

# 지식 관리를 위한 온톨로지 맵 개발†

## Ontology Map Development for Knowledge Management

신광섭<sup>1</sup>, 정재윤<sup>1</sup>, 이우기<sup>2</sup>, 강석호<sup>1</sup>

<sup>1</sup>서울대학교 산업공학과, <sup>2</sup>성결대학교 컴퓨터공학부

### Abstract

최근 산업경쟁이 심화되면서 지식은 기업 가치에서 점점 더 많은 비중을 차지하고 있다. 이러한 지식을 기업의 자산으로서 체계적으로 관리하고 확산시키기 위하여 지식관리 시스템이 등장하였다. 본 논문은 문서 검색을 이용하는 지식관리 시스템의 설계를 대상으로 한다. 특히, 지식관리 시스템의 효과적인 문서 검색을 위하여 온톨로지 맵을 구성하고, 지식 검색에 이용하고자 한다. 온톨로지 맵의 구축과 수정 및 확장 방안을 제안하고, 질의에 대해 적합한 문서를 추출하기 위하여 질의어와 온톨로지 맵 간, 온톨로지 맵과 문서 간의 유사성을 측정하고 이를 결합하는 방법을 설명한다. 이러한 온톨로지 맵을 이용한 지식관리에 관한 연구는 지식의 체계적인 유지 및 확장성뿐 아니라, 지식 검색의 정확성과 유연성을 배가할 수 있을 것이다.

### 1. 서론

기업 활동이 세계화되고 경쟁이 심화되면서, 무형자산에 대한 관심과 투자가 늘어가고 있다. 특히 기업은 타 기업에 대한 경쟁우위를 점하기 위하여 자신들이 보유하고 있는 지식을 업무에 적극적으로 활용하기 위한 방안을 모색하게 되었다. 이제 지식은 단순히 조직이 보유하고 있는 자산의 일부가 아니라, 가장 큰 경쟁우위로 인식되고 있으며, 기업 자산 평가에 있어서도 다른 어떠한 물리적 자산보다 높은 중요도를 가진다[15]. 따라서 기업이나 연구소, 정부 등을 막론하고 그들의 보유 지식을 저장하고 공유하는 것은 가장 중요한 이슈가 되었다. 이러한 관점에서 조직 내 저장되어 있는 데이터와 정보를 지식의 형태로 전환하고, 이를 소프트웨어 혹은 데이터베이스에 저장하거나 효율적으로 관리할 수 있는 지식관리 시스템(KMS; Knowledge Management System)의 필요성이 부각되었다[9, 12].

지식은 크게 암묵지(tacit knowledge)와 형식지(explicit knowledge)의 두 가지 형태로 나눌 수 있다[13]. 암묵지는 전문가의 경험과 같이 어떠한 형태로든 표현하기 힘든 것을 의미하며, 형식지는 부호화된 지식(codified knowledge)이라고도 불리며, 일반적인 정보 시스템에 적합한 형태의 언어로 표현하기 쉬운 형태의 지식을 의미한다[14]. 형식지로 저장하기 위하여 문서, 웹 페이지, 이미지, 음성 데이터 등과 같은 여러 가지 매체가 사용될 수 있지만, 현재 대부분의 조직에서는 작성, 보관 및 공유의

편의성, 그리고 저렴한 비용으로 인하여 지식을 저장하는 매체로서 문서를 가장 많이 사용하고 있다. 따라서 본 논문에서는 문서를 기본적인 지식 저장 매체로 한정하고, 이를 체계적이고 효과적으로 저장할 수 있고, 사용자가 원하는 문서를 빠르고 정확하게 유연하게 검색할 수 있는 지식관리 시스템을 구축하고자 한다.

### 2. 지식관리 시스템

일반적인 지식관리 시스템은 Figure 1과 같이 휴먼 인터페이스, 검색 시스템, 지식 기반(knowledge base)으로 구성되어 있다[1]. 휴먼 인터페이스는 사용자 인터페이스와 전문가 인터페이스로 나뉜다. 사용자 인터페이스는 시스템 사용자부터 질의어(query)를 입력받아 검색 결과를 보여주며, 전문가 인터페이스는 지식 기반에 지식을 추가하고 통제하는 데 사용된다. 지식 기반은 지식을 저장하는 기능을 담당한다. 지식관리 시스템의 주요 기능은 질문에 대한 대답을 찾아주는 것과 사용자가 입력한 질의어를 바탕으로 사용자가 원하는 지식을 찾아 답해주는 것으로 나눌 수 있다. 전자는 추론 시스템에 의해서, 후자는 검색 시스템에 의해 수행된다.

#### 2.1. 지식 맵

현재 지식관리의 주요 기능을 위해서 널리 사용되고 있는 것이 지식 맵(Knowledge Map)이다. 지식 맵은 하나의 문제를 여러 개의 작은 문제로 분할하여 해결할 수 있고, 모델의 구축이나 사용이 단순하다는 장점을 가지고 있다. 그러나 검색 기능보다는 문제에 답을 해주는 것에 적합하기 때문에 사용자가 원하는 문서를 검색하는데 있어서는 정확도가 떨어진다. 그리고 한번 작성되고 난 후에는 다시 수정하기가 어려울 뿐만 아니라 동의어나 유의어가

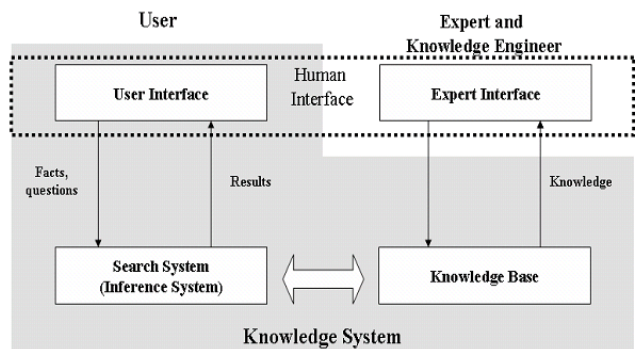


Figure 1. 전통적 KM system 구조[1]

† 본 연구는 한국과학재단 목적기초연구(R01-2002-000-00155-0) 지원으로 수행되었음.

검색어로 입력되었을 경우 이를 처리할 수 있는 유연성이 떨어진다. 또한 확장 가능성 면에서도 한계를 가지고 있다[14].

### 2.2. 온톨로지

온톨로지는 인공지능 분야에서 비롯하여 여러 분야에서 중요한 도구로 사용되고 있다. 특히 데이터베이스에서의 지식 발견(KDD: Knowledge Discovery in Database)과 지능형 인터넷 검색(IIS: Intelligent Internet Searching)에 있어 새로운 연구 방향을 제시하고 있다. 온톨로지를 작성하는 방법으로는 여러 가지가 있으나, 가장 기본적인 것은 개념들의 계층적 구조화이다. 이러한 구조에서는 상위 노드와 하위 노드가 서로 'is-a' 관계를 가지게 되며, 의미의 특수화 및 일반화를 통해 보다 유연하게 문서를 검색할 수 있게 된다[3]. 그리고 검색을 위한 도구로서 온톨로지를 사용할 경우, 검색의 유연성과 투명성이 보장된다[16].

따라서 본 논문에서는 지식 맵을 사용하는 지식 관리 시스템의 비유연성을 극복하기 위하여, 온톨로지 맵을 새롭게 정의하고 이를 지식 관리에 이용하고자 한다.

### 3. 온톨로지 맵을 이용한 지식관리

본 논문에서 제안하는 지식관리 시스템은 2절에서 언급한 전통적인 지식관리 시스템의 틀을 따르며, 문서 검색을 중심기능으로 한다.

온톨로지 맵은 특정 조직 내에서 사용하는 개념들의 계층 구조로 정의한다. Figure 2는 온톨로지 맵을 이용한 지식관리 과정을 보여준다. 사용자 질의와 새로운 문서는 온톨로지 맵을 매개로 지식 기반과 비교된다. 사용자로부터 입력된 질의어는 ①에서 ⑥의 지식 검색 과정을 통해 그 결과를 보여준다. 먼저, 입력된 질의어는 몇 개의 용어로 이루어진 벡터( $\vec{q}$ )로 처리된다. 질의어 벡터가 WordNet으로 입력되며, 온톨로지 맵 내의 개념과 비교된다. 벡터 성분 중 온톨로지 맵의 개념과 일치하는 것이 없는 경우, WordNet을 참조하여 온톨로지 맵 내의 가장 유사한 개념으로 변환된다. 이 변환된 질의어( $\vec{q}$ )가 온톨로지 맵으로 전송된다.

변환된 질의어 벡터와 온톨로지 맵 사이의 유사도를 계산하고, 지식 기반 내의 순위 테이블, 유사도 테이블을 이용해 질의어와 각 문서 간의 유사도를 계산한다. 유사도가 높은 순서대로 문서 테이블로부터 문서에 대한 정보를 추출하여 사용자에게 전달한다.

새로운 지식이 입력되는 경우, ③과 ④의 과정

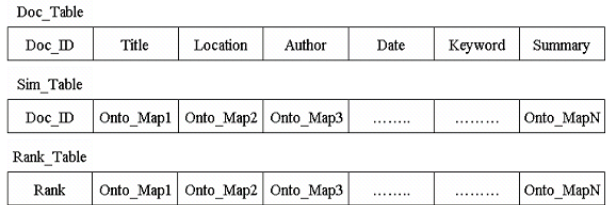


Figure 2. 지식 기반 Table 구조

을 거쳐 지식 기반에 저장된다. 온톨로지 맵으로 전송된 새로운 문서와 각 온톨로지 맵 간의 유사도를 측정하고, 문서 테이블, 유사도 테이블 그리고 순위 테이블을 갱신한다. 문서 내에 온톨로지 맵 내에는 존재하지 않는 용어들은 후보 용어 목록에 저장되며, 추후 온톨로지 맵 수정에 사용된다.

기존의 지식관리 시스템과는 달리 본 시스템의 지식 기반은 문서 자체를 저장하지 않는다. 문서 자체를 저장할 경우, 저장 공간의 제약과 전체 문서의 관리의 어려움이 따른다. 또한 사용자로부터 질의어가 입력될 때마다 온톨로지 맵과 문서간의 유사도를 계산해야 하므로, 문서 검색 속도가 현저하게 떨어진다. 따라서 지식 기반에 문서 자체를 저장하는 대신, 문서 테이블, 유사도 테이블, 순위 테이블을 통해 문서의 메타 데이터(meta data)만을 저장한다.

Figure 3은 각 테이블에 저장될 내용을 보여준다. 문서 테이블은 각 문서에 대한 정보, 예를 들어 문서ID, 저장위치, 제목, 저자, 작성일자, 키워드, 요약문과 같은 정보를 저장하며, 유사도 테이블은 각 온톨로지 맵과 문서 간의 유사도를 계산한 값을 저장한다. 순위 테이블은 각 온톨로지 맵별로 가장 큰 유사도를 가지는 상위 'k'개의 문서 인덱스를 저장한다. k의 크기는 조직의 규모나 전문성 등을 고려하여 지식관리 전문가가 결정한다. 순위 테이블과 유사도 테이블은 온톨로지 맵의 수정 혹은 확장이 이루어지거나 새로운 문서가 입력될 때마다 갱신된다.

#### 3.1. 온톨로지 맵 설계

Jansen, Spink 그리고 Saracevic의 연구에 따르면 사용자가 검색엔진에 입력하는 용어의 수는 평균 2.1개이다[4]. 즉, 대부분의 사용자들이 자신이 원하는 정보를 한 두 개의 키워드로 표현하고 있으며, 용어들 간의 관계를 정의할 수 있는 용언의 사용은 미약하다는 것을 의미한다. 만약 동사가 입력되었다고 키워드 간의 관계를 정의하기보다는 그 자신이 하나의 키워드로서 사용된다. 많은 검색 시스템들이 이런 관점에서 키워드 매칭(keyword matching) 혹은 이진 매칭(binary matching) 방법을 문서 간 유

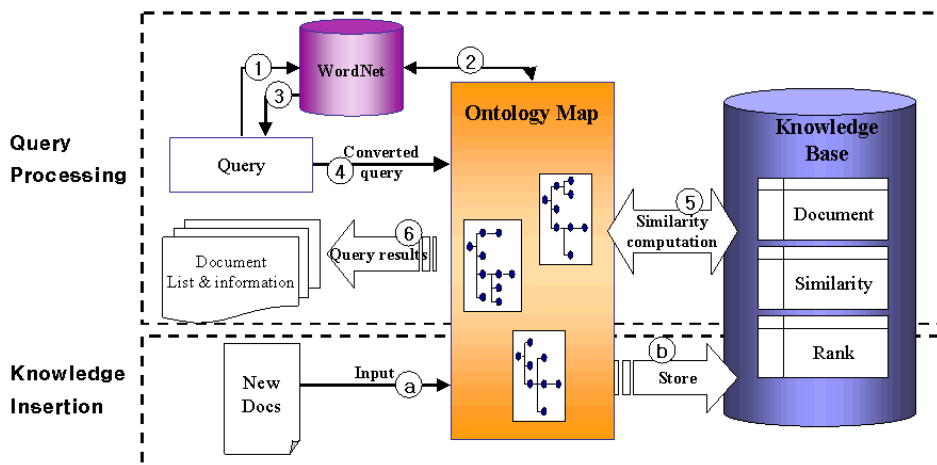


Figure 3. 온톨로지 맵을 이용한 지식관리

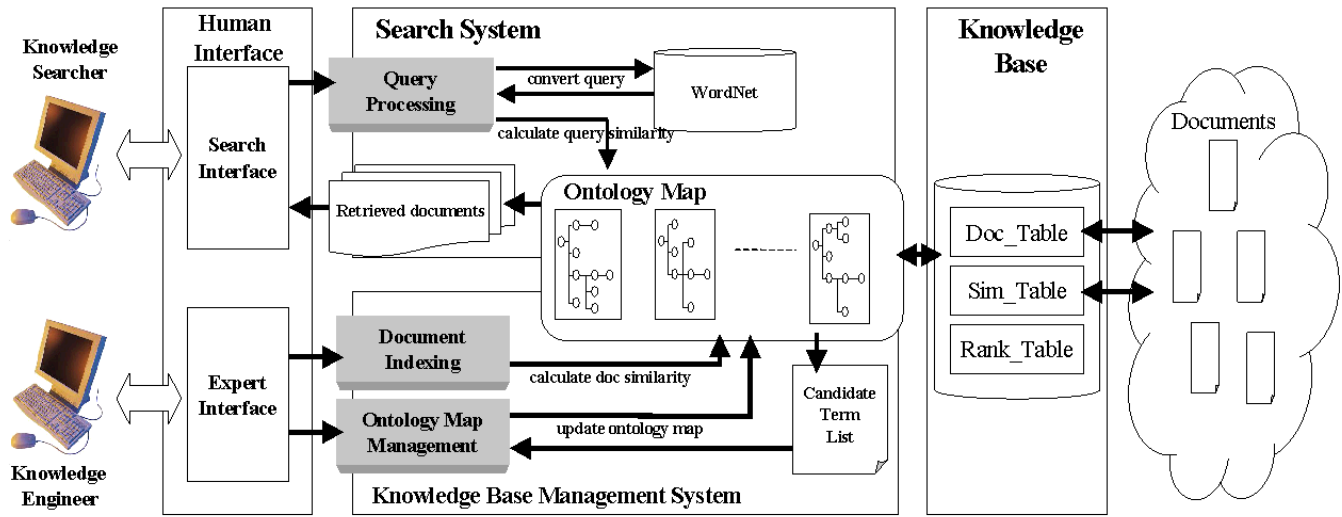


Figure 4. 지식관리 시스템 설계

사도 측정에 사용하고 있지만, 이러한 방법은 질의어와 문서의 문맥(context)을 고려하지 못한다는 한계를 가진다[5]. 본 연구에서는 이러한 제약을 극복하기 위하여 개념들 간의 계층적 구조를 나타내는 온톨로지 맵을 사용한다.

초기 온톨로지 맵을 생성하는 방법으로는 군집화를 이용하는 방법과 Open directory project(ODP)에서 정의한 카테고리[8]를 이용하는 두 가지 방법이 사용될 수 있다. 전자는 현재 저장된 문서를 군집화를 통해 몇 개의 군집으로 나누고, 각 군집에 속한 문서에서 대표적인 키워드를 추출한다. 지식관리 전문가들은 이러한 키워드를 바탕으로 각 군집을 표현하는 온톨로지 맵을 작성한다. 군집화 알고리즘으로는 DBSCAN[6]을 사용하며, 문서간의 거리는 다음 식(1)의 Jaccard 계수를 사용한다.

$$dis(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

군집화를 이용하는 방법은 지식관리 시스템이 설치되기 이전에 보유하고 있는 문서가 있어야 하지만, 그렇지 못한 경우에는 ODP의 카테고리를 이용하여 지식관리 전문가가 직접 해당 조직에 적합한 초기 온톨로지 맵을 작성할 수 있다.

지식 기반에 입력된 새로운 문서 내부의 키워드들 중에서 온톨로지 맵 내의 개념들과 일치하지 않는 키워드는 후보 용어 목록에 추가된다. 온톨로지 맵 내부의 개념들을 후보 용어들로 대체하거나, 후보 용어들을 새롭게 추가하여 기존의 맵을 수정할 수 있다. 혹은 완전히 새로운 온톨로지 맵을 작성하여 추가할 수 있다. 온톨로지 맵을 수정하는 시기는 상황에 맞게 지식관리 전문가가 결정한다. 예를 들어 일정 시간이 지났을 때, 목록 내의 용어 개수가 일정 수준을 넘었을 때, 혹은 해당 조직의 관련 분야가 바뀌거나 혹은 관련 분야 자체의 급격한 변화로 인해 지식관리 전문가가 온톨로지 맵의 수정이 필요하다고 판단되는 시기에 온톨로지 맵을 수정하거나 추가한다. 이 과정에서 사용되는 후보 용어 역시 지식관리 전문가에 의해 취사선택되어야 한다.

### 3.2. WordNet을 이용한 온톨로지 맵의 확장

앞 절에서 언급한 바와 같이 온톨로지를 이용한 지식 검색은 이전 매칭의 한계를 극복할 수 있으나, 동의어 및 유의어에 대한 처리가 불가능하다. 즉, 사용자가 입력한 질의어 혹은 새로운 문서의 키워드가 온톨로지 맵에 직접 존재하지 않지만 동일하거나 유사한 의미를 가지는 용어를 포함할 수

있다. 본 연구에서는 WordNet을 이용하여 이러한 상황을 처리한다. WordNet은 특정 용어에 대한 정의와 함께, 동의어, 유의어, 반의어 등과 같은 정보를 제공해준다[2]. 사용자가 입력한 질의어 중 온톨로지 맵 내에 존재하지 않는 용어가 입력되는 경우, WordNet를 참조하여 이 용어와 가장 유사한 의미를 가지는 개념과 유사도를 나타내는 가중치 값을 온톨로지 맵과 질의어 사이의 유사도 측정에 이용할 수 있다[3]. 이를 통해 개념의 특수화와 일반화를 이룰 수 있다.

## 4. 유사도 측정

본 연구에서는 지식 검색에 적합한 문서들을 제공하기 위하여, 입력된 질의어와 온톨로지 맵과의 유사도, 온톨로지 맵과 문서와의 유사도를 측정한다. 최종적으로 이 두 유사도를 결합하여 질의어 가장 적합한 문서 목록을 제공한다. 사용자로부터 입력되는 질의어는 벡터의 형태로, 온톨로지 맵은 여러 개념들의 나무구조를 가정한다.

### 4.1. 온톨로지 맵과 질의어 간의 유사도

질의어는  $k$ 개의 용어로, 온톨로지 맵은  $l$ 개의 개념들로 구성된 집합으로 가정한다. 앞 절에서 설명한 방법대로 질의어 내의 용어( $t_j$ )은 온톨로지 맵 내의 한 개념( $c_i$ )으로 가중치( $s_j$ ) 값을 가지고 사상된다. 사상된 개념들 간의 유사도는 Wu&Palmer 척도[11]( $S_{W\&P}$ )를 이용하며, 다음 식(2)를 이용하여 계산한다.

$$S_{W\&P}(a, b) = \frac{depth(c)}{depth(a) + depth(b)} \quad (2)$$

$c$ 는  $a$ 와  $b$ 의 공통 부모 노드 중 깊이 값이 가장 큰 노드를 의미한다. 온톨로지 맵( $O$ )과 변환된 질의어( $\vec{q}$ ) 사이의 유사도( $w(\vec{q}, O)$ )는 다음 식(3)을 이용해 계산한다[3].

$$w(\vec{q}, O) = \frac{1}{k C_2} \left[ \sum_{(i,j)} \frac{s_i + s_j}{2} S_{W\&P}(c_i, c_j) \right] \quad (3)$$

### 4.2. 온톨로지 맵과 문서와의 유사도

온톨로지 맵( $O$ )은  $l$ 개의 개념들로 이루어진 나무 구조를 가지며, 문서( $D$ )는  $m$ 개의 키워드들로 이루어진 집합으로 가정한다. 온톨로지 맵과 문서와의 유사도는 F-measure를 이용한다. F-measure

는 precision과 recall값의 조화평균값을 나타내며, precision, recall 그리고 F-measure 값은 다음 식(4)와 같이 계산된다[10].

$$p: precision(O, D) = \frac{|O \cap D|}{|D|}$$

$$r: recall(O, D) = \frac{|O \cap D|}{|O|} \quad (4)$$

$$F(O, D) = \frac{2pr}{p+r}$$

$|D|$ 는 문서 내 포함된 키워드의 수,  $|O|$ 는 온톨로지 맵 내의 개념의 개수를 의미한다.

#### 4.3. 질의어에 대한 문서의 유사도

입력된 질의어와 문서 사이의 유사도는 식 (5) 같이 좌앞에서 계산된 질의어와 온톨로지 맵 사이의 유사도  $w(\vec{q}, O)$ 와, 온톨로지 맵과 문서 사이의 유사도  $F$ 의 선형 결합을 통해 계산한다.

$$sim(\vec{q}, D_i) = \sum_j w(\vec{q}, O_j) F(O_j, D_i) \quad (5)$$

이 유사도 값이 큰 순서대로 문서를 추출하고, 검색 시스템의 결과로 문서 목록과 각 문서의 정보를 사용자에게 제공한다.

### 5. 시스템 설계

Figure 4는 온톨로지 맵을 이용한 시스템의 구조를 보여준다. 전체 시스템은 크게 휴먼 인터페이스, 검색 시스템, 지식 기반 관리 시스템, 지식 기반으로 나눌 수 있다. 일반 사용자와 지식관리 전문가는 웹 기반의 사용자 인터페이스와 전문가 인터페이스를 통해 시스템을 이용하거나 통제할 수 있다. 검색 시스템은 사용자로부터 입력된 질의어를 처리하는 모듈과 WordNet 그리고 온톨로지 맵으로 구성된다. 지식기반 관리 시스템은 새롭게 입력된 문서를 처리하는 모듈과 온톨로지 맵을 직접 수정하거나 확장시키기 위한 모듈, 온톨로지 맵 그리고 후보 용어 목록으로 구성된다. 지식 기반은 문서 테이블, 유사도 테이블, 그리고 순위 테이블로 이루어져있다. 온톨로지 맵은 검색 시스템과 지식기반 관리 시스템 그리고 지식 기반을 연결하는 매개체가 된다.

### 6. 결론

본 논문에서는 지식관리 시스템의 중요성을 인식하고, 암묵지의 가장 일반적인 저장매체인 문서를 효율적으로 관리하고 검색할 수 있는 방안을 제안하였다. 시스템의 확장성과 문서 관리 및 검색에 있어 효율성을 높이기 위해 온톨로지 맵을 새롭게 정의하고 이를 이용하는 방법을 제안하였다. 각 문서와 온톨로지 맵 간의 유사도를 질의 단계 이전에 미리 계산하고, 상위 k개의 문서 정보만 보유하는 방법을 선택하였다. 시스템의 공간 복잡도를 줄이기 위해 각 문서에 대한 메타 데이터만을 저장한다.

본 연구에서 제안하는 시스템을 통해, 기존의 지식관리 시스템에 비해 사용자에게 보다 더욱 정확하고 빠른 문서 검색 서비스를 제공할 수 있다. 그리고 시스템 관리 측면에서 온톨로지 맵의 확장이 용이하기 때문에 조직의 규모가 커질 경우 혹은 관련 분야가 바뀌거나 해당 분야 내의 급격한 변화가 있을 경우에 유연하게 대처할 수 있다.

앞으로 본 논문에서 설계된 시스템을 실제로 구축하여 시스템의 성능을 측정하고, 기존 지식관리

시스템에 비해 성능향상 정도를 평가하고자 한다. 그리고 문서뿐만 아니라 이미지, 음성 및 영상 자료 등의 다양한 지식 저장 매체로 확장할 수 있는 방법을 연구하여 시스템에 적용하고자 한다.

### References

- [1]Mark Stefik(1995), *Introduction to knowledge system*, Morgan Kaufmann Publishers, 296-299,
- [2]George A. Miller(1995), WordNet: A Lexical Database for English, *Communications of the ACM*, Volume 38, No.11
- [3]Maria Halkidi, Benjamin Nguyen, Iraklis Varlamis and Michalis Vazirgiannis(2003), THESUS: Organizing Web document collections based on link semantics, *The VLDB Journal* (2003) 12, 320-332
- [4]Bernard J. Jansen, Amanda Spink and Tefko Saracevic (Jan.2000), Real life, real users, and real needs: a study and analysis of user queries on the web, *Information Processing and Management: an International Journal*, Volume 36 (2), 207-227
- [5]Brin S. and Page L.(1998), The anatomy of a large scale hypertextual Web search engine, Proceedings of the 7th International World Wide Web conference, Computer Network ISDN System, Volume 30, 107-117
- [6]Martin Ester, Peter Kriegel, Jörg Sander and Xiaowei Xu (1996), A density based algorithm for discovering clusters in large spatial databases with noise, Proceedings of the 2nd international conference on knowledge discovery and data mining, ACM-SIGKDD, 226-231
- [7]Eiter T. and Mannila H. (1997), Distance measures for point sets and their computation, *Acta Informatic*, Volume 34 (2), 109-133
- [8]Open directory project: <http://www.dmoz.org>
- [9]Kathy A. Stewart, Richard Baskerville, Veda C. Storey, James A. Sem, Arjan Raven, Cherie Long(2000), Confronting the Assumptions Underlying the Management of Knowledge: An Agenda for Understanding and Investigating Knowledge Management, *The Database for Advances in Information Systems*, Volume 31, No.4
- [10]Ricardo Baeza-Yates, Berthier Ribeiro-Neto(1999), *Modern Information Retrieval*, ACM Press, New York
- [11]Wu Z., Palmer M.(1994), Verb semantics and lexical selection, Proceedings of the 32nd annual meetings of the associations for computational linguistics, Las Cruces, NM, 133-138
- [12]Atreyi Kankanhalli, Fransiska Tanudidjaja, Juliana Sutanto, Bernard C.Y. Tan(2003), The Role of IT in Successful Knowledge Management Initiatives, *Communications of the ACM*, Volume 46, No.9, 69-73
- [13]M. Polanyi(1966), *The Tacit Dimension*, Doubleday, Garden City
- [14]Jeong-Han Woo, Mark J. Clayton, Robert E. Johnson, Benito E. Flores, Christopher Ellis(2004), Dynamic Knowledge Map: reusing experts' tacit knowledge in the AEC industry, *Automation in Construction*, Volume 13, 203-207
- [15]Thomas A. Stewart(1997), *Intellectual Capital: The New Wealth of Organization*, New York: Currency Doubleday
- [16]R. Braga, M. Mattoso, C. Werner,(2000), Using Ontologies for Domain Information Retrieval, DEXA 2000 DomE Workshop, 100-104