

# 의미정보기반 검색시스템의 설계 및 구현

## Design and Implementation of Information Retrieval System based on Semantic Information

박창근, 양기철

목포대학교 멀티미디어공학과

Park Chang-Keun, Yang Gi-Chul

Dept. of Multimedia Engineering, Mokpo University

### 요약

현재 대부분의 정보검색시스템에서 사용되고 있는 키워드 매칭기법은 기하급수적으로 늘어나는 정보를 효과적으로 처리하기에는 부적합하다. 이러한 문제는 의미정보를 활용하여 해결할 수 있으며, 본 논문에서는 의미정보를 효과적으로 활용할 수 있는 한 방법을 제시한다. 본 논문에서 제안한 기법은 의미정보를 개념그래프를 이용하여 표현하고, 이를 정보검색에 활용한다. 구현된 시스템은 완전 매칭과 부분 매칭이 가능한 시스템이다. 부분 매칭은 구문적 부분 매칭과 의미적 부분 매칭 두 가지가 있다. 의미적 유사도는 온톨로지 내의 서브클래스 관계로 계산된다. 이러한 기법은 정보검색 뿐만 아니라 동적 하이퍼링크의 구현 등 다양한 분야에서 적용될 수 있다.

### Abstract

Keyword matching technique which is used in most information retrieval systems is unfit for efficient processing of geometrically increasing information. The problem can be solved by using semantic information and an efficient method of semantic processing is introduced in this paper. The technique uses conceptual graph to represent the semantic information and apply it for information retrieval. The implemented system can perform exact matching and partial matching. Partial matching has two different types. One is syntactic partial matching and the other is semantic partial matching. The semantic similarities are measured by the subclass relations in the ontology. The introduced technique can be used not only information retrieval but also in various applications such as an implementation of dynamic hyperlinks.

## I. 서론

과거에는 정보의 획득을 위해서 주로 서적을 이용하였다. 최근에는 컴퓨터와 인터넷이 보급됨에 따라 정보의 교환 및 습득에 큰 영향을 미쳤고, 인터넷의 출현과 함께 온라인 정보가 우리의 생활 속에 깊숙이 자리잡게 되었다. 이는 대량의 정보를 생산하게 하였

으며, 정보의 획득을 위해 검색시스템을 사용 하여야 하게 되었다. 기존의 검색시스템은 정보 검색 방법 중 하나로 키워드 매칭 검색방법을 이용하였다. 키워드 매칭은 동일한 키워드를 가지지만 의미가 다른 두 개 이상의 문장을 모두 검색해 내는 문제점을 가지고 있으며, 이는 정확하고 효율적인 검색 결과를 제공해주지 못하고 있다. 본 논문에서는 키워드 매칭이 아

닌, 의미정보를 이용한 검색방법을 제안하며, 이는 의미정보를 이용하므로써, 정보 검색에 있어서 효율적이고, 보다 정확한 검색 결과를 제공 할 수 있도록 한다. 의미정보의 표현은 개념그래프를 이용하고, 사이트의 내용을 설명해 주는 개념그래프와 URL을 데이터베이스에 저장하게 되며, 개념그래프의 구문을 통해 온톨로지 내의 인스턴스를 검색하여 의미적 상하관계를 검색을 하게 된다. 구현하고자 하는 검색 시스템은 완전매칭과 부분매칭의 기능을 제공하게 된다. 완전매칭의 경우에는 정확한 그래프를 검색하게 되며, 부분매칭의 경우에는 어느 정도 비슷한 것까지를 매칭해 내느냐 하는 것을 온톨로지에서 상하관계를 적용하여 의미적 검색 범위를 결정하도록 한다. 본 논문의 구성은 다음과 같다. II장에서는 관련연구방법을 살펴보고, III장에서는 시스템의 설계와 구현에 대해 설명하고, IV장에서는 결론을 맺으며, 향후 연구과제에 대해 소개한다.

## II. 관련연구

### 1. 개념그래프

J. F. Sowa가 제안한 개념그래프는 인공지능의 여러 의미망을 통합한 지식표현 언어이며, 개념 도식을 이용해 인간이 쉽게 읽을 수 있도록 논리적으로 간결하면서도 풍부한 표현력이 있으며, 컴퓨터에 의한 자연언어 처리 등에서도 쉽게 이용할 수 있다. 개념그래프는 노드들이 아크로 연결된 유한 이분 그래프로, 개념을 표현하는 개념노드와 개념노드 간의 관계를 표현하는 관계노드가 된다. 추상적 또는 사실적 실체를 나타내는 개념노드는 그래프 상에서 사각형으로 표현되는데, 표기의 간결성을 위하여 그래프를 선형 표현으로 표현할 때는 대괄호 '[' ]'를 사용한다. 하나의 개념노드는 콜론 ':'에 의하여 두 부분으로 나뉘어 있는데 이때 콜론의 왼쪽 부분은 그 개념노드의 유형을 나타내고, 콜론의 오른쪽 부분에는 그 노드에 제약을 가하는 정보가 들어간다. 관계노드는 원으로

표시하나, 선형표현에서는 소괄호 '(' )'를 쓴다. 그림 1은 하나의 개념그래프의 예를 보인 것이다.

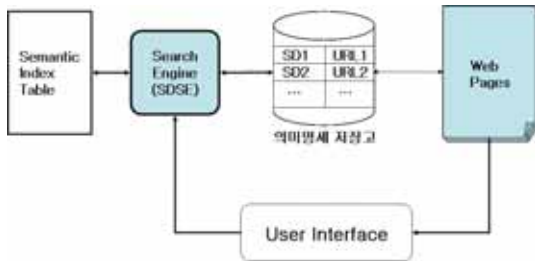
[ EAT ] → ( AGNT ) → [ PERSON : John ]

▶▶ 그림 1. 개념 그래프

## III. 의미정보기반 검색 시스템

정보검색(Information Retrieval) 시스템은, 정보추출(Information Extraction) 시스템 등과 관련이 있는 시스템이지만 이들과는 차이점이 있다. 정보검색 시스템은 많은 양의 정보와 여러 분야의 정보를 대상으로 하고 대부분 문서전체를 검색결과로 제시한다. 여기에는 SMART나 SPIDER 같은 시스템이 있다. 하지만 이들은 문서전체 또는 많은 분량의 텍스트를 사용자에게 제시하여 사용자에게 크게 도움을 주지 못할 때가 많다. 이를 개선하기 위해 중요한 단어가 많이 들어있는 문장을 찾아주기도 하지만 여기에도 문제점들이 많다. 중요단어를 사용할 때 기능어는 빼고 내용어만을 사용하는 문제도 있고 의미정보를 이용하지 못하는 대서오는 여러 가지 문제들이 있다. 따라서 필요 없는 부분이 검색되기도 한다. 본 논문에서는 정보검색을 수행할 때 원하는 정보만을 신속히 검색할 수 있는 시스템을 개발하고 이에 관련된 이론을 정립한다. 즉, 문서의 내용을 검색에 활용하여 사용자에게 관련이 없는 문서는 검색되지 않고 사용자가 필요로 하는 정보만을 검색할 수 있는 검색 시스템을 개발 하고자 하는 것이다. 따라서, 개발되는 검색시스템은 단어나 스트링이 아닌, 내용에 의한 검색이 가능한 시스템이 되도록 한다. 이를 위하여 기계가 이해할 수 있는 형태의 의미명세를 문서 작성자가 작성하여 일종의 헤더 형식의 간단한 메타 정보로 이용하여 파일(문서)의 의미정보를 검색에 활용 할 수 있게 한다. 특히, 검색시스템은 메타 정보의 완전매칭(Exact Matching)과 부분매칭(Partial Matching)

모두를 수행할 수 있도록 한다. 의미명세(Semantic Description)에서 개념간 의미적 관계는 유형계층 (Type Hierarchy)상의 관계를 이용하며 사용자가 의미적 검색 범위(Semantic Range of Search)를 결정하도록 한다. 즉, 부분매칭인 경우 어느 정도 비슷한 것까지를 매칭해 내느냐 하는 것을 사용자가 결정할 수 있도록 하는 것이다. 이는 유형계층(Ontology)에서 상하관계만을 적용하여 숫자로 의미적 검색 범위를 간단히 표현할 수 있게 하면 된다. 또한 검색속도나 신뢰도를 증가시키기 위하여 상용의 데이터베이스를 이용하여 시스템을 구현한다. 시스템은 크게 두 부분으로 나누어진다. 첫째는 다수의 의미명세를 효율적으로 저장할 수 있는 저장시스템(Storage System)이고 둘째는 저장된 의미명세를 신속히 검색해 낼 수 있는 검색시스템(Retrieval System)이다. 저장시스템은 기억장치의 자리를 조금 사용하면서 많은 의미명세를 저장할 수 있도록 하고 검색 시스템은 의미적 매칭을 효율적으로 수행할 수 있도록 한다. 시스템의 전체적인 구조 [그림2]는 다음과 같다.



▶▶ 그림 2. 시스템의 전체적인 구조

검색엔진은 사용자 인터페이스, Conceptual Graph, Acceptor, Ontology Manager, SIT, SDDB, 검색기(Semantic Matcher) 등으로 구성된다. CG Acceptor를 통해, 들어온 의미명세는 SDDB에 저장되고 인덱스화 되어 SIT에 저장된다. Ontology Manager는 유형계층을 생성할 때 사용하는 도구로 유형계층의 생성은 물론이고, 그 변경이나 검색 등 유지보수에 사용할 수 있다. 검색기는 앞서도 설명

한 바와 같이 의미매칭을 수행하는데 완전매칭뿐만 아니라 부분매칭도 할 수 있도록 한다. 검색시스템의 인터페이스 [그림3]은 다음과 같다.

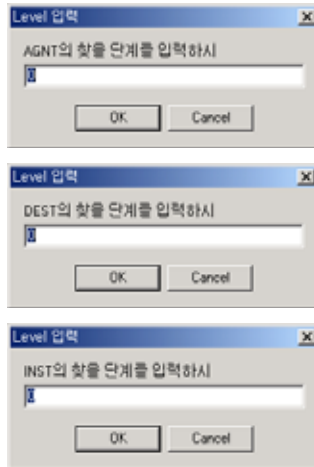


▶▶ 그림 3. 검색 인터페이스

위 검색 인터페이스에서 Exact Matching에 경우, 검색할 내용을 입력하고, Search Start 버튼을 누르면, 검색된 문장 번호와 현재 선택된 문장 번호에 맞는 개념 그래프가 출력된다. Partial Matching을 선택시, 매칭 비율부분이 활성화되고, 이곳에 찾을 문자의 비율을 선택하여 입력하고, 검색을 시작하면, 검색된 문장번호와 개념그래프가 출력된다. 찾는 개념그래프의 노드 수가 4개 일 경우에 매칭비율이 25%이 하이면 하나의 노드만 같아도 같은 문장으로 찾아진다. 그리고 매칭 비율이 50%이상이면 두 개의 노드가 같을 경우에 같은 문장으로 간주하여 찾아준다. 매칭 비율이 75% 이상이면 3개의 노드가, 매칭 비율이 100%이면 4개의 노드가 같은 문장을 찾는다. 이때 찾아지는 문장은 노드의 수와는 무관하다.

Semantic Matching의 경우에 Syntatic Matching과 다른점은 개념의 계층 구조에서 찾을 개념들의 상위 개념의 단계를 입력받아서 그 단계에 해당하는 개념의 하위개념들이 포함하고 있는 개념그래프도 찾아 준다는 것이다.(AGNT [GO][PERSON:tom])(DEST [GO][CITY:Boston])(INST [GO][BUS]) 이런 문장을 찾을 경우 다음과

같은 세 개의 단계를 묻는 대화상자가 나타난다.



▶▶ 그림 4. 대화상자

AGNT 부분인 tom을 1단계 올리면 개념의 계층구조에서 MAN과 TEACHER가 나오는데 MAN이나 TEACHER의 하위개념인 STEVE가 들어간 문장인 (AGNT [GO][PERSON:steve])(DEST [GO][CITY:Boston]) (INST [GO][BUS])도 찾아 준다.

#### IV. 결론 및 향후 연구과제

본 논문에서는 단어나 스트링이 아닌 내용(i.e., Semantic Description)에 의한 검색이 가능한 의미정보 기반 정보검색 시스템을 설계하고 구현하였다. 이를 위하여 기계가 이해할 수 있는 형태의 의미명세를 문서 작성자가 작성하여 일종의 헤더 형식의 간단한 메타 정보로 이용하여 의미정보를 검색에 활용 할 수 있게 한다. 특히, 검색시스템은 메타 정보의 완전매치(Exact Matching)과 부분매칭(Partial Matching) 모두를 수행할 수 있도록 하였다. 소개한 검색 기법은 기존의 Semantic Web과는 달리 웹 문서 전체에 의미 마크업을 하지 않고도 의미명세를 이용하여 문서의 의미를 검색에 활용할 수 있고, 검색 절차는 기존의 웹 검색엔진을 사용하는 시스템보다 간단하여 사

용자가 쉽게 사용할 수 있다.

#### ■ 참고문헌 ■

- [1] 오행언, 양기철, 의미 정보를 이용한 문헌 검색, 한국정보처리학회추계학술표논문집, 1997.
- [2] 양기철, 개념그래프의 소개, 정보과학회지 12-9, pp 72-79, 1994.10.
- [3] John F. Sowa, Conceptual Structures: Information Processing in Mind and Machine, Addison-Wesley, Massachusetts, 1994.
- [4] Poole, J., J.A. Campbell. A novel algorithm for matching conceptual and related graphs, Conceptual structures : applications, implementation and theory, Eds. G. Ellis et al., New York, (LNAI954), 1995.
- [5] Yang, G-C., Y. Choi, & J. Oh. CGMA : A novel conceptual graph matching algorithm, Conceptual structures : theory and implementation, Eds. H.D.Pfeiffer, T.E.Nagle, New York,(LNAI754), 1993.