
- [16] 장명길, 김현진, 장문수, 최재훈, 오효정, 이충희, 허정, “의미기반 정보검색”, 정보과학회지, 제19권, 제10호, 2001, pp.7-18

**이 덕 남(Deok-Nam Lee)**

[정회원]



- 1996년 호원대학교 전자계산학과 학사
- 1998년 수원대학교 전자계산학과 석사
- 2000년 전북대학교 전산통계학과 박사수료
- 2002년~2005년 전주대학교 정보기술컴퓨터공학부 객원교수

- 2006년~현재 한국정보통신대학원대학교 청소년교육 분과 전주분원 전임강사

<관심분야>

한국어정보처리, 정보검색

**박 인 철(In-Chol Park)**

[정회원]



- 1984년 전북대학교 전산통계학과 이학사
- 1986년 전북대학교 전산통계학과 이학석사
- 1998년 전북대학교 전산통계학과 이학박사
- 1992년~현재 호원대학교 컴퓨터게임학부 교수

<관심분야>

관심분야 : 한국어정보처리, 정보검색, 시맨틱 웹, 임베디드 소프트웨어