
도메인 온톨로지를 이용한 검색 시스템 설계 및 구현

강래구* · 정채영**

Design and Implementation of Search System Using Domain Ontology

Rae-Goo Kang* · Chai-Yeoung Jung**

본 연구는 문화관광부 및 한국문화콘텐츠진흥원의 문화콘텐츠기술연구소(CT)육성사업의 연구결과로 수행되었음.

요 약

지금까지의 상품 검색 방법으로는 찾고자하는 정보를 검색할 때 주로 단어의 빈도수나 어휘 정보를 이용하는 키워드 기반의 검색이 주로 쓰이고 있었다. 키워드 기반의 검색에서는 사용자의 질의와 관련이 없는 문서들까지도 같은 결과로 나타내 주고 이로 인해 사용자는 제시된 결과를 한번 더 수동적으로 검색해야하는 부담을 앓게 되었다. 이러한 문제점을 해결하기 위해 온톨로지가 대두되었다. 본 논문에서는 온톨로지를 이용한 상품 검색 시스템을 직접 구축하여 분류별 검색을 통해 얼마나 정확한 검색을 하는지 실험하였다. 실험을 위해 전국적으로 On/Off라인 할인점을 운영 중에 있는 A할인점의 상품 데이터 약 40,000여개를 데이터베이스로 구축하였고 User Interface 개발환경은 JSP와 PowerBuilder9.0을 사용하여 검색 시스템을 개발하여 실험하였다. 그 결과 본 논문에서 제안하고 설계한 상품 도메인 온톨로지를 이용한 검색 방법이 기존의 키워드 기반의 검색 방법보다 우수한 결과를 나타내고 있음을 입증하였다.

ABSTRACT

TSP(Traveling Salesman Problem) is a problem finding out the shortest distance out of many courses where given cities of the number of N, one starts a certain city and turns back to a starting city, visiting every city only once. As the number of cities having visited increases, the calculation rate increases geometrically. This problem makes TSP classified in NP-Hard Problem and genetic algorithm is used representatively. To obtain a better result in TSP, various operators have been developed and studied. This paper suggests new method of population initialization and of sequential transformation, and then proves the improvement of capability by comparing them with existing methods.

키워드

Ontology, 지능형 검색, 분류별 검색, 상품 도메인

* 조선대학교 전산통계학과
** 교신저자, 조선대학교 전산통계학과

I. 서론

1989년 Tim Berners-Lee에 의해 처음 제안된 월드와이드웹은 HTML을 이용하여 누구나 쉽게 하이퍼텍스트 정보에 접근할 수 있게 편리성에 중점을 둠으로써 결과적으로 정보의 증가를 가져왔다. 이처럼 사용하기 편리하다는 점이 웹의 주요 성장 요인이기도 하지만 반면에 많은 양의 정보가 폭발적으로 생성되는 현 상황에서 정보의 저장과 검색에 있어서 많은 문제점이 생겨나고 있다. 현재의 웹 사이트에서의 검색은 사용자가 자주 사용하는 용어와는 관계없이 웹 사이트 개발자나 데이터베이스 설계자가 일방적으로 정해 놓은 키워드를 통해서만 정보를 검색할 수 있다는 문제점을 가지고 있다. 이러한 문제점을 극복하기 위해서 1999년 W3C (World Wide Web Consortium)에서는 시맨틱 웹(Semantic Web)을 제안하게 되었고 그 중심에 온톨로지(Ontology)가 있다.[1]

지금까지의 검색은 주로 단어의 사용 빈도수나 어휘 정보를 이용하여 문서의 유사도를 측정하고 순위를 부여하기 때문에 사용자가 찾고자 하는 정보와 관계가 없는 동형어의어나 동음이의어의어 같은 단어가 포함된 문서들까지도 동일한 결과로 나타내었고 이로 인해 사용자는 제시된 결과를 한번 더 검색을 해야 하는 부담을 안게 되었다. 반면 온톨로지를 이용하여 정보를 검색할 경우엔 사용자가 찾고자 하는 정보를 보다 빠르게 찾을 수 있고 또한 검색된 결과의 정확도를 향상시킬 수 있다는 장점이 있다.[2]

II. 온톨로지(Ontology)

온톨로지는 넓은 의미로는 데이터베이스라고 할 수 있지만 데이터베이스보다 복잡한 형태의 지식과 관련되어 있다는 의미에서 지식베이스라고 부르기도 한다. 지식베이스 구축을 위한 하나의 방안으로 활발히 연구되고 있는 온톨로지는 정보검색은 물론 추론을 지원함으로써 지식을 수집하고 표현하는 유용한 수단으로 이용되고 있다. 온톨로지는 원래 철학 용어이나 지금은 철학 이외의 여러 다른 분야에서도 널리 사용되고 있으며 인공지능 분야에서의 온톨로지는 세상의 어떤 부분을 모델링하여 표현할 수 있는 용어의 집합으로 정의하고 있다.[1][3]

또한 온톨로지는 특정분야의 용어 그리고 용어의 관계정의는 물론 용어의 조합규칙과 용어의 확장에 대한 관계도 정의한다. 즉, 온톨로지는 영역 지식 표현을 위한 형식 체계인 것이다.[4][5][6]

2.1 온톨로지 언어의 개요

웹에서는 서로 다른 사전 체계를 가진 문서들 간에도 변환이 쉽게 이루어지도록 개념 간의 계층관계나 개념 정의 간의 정합성 등을 자동으로 계산할 수 있는 온톨로지 언어를 제공하고 있다. 웹 환경에서 효과적으로 지식을 표현하기 위해 W3C에서는 다양한 언어를 개발하여 사용하고 있으며, 2004년 2월에는 개정된 자원 프레임워크(RDF)와 웹 온톨로지 언어(OWL)를 웹 기술 언어로 인정한다고 발표했다.

RDF(Resource Description Framework)는 XML의 문제점을 해결하고 의미에 초점을 맞추기 위해 제시된 기반 구조로서 W3C에서 발표한 메타데이터를 기술하는 포맷으로 웹의 리소스에 관한 메타데이터의 기술과 교환을 위한 표준으로 사용되고 있다.[7]

OWL은 문서에 포함된 정보를 어플리케이션을 이용하여 자동으로 처리하고자 할 때 사용하는 언어이다. OWL을 이용하면 임의의 어휘를 구성하는 용어의 의미와 용어들 간의 관계를 명시적으로 표현할 수 있다. 이처럼 용어와 용어들 간의 관계를 표현한 것을 온톨로지(Ontology)라 한다. [8]

III. 온톨로지의 설계

대부분의 인터넷 쇼핑몰은 상품 검색시 주로 키워드를 기반으로 검색을 함으로써 사용자가 입력한 검색어가 데이터베이스의 상품 목록과 일치해야만 정확한 검색이 이루어지며 또한 유사한 상품을 검색함에 있어서 키워드가 서로 다를 경우엔 전혀 검색할 수 없는 단점이 있었다. 이는 기존 웹에서 주로 사용되고 있는 HTML의 의미 정보를 제공하지 못하기 때문이다.[9]

본 논문에서는 웹상에서 보다 정확한 상품 검색을 위해 상품 도메인 온톨로지를 이용하여 사용자가 입력한 질의를 의미적인 해석을 통해 원하는 결과를 보다 쉽고 정확하게 검색할 수 있도록 하였다.

3.1 온톨로지의 개념 설계

개념 설계는 구축하고자 하는 온톨로지에 필요한 클래스를 결정하는 것으로 시작한다. 하지만 클래스를 결정하고 설계한다는 것은 매우 어렵고 불명확하다. 이 과정은 온톨로지 개발자와 관련 분야 전문가의 경험과 직관에 강하게 의존하게 된다.

클래스의 이름을 정의하는 것은 온톨로지를 구축하는데 있어서 매우 중요하다. 만약 어떠한 클래스의 이름이 정확하지 않다면 그에 따른 속성이나 다른 클래스와의 관계를 명확하게 정의하는 것은 불가능할 것이다. 일반적으로 클래스의 이름을 정의하고 계층 구조를 정의하는 데 있어서 다음과 같이 몇 가지 고려할 규칙이 있다.

- 클래스 이름을 명확하게 정의한다.
- 클래스 이름에 대한 다국어 표현을 고려한다.
- 유사어는 통합하여 하나의 클래스로 정의한다.
- 형제 관계는 같은 단계로 정의한다.
- 하나의 클래스에서 서브 클래스가 12개 이상일 경우엔 새로운 부모 클래스를 만든다.
- 다중 상속을 고려한다.

위의 규칙을 고려하여 클래스의 이름을 정의하여야 한다. 클래스는 다음과 같이 세 단계 트리구조로 분류할 수 있다. [10]

표 1. 온톨로지 클래스 사전의 일부
Table 1. A part of ontology class dictionary

클래스	상위 클래스	하위 클래스
유음료	Daily식품	우유, 두유, 유산균음료
디저트	Daily식품	호상, 젤리, 푸딩, 무스
유제품	Daily식품	버터, 마아가린, 치즈
국산주류	주류	소주, 맥주, 전통주, 탁주
수입주류	주류	위스키, 브랜디, 와인

표 1은 본 논문에서 구축한 온톨로지 클래스 사전의 일부를 나타내고 있다.

- 최상위 클래스 : 부모를 갖지 않는 가장상위에 존재하는 클래스
- 핵심 클래스 : 최상위 클래스 속성을 물려 받은 자

식 클래스

- 행위 클래스 : 실제 세계에서 지속적으로 데이터가 생성되는 내용을 갖는 클래스

3.2 온톨로지의 구현

온톨로지를 구축함에 있어서 최상위 클래스는 부모를 갖고 있지 않기 때문에 제일 상위에 위치하게 된다. 핵심 클래스란 부모를 가지고 있는 클래스로서 각 분류의 조상 클래스가 된다. 보편적으로 온톨로지를 구축함에 있어서 핵심 클래스는 2단계 혹은 3단계 수준에 위치하게 된다. 마지막으로 행위 클래스는 트리구조에서 가장 하위에 위치해 있고 자손으로 비유된다. 최상위 클래스와 핵심 클래스를 제외한 모든 것을 행위 클래스로 간주한다.[11]

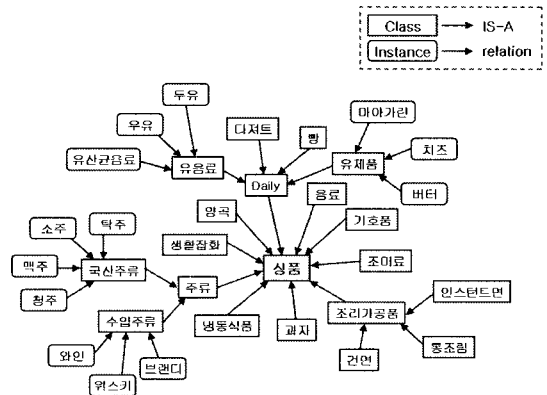


그림 1. 온톨로지 클래스의 개념과 관계 일부
Fig. 1. A part of notions and relation of ontology class

그림 1은 본 논문에서 구축한 상품 도메인 온톨로지 클래스의 개념과 관계를 축약하여 보여주고 있다.

이와 같은 개념과 관계를 이용하여 상품 도메인 온톨로지를 설계 하였고 온톨로지에 존재하는 개념과 그들의 관계는 Protégé 3.1을 이용하여 OWL로 표현하였다.

3.3 유의어 처리기능

상품 검색에 있어서 “가격” 이란 단어와 “판매 금액”이라는 단어는 서로 다른 단어이지만 하나의 뜻으로 쓰이는 유의어이다. 이러한 유의어를 사용자가 검색 조건으로 사용할 경우 같은 의미로 판단할 수 있도록 온톨로

지에 유의어 목록이 정의되어 있어야 한다. 사용자의 단어 선택 성향에 따라 같은 의미의 단어일지라도 서로 다른 단어를 선택하여 검색하게 되는 경우가 많이 있다. 이러한 경우 검색어가 서로 다르다 할지라도 검색된 결과는 항상 같은 결과가 나올 수 있도록 유의어 목록이 정의되어 있어야 한다.

IV. 시스템 구현 및 평가

지금까지의 웹에서 검색 방법으로는 주로 키워드 검색이 이용되어 왔다. 하지만 차세대 웹으로서 시맨틱 웹이 대두되기 시작하면서 온톨로지를 활용한 의미적인 검색이 가능하게 되었다.

이 장에서는 본 논문에서 구축한 상품 도메인 온톨로지를 이용하여 상품 검색을 했을 경우 얼마나 정확하게 원하는 결과를 검색해 내는지 실험하기 위해 직접 시스템을 구축하여 실험하였다.

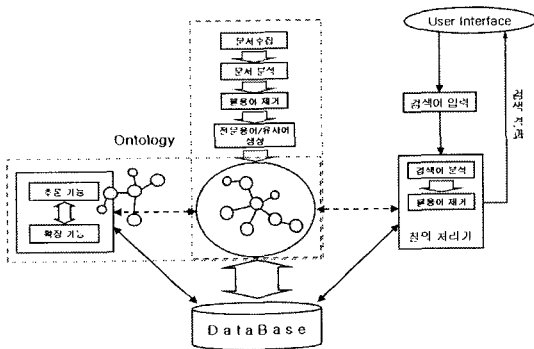


그림 2. 온톨로지를 이용한 검색 시스템의 구조
Fig. 2. Structure of ontology-based product search system

본 논문에서 구축하고자 하는 검색 시스템의 구조는 그림 2와 같다. 사용자 인터페이스(UI : User Interface)는 일반적으로 사용하고 있는 GUI 환경에서 사용자가 원하는 검색어를 입력하고 검색 결과를 통보 받을 수 있도록 시스템을 설계하였다.

실험을 위해 사용할 상품 데이터는 현재 전국적으로 온·오프라인 할인점을 운영 중에 있는 A할인점의 상품 데이터 약 40,000여개의 품목을 Oracle 9i로 구축하여 사

용하였다.

그림 3은 본 논문에서 상품 검색을 위해 사용하게 될 사용자 인터페이스 화면을 보여주고 있다.

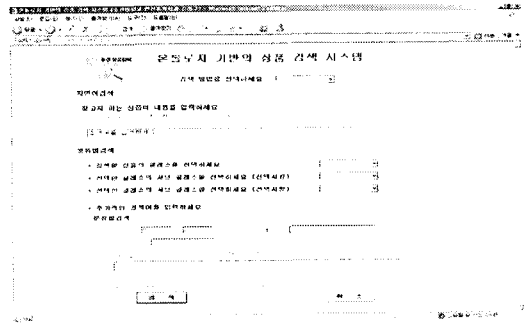


그림 3. 사용자 인터페이스(UI) 화면
Fig. 3. User interface

사용자 인터페이스는 JSP와 PowerBuilder9.0을 이용하여 웹상에서 사용 가능하도록 직접 구현하였다. 검색 방법은 검색하고자 하는 상품의 대·중·소분류를 선택하여 선택된 분류를 검색어로 사용하는 방법이다. 선택된 검색 조건을 온톨로지를 이용해 의미적인 내용을 부여함으로써 원하는 상품을 데이터베이스에서 정확하게 검색할 수 있으며 검색 조건의 추론 및 확장을 통해 유사 상품이나 비슷한 분류의 상품까지도 검색할 수 있다.

5.1 분류별 검색 실험 및 평가

찾고자 하는 상품에 대한 기본적인 지식이 있을 경우 사용하는 검색 방법으로서 상품이 해당하는 분류를 선택하고 추가적인 검색 내용을 입력하면 입력된 값들이 파라미터로 전달되고 전달된 값을 가지고 데이터베이스에서 상품을 찾게 된다.

그림 4는 분류별 검색 방법을 나타내고 있다.

①과 ②에 해당하는 부분이 온톨로지로 구축된 부분으로서 사용자가 쉽게 사용할 수 있도록 재구성하여 사용자 인터페이스 화면에 디자인 하였다.

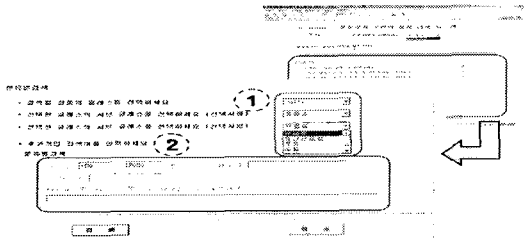


그림 4. 분류별 검색 방법
Fig. 4. Search method

①의 경우 각각의 상품 정보를 분석하여 구축한 온톨로지를 사용자로 하여금 검색하고자 하는 상품의 대·중·소분류로 활용할 수 있도록 클래스를 도식화 하였다. 즉, 검색을 원하는 상품에 대한 기본적인 지식을 알고 있을 경우 찾고자 하는 상품의 대·중·소분류를 메뉴에서 선택하게 되면 선택된 각각의 분류와 구축된 온톨로지의 상·하위 클래스를 1:1로 매칭을 시켜 내부적으로는 사용자가 클래스를 직접 선택하여 상품을 검색하는 것과 같은 개념이 된다.

②의 경우는 좀 더 구체적이고 세분화 된 검색을 하고자 할 때 추가적으로 입력이 가능한 부분으로서 주로 유의어 목록과 추론기능을 이용하게 된다. 가격, 제조사, 제품 용량, 수량뿐 만 아니라 각 상품의 특징을 추가적으로 입력함으로써 상품 검색을 보다 상세하게 할 수 있도록 해 준다.

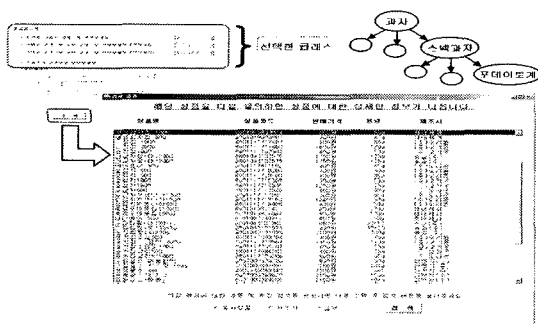


그림 5. 분류별 검색 결과 1
Fig. 5. Search result 1

그림 5는 분류별 검색 결과를 보여주고 있다. 그림 5에서 보면 사용자는 검색하고자 하는 상품의 대·중·소분류를 “과자”->“스넥과자”->“포테이토계” 순으로 선택하였고, 이렇게 선택된 분류를 가지고 상품 도메인

온톨로지를 이용하여 상품을 검색하게 된다. 이 때 선택된 대·중·소분류와 온톨로지의 상·하위 클래스는 1:1로 서로 매칭되어 검색이 진행된다. 이러한 과정을 통해 검색된 상품들은 모두 “과자” 클래스의 하위 클래스로서 그 중 “스넥과자” 클래스에 속하고 “스넥과자” 클래스의 하위 클래스인 “포테이토계” 클래스에 속하는 상품들이다.

이처럼 내부적으로 구축되어 있는 온톨로지의 상·하위 클래스 계층을 사용자가 인터페이스 화면에서 대·중·소분류를 이용하여 직접 선택할 수 있도록 디자인함으로써 사용자로 하여금 찾고자 하는 상품이 속해 있는 온톨로지 클래스를 직접 선택할 수 있는 효과를 주는 것이다.

그림 6은 그림 5의 검색 결과에 추가적인 검색어 입력을 통해 좀 더 세부적인 상품 검색을 실행한 결과이다. 검색 조건에 나타나 있듯이 가격대와 제조사를 추가 검색어로 입력하여 그림 6에 나타난 결과 중 가격대가 “1000원~2500원” 사이에 해당하는 상품들 중 제조사가 “P&G”에 해당하는 상품들만을 검색하도록 하였다. 이 때 기존의 키워드 검색에서는 추가적으로 입력된 검색어 중 가격과 관련된 검색어는 특별한 문제가 없지만 제조사와 관련된 검색어에서는 문제가 발생하게 된다.

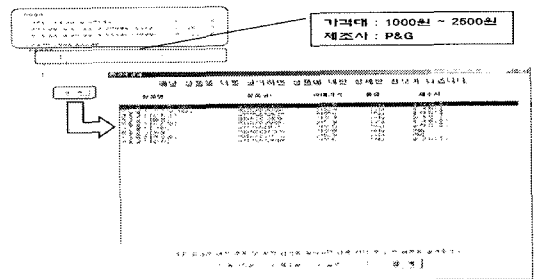


그림 6. 분류별 검색 결과 2
Fig. 6. Search result 2

키워드 검색에서는 검색어로 입력한 제조사의 영문 명칭인 “P&G”와 한글 명칭인 “피앤지(주)”를 같은 의미로 판단하지 못한다. 그렇기 때문에 “P&G”와 “피앤지(주)”가 동일한 제조사임에도 불구하고 데이터베이스에 저장된 상품의 제조사 키워드가 어떠한 명칭으로 저장되어 있는가에 따라 서로 다른 검색 결과를 나타내게 된다.

표 2. 온톨로지에 정의된 제조사 유의어 목록의 일부
Table 2. A part of synonym list defined in ontology

대표어	유의어
P&G	피앤지, 피앤지(주), 피엔지, 피엔지(주)
Unilever	유니레버, 유니레버코리아, 유니레버, 유니레버코리아
오리온	동양제과, (주)오리온
매일유업	매일, 매일우유

하지만 그림 6에서처럼 온톨로지를 이용한 상품 검색에서는 표 3과 같이 온톨로지에 구축되어 있는 제조사 관련 유의어 목록을 통해 “P&G”와 “피앤지(주)”가 같은 의미로 사용되고 있음을 추론하여 검색하기 때문에 둘 중 어떠한 명칭을 사용하더라도 항상 동일한 검색 결과를 나타내게 된다.

이는 이미 각 상품들의 특징과 자주 쓰이는 단어 및 유의어 등을 분석하여 구축해놓은 온톨로지를 통해 검색어의 의미적인 부분까지 파악하여 검색하기 때문에 가능한 일이다.

V. 결론 및 향후 과제

본 논문에서는 온톨로지를 상품 검색에 활용하기 위해 상품 도메인 온톨로지를 직접 구축하여 분류별 검색을 통해 얼마나 정확한 검색을 하게 되는지 실험하였다.

이 실험을 위해 Protégé 3.1을 이용하여 OWL로 온톨로지를 구축하였고, JSP와 PowerBuilder9.0을 이용하여 사용자 인터페이스를 만들어 상품 검색을 하였다. 그 결과, 온톨로지에 구축된 클래스 개념과 속성, 유의어, 추론 등 다양한 온톨로지 기능을 통해 사용자가 원하는 상품을 정확히 찾아내는 걸 확인하였고, 사용자가 예상치 못한 상품에 대해서도 추론을 통해 상품 범위를 확장시켜 검색을 하였다.

이처럼 온톨로지를 이용하여 검색할 경우 개념과 추론 등을 통해 스스로 검색어 자체에 의미를 부여함으로써 사용자가 원하는 상품을 보다 정확하게 검색할 수 있고, 사용자가 예상하지 못한 상품에 대해서도 검색이 가능하다는 장점이 있다. 매년 새로운 신상품이 다량으로 출시되고 있는 현 시점에서 초기에 구축해 놓은 온톨로

지를 영구적으로 이용한다는 것은 불가능한 일이다. 정확한 상품 검색을 처리하기 위해서는 신상품과 새로운 분류의 상품들에 맞게 지속적으로 온톨로지를 수정 및 확장해 나가야 할 것이며, 상품 데이터베이스 설계에 있어서도 보다 간단하고 명확한 설계가 이루어져야 할 것이다.

참고문헌

- [1] 김홍기, 김학래, 이강찬, 정지훈, 이재호 외, “월드 와이드 웹에서 시맨틱 웹으로,” 『마이크로소프트 웨어』, pp.242-301, 2002.
- [2] Guarino, N., “Formal Ontology and Information Systems,” Proceedings of FOIS’98, pp.3-15, 1998.
- [3] Kuhanandha Mahalingam and Michael N. Huhns “An Ontology Tool for Query Formulation in an Agent-Based Context”, 1997.
- [4] Robert Neches et. al. “Enabling Technology for Knowledge Sharing” AI Magazine, Winter, 1991.
- [5] Rudi Studer, V. Richard Benjamins, and Dieter Fensel. “Knowledge Engineering : Principles and Methods.” Data & Knowledge Engineering, Vol.2 5, No.1-2, pp.184-185, 1998.
- [6] 김홍기, 김학래, “생명이 없는 공간과 의사소통을 : 온톨로지”, 『(주)마이크로소프트웨어』, 2002.
- [7] Stefan Decker, Dan Brickley, Janne Saarela, and Jurgen Angele, A Query and Inference Service for RDF, The Query Languages Workshop(QL’98), W3C, 1998.
- [8] WD-RDF-Schema Website, <http://www.w3.org/TR/2004/REC-owl-features-20040210/>
- [9] 김영민, 이상춘, “시맨틱을 이용한 연구논문 검색 시스템”, 『인터넷 정보학회 논문지』 Vol.4 No.3, pp.15-22, 2003.
- [10] 한국전산원, “웹 온톨로지 개발 지침 연구”, 2004.
- [11] <http://protege.stanford.edu>.

저자소개



강 래 구(Rae-Goo Kang)

2007년 조선대학교
전산통계학과석사졸업

2007년 - 조선대학교
전산통계학과박사과정

※관심분야: 신경망,인공지능,정보보호,Bioinformatics



정 채 영(Chai-Yeoung Jung)

1989년 조선대학교 컴퓨터공학과
공학박사

1986년 - 조선대학교
컴퓨터통계학과 교수

※관심분야: 신경망,인공지능,정보보호,Bioinformatics