

## 문서 유사도 기반의 웹 마이닝 시스템 개발

이강찬\*, 민재홍\*, 박기식\*, 임동순\*\*, 우훈식\*\*\*

### Development of A Web Mining System Based On Document Similarity

Kangchan Lee, Jaehong Min, Kishik Park, Dong-Soon Yim, Hoon-Shik Woo

#### Abstract

In this study, we proposed design issues and structure of a web mining system and develop a system for the purpose of knowledge integration under world wide web environments resulted from our developing experiences. The developed system consists of three main functions: 1) gathering documents utilizing a search agent; 2) determining similarity coefficients between any two documents from term frequencies; 3) clustering documents based on similarity coefficients. It is believed that the developed system can be utilized for discovery of knowledge in relatively narrow domains such as news classification, index term generation in knowledge management.

*Keyword : Web Mining, Knowledge Base, Document Similarity, Knowledge Discovery*

---

\* 한국전자통신연구원 정보화기술연구소 표준연구센터  
\*\* 한남대학교 공과대학 산업공학과  
\*\*\* 대전대학교 공과대학 컴퓨터정보통신공학부

## 1. 서론

월드 와이드 웹은 사용자와 웹 간의 상호 작용이라는 독특한 기능을 바탕으로, 하나의 거대한 정보 원천으로써 정보의 수집 및 유통 그리고 확산을 매우 보편화시켰다. 하지만, 이렇게 정보가 일반 사용자에게 보편화되는 동안, 너무나 많은 정보 때문에 사용자 입장에서는 자신이 원하는 정보를 정확하게 획득할 수 없는 소위 정보의 홍수 현상이 발생하게 되었다.

일반적으로 사용자가 웹에서 특정 정보를 찾고자 할 때는, 서치 서비스 및 웹 디렉토리를 이용하거나 혹은 특정 사이트를 중심 축으로 사용자 스스로 브라우징을 한다. 서치 서비스를 사용하는 경우, 사용자는 단순한 형태의 키워드 질의를 입력하며, 서치 서버는 이러한 질의를 기반으로 상대적으로 높은 유사성을 갖는 웹 문서를 화면에 표시한다.

하지만, 현재 사용 가능한 서치 서비스는 서치 결과의 정확도에 많은 문제점을 안고 있다. 즉, 서치 서비스가 단순히 사용자의 키워드 질의만을 검색의 기준으로 정하다 보니, 실제적으로 문서에 해당 키워드가 나타나기는 하지만 관련 문서 수가 대단히 많기 때문에 전체적으로 사용자에게 정보 검색의 의미를 전달하지 못하는 상황이 발생하는 것이다. 이러한 결과는 사용자로 하여금 서치 서비스에 대한 불신으로 이어지며, 사용자의 입장에서는 직접적으로 관계되는 정보를 찾기 어렵게 되는 결과를 초래한다.

이러한 서치 엔진의 문제점에 대한 해결 대안으로 월드 와이드 웹을 대용량의 정보

베이스로 인식하여 자동으로 정보를 발견하고 추출하여 사용자 중심의 지식 베이스를 제공하는 문서 유사도 기반의 웹 문서 마이닝이 부상하고 있다. 웹 문서 마이닝은 월드 와이드 웹 상에 수 없이 존재하는 정보를 효과적으로 인덱싱하여 사용자가 원하는 정보를 즉각적으로 지식 베이스의 형태로 제공하는 것으로 지식 정보의 통합을 위해 매우 중요하다.

지식 베이스는 웹과 같이 정보가 대량으로 유지되고 있는 상황에서, 이렇게 유지되는 보유 데이터 중 유용한 지식이 될 가능성이 높은 것을 추출하는 것에 초점이 맞추어져 있다. 즉, 효과적인 지식 베이스의 구축을 위해서는 단순한 질의어로는 추출이 불가능한 숨겨진 형태의 정보를 찾아내어 이를 바탕으로 데이터에 대한 해석 및 의사 결정을 지원하는 웹 마이닝이 필수적인 것이다.

## 2. 웹 마이닝 시스템

웹 마이닝은 웹 문서의 내용을 분석하여 특정한 주제를 포함하는 문서를 검색하거나 혹은 주제와 관련된 문서를 기준에 따라 그룹화하여 사용자 중심의 지식을 제공하는 것이다. 이러한 웹 마이닝의 기본적인 개념은 기존의 대용량 데이터베이스 분야에서 지식 추출에 대한 유용성이 검증된 데이터 마이닝 방법을 월드 와이드 웹에 적용하고자 하는 것이다. 즉, 전통적인 데이터 마이닝이 구조화되어 있는 데이터베이스를 연구의 대상으로 하는 반면에 웹 문서 마이닝은 웹을 대상으로 데이터 마이닝 기술을 사용

하여 자동으로 월드 와이드 웹으로 부터 정보를 발견하고 추출하여 의사 결정에 필요한 유용한 정보를 지식 베이스의 형태로 제공하는 것이다.

이렇게 웹 마이닝 연구 분야는 마이닝의 결과가 웹 지식 베이스나 웹 웨어하우스의 형태로 나타나는 정보 통합 분야이며, 정보 통합 분야는 기존의 데이터베이스, 정보 검색, 통계학, 인공 지능, 그리고 기계 학습 등의 연구 분야가 종합적으로 연구되는 학문 분야이다 [1].

웹 정보 통합을 위한 웹 마이닝 분야의 연구는 두 가지로 구분된다. 즉, 정보 검색 중심과 데이터베이스 중심의 연구로 나뉘어진다. 정보 검색 중심의 연구는 사용자에게 정보 발견 혹은 필터링을 지원하는 것으로 지능형 웹 에이전트와 같은 정보 검색을 위한 보다 지능적인 도구를 사용하는 것이다.

반면에, 데이터베이스 중심의 연구는 웹 상의 데이터를 모델링하거나 이를 통합함으로써 일반적인 키워드 질의보다 정교한 질의가 가능하도록 하는 것이다. 웹 마이닝의 주요 연구 분야는 다음과 같다.

## 2.1 정보 검색 접근 방법

웹 콘텐츠 정보를 나타내는 구조 정보가 부족하기 때문에 웹 문서를 자동으로 발견하는 것은 어려운 일이다. 기존의 서치 엔진이 사용자에게 일부 정보는 검색을 통하여 전달하지만, 웹 문서의 구조 정보, 카테고리, 필터, 해석 등의 보다 정교한 정보는 제공하지 못한다. 따라서, 보다 효율적인 웹 문서 마이닝을 위하여 현재 지능형 웹 에이전트

와 같은 정보 검색을 위한 보다 지능적인 도구의 개발이 진행되고 있으며, 웹 상에 존재하는 반구조 데이터에 대한 마이닝 기술이 개발되고 있다.

### - 지능형 서치 에이전트

사용자 프로파일과 도메인 특성을 사용하여 발견된 정보를 조직화하고 해석하여 관계 정보를 서치하는 다수의 지능형 웹 에이전트가 개발되었다. Harvest, FAQFinder, Information Manifold, OCCAM, ParaSite 등의 웹 에이전트는 문서의 특정 형태에 대한 사전에 정의된 도메인 정보나 혹은 문서의 검색 및 해석을 위한 정보 소스 모델에 기반하여 정보를 검색한다. ShopBot [3], ILA (Internet Learning Agent) [11]와 같은 에이전트는 정보 소스의 구조에 대하여 기계 학습을 하거나 상호 작용한다. ShopBot 은 상품 도메인에 대한 일반적인 지식을 이용하여 다양한 공급자 사이트로 부터 상품 정보를 검색한다. ILA 는 다양한 정보 소스의 모델을 학습하여, 이를 고유의 개념 구조에 매핑하는 방법을 사용한다.

### - 정보 필터링/범주화

다수의 웹 에이전트가 웹 문서의 검색, 필터링, 분류를 위하여 정보 검색 기술과 자동 검색을 위한 하이퍼텍스트 문서의 특성을 사용한다. Hypursuit [14]는 하이퍼 문서의 클러스터 계층 구조를 생성하기 위하여 링크 구조와 문서 콘텐츠에 내재되어 있는 의미 정보를 이용한다. BO (Bookmark Organizer) [9]는 계층적 클러스터링 기술과 사용자 상호 작용을 이용하여 개념 정보에

기반한 웹 문서 클러스터링을 구축하였다.

#### - 개인형 웹 에이전트

개인형 웹 에이전트는 개인 사용자의 선호도를 학습하고 이러한 선호도에 따른 웹 정보 소스를 검색한다. 또한, 유사한 선호도를 갖는 사용자 간의 협업 필터링을 사용하여 정보 검색의 정확도를 높이고자 한다. 이러한 웹 에이전트에는 WebWatcher, PAINT, GroupLense 등이 있다.

## 2.2 데이터베이스 접근 방법

웹 마이닝에 대한 데이터베이스 접근 방법은 웹 상의 정보를 보다 구조화된 형식으로 조직화하는 기술과 데이터베이스 질의 매커니즘을 개발하는 것이다.

#### - 멀티 레벨 데이터베이스

멀티 레벨 데이터베이스의 기본적인 개념은 가장 낮은 수준의 데이터베이스가 다양한 웹 리퍼지토리에 저장된 반구조 문서 즉 하이퍼텍스트 문서를 보관하는 것이며 상위 레벨에서는 하위 레벨의 정보에 대한 메타 데이터 혹은 일반화 정보가 관계형 데이터베이스 혹은 객체 데이터베이스와 같은 구조화된 형태로 저장되는 것이다. 이렇게, 정보를 다층으로 구성하여 상위 레벨에서의 정보 구조화를 추구하였다. Zaiane [15]는 각 계층이 하위 계층의 일반화로 생성되는 다계층 데이터베이스를 제안하였다. King 과 Novak [5]은 하나의 이질 글로벌 데이터베이스 스키마 대신에 각 정보 소스로부터 점진적으로 글로벌 스키마를 생성해가는 스

키마 통합 방법을 제안하였다. ARANEUS [10] 시스템은 하이퍼텍스트 문서로부터 관련 정보를 추출하고 이 정보를 상위 레벨의 유도된 웹 하이퍼텍스트에 통합하는 방법을 제안하였다.

#### - 웹 질의 시스템

웹 질의 시스템은 웹 상에서 SQL 형식의 데이터베이스 질의어, 웹 문서의 구조 정보, 자연어 기반 질의 처리 언어를 사용하는 것이다. W3QL[6]은 하이퍼텍스트 문서의 조직에 기반한 구조 질의와 정보 검색 기술에 기반한 내용 질의를 연동한다. WebLog [7]는 웹 정보 소스로부터 정보를 추출하기 위하여 로직 기반 질의 언어를 사용한다. Lorel [12]는 라벨드 그래프 데이터 모델을 사용하여 웹 상의 이질 반구조 정보를 질의한다. TSIMMOS [2]는 이질 반구조 정보 소스로부터 데이터를 추출하여 이렇게 추출한 정보를 통합 데이터베이스 관점에서 표현하기 위하여 상호 연동을 추구하는 것이다.

이렇게 웹 마이닝 연구는 데이터 마이닝 개념을 이용하여 웹을 지식 베이스로 활용하는 것에 초점이 맞추어져 있으며, 최근의 연구는 서치 에이전트에 의해 획득된 웹 문서 정보를 사용자의 요구를 기초로 문서 간의 유사도에 따라서 필터링과 그룹핑을 수행하여 지식을 체계화하는 연구가 진행되고 있다. 즉, 웹의 정보를 지식 베이스로 변환하기 위해서 서치 에이전트를 이용하여 정보를 검색하고 검색된 결과를 유사도 측정을 이용하여 서로 연관된 웹 문서 간의 주요 특징을 추출하여 체계화하는 것이다.

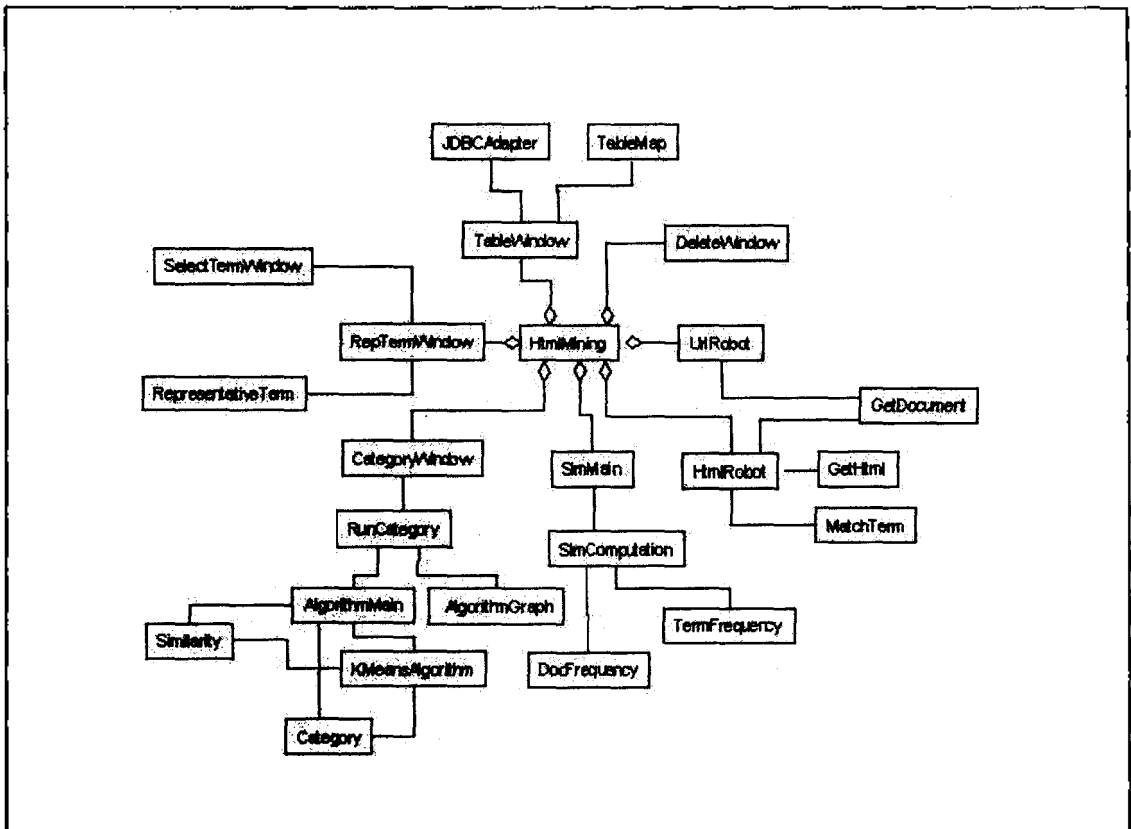
### 3. 시스템 설계

문서 유사도 기반의 웹 마이닝 시스템의 개발을 위하여 객체 지향 분석 및 설계 방법의 OMG (Object Management Group) 표준인 UML(Unified Modeling Language)을 이용하였다. UML 은 업무 모델링, 데이터 모델링, 컴포넌트 모델링, 객체 모델링과 같은 기존의 다양한 모델링 기법을 결합한 통합 모델링 언어로써 소프트웨어 중심적인

시스템의 시각화, 상세화, 구축 및 문서화를 지원하기 위한 표준 언어이다 [4,8].

본 시스템 설계에서는 웹 마이닝 시스템 구현을 위한 클래스 다이어그램을 중심으로 작성하였으며 주요 내용은 다음과 같다.

클래스 다이어그램(Class Diagram)은 객체 지향 모델의 기본이 되는 클래스를 정의하고 이들 간의 연관성, 상속성을 정의하는 것으로, 시스템에 존재하는 객체의 타입과 정적인 관계를 설명한다.



<그림 1> 클래스 다이어그램

<그림 1>은 웹 마이닝 시스템에 대한 클래스 구조를 나타낸 것으로 주요 클래스는 다음과 같다:

- HtmlMining: 웹 HTML 문서 마이닝을 위한 주 클래스
- TableWindow: 데이터베이스 조회를 위한 유저 인터페이스 클래스
- DeleteWindow: 시간이 지난 HTML 문서를 제거하기 위한 유저 인터페이스 클래스
- UriRobot: 해당 URL 사이트의 정보를 가져오기 위한 유저 인터페이스 클래스
- HtmlRobot: HTML 문서를 가져오기 위한 유저 인터페이스 클래스
- SimMain: 유사성계수를 계산하기 위한 유저 인터페이스 클래스
- CategoryWindow: 유사성 카테고리 유저 인터페이스 클래스
- RunCategory: 카테고리 주 클래스
- SimComputation: 유사성 계수 계산 클래스
- AlgorithmMain: 유사성 알고리즘 주 클래스
- RepresentativeTerm: 유사성 기준을 위한 대표 용어를 구하고, 모든 대상 용어를 화일로 저장
- RepTermWindow: 대표 용어를 구하기 위한 유저 인터페이스 클래스
- SelectTermWindow: 대표 용어를 수정하기 위한 유저 인터페이스 클래스
- Similarity: 유사성 계수 클래스
- JDBCAdapter: JDBC/DB 어댑터 클래스
- TableMap: DB 테이블 및 자바 테이블 매핑 클래스
- GetDocument: 주어진 URL 의 html 파일

을 가져오는 클래스

- GetHtml: Html 문서를 읽어 용어를 추출하는 클래스
- MatchTerm: 주어진 단어가 용어에 속하는지를 결정하는 클래스
- ReadDocTerm: 한 용어를 포함하는 모든 문서를 가져오는 클래스
- ReadSiteUrl: DB 에서 URL 사이트를 읽는 클래스
- ReadTermDoc: DB 에서 한 문서에 포함된 모든 용어를 읽는 클래스
- ReadTermGlobal: DB 의 TermGlobal 테이블을 읽는 클래스

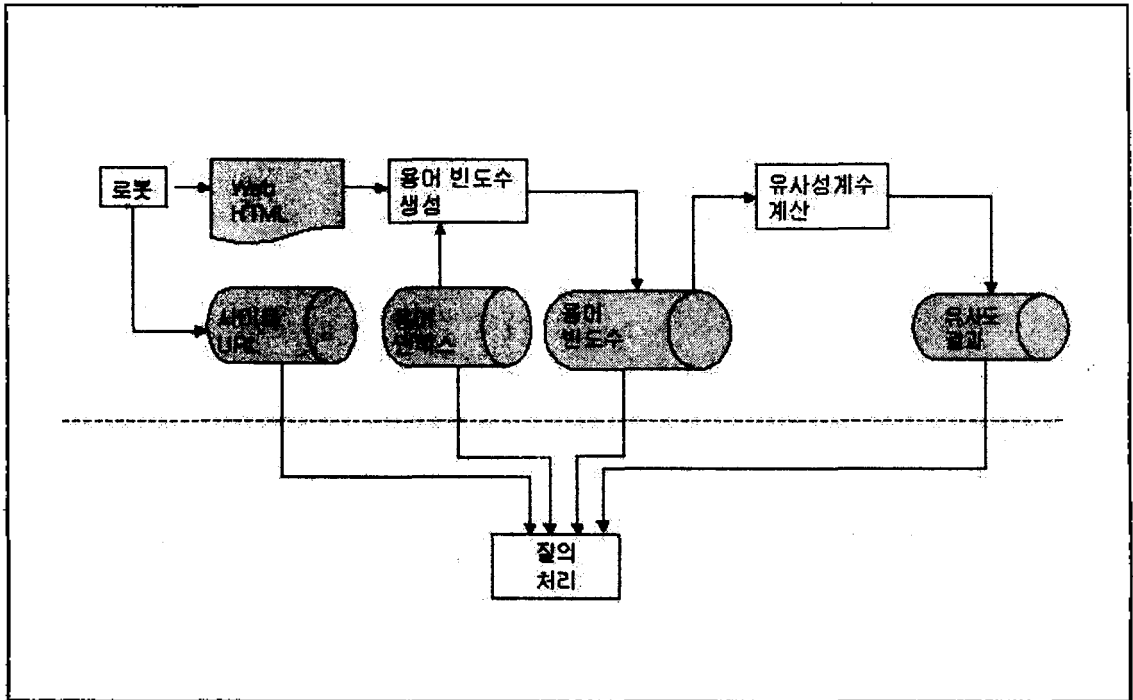
#### 4. 시스템 개발

본 연구에서 개발된 웹 마이닝 시스템의 구성은 <그림 2>와 같다. 시스템을 구성하는 주요 모듈은 로봇, 용어 빈도수 생성, 문서 유사성 계수 계산으로, 자바 언어를 이용하여 개발하였다. 주요 개발 내용은 다음과 같다:

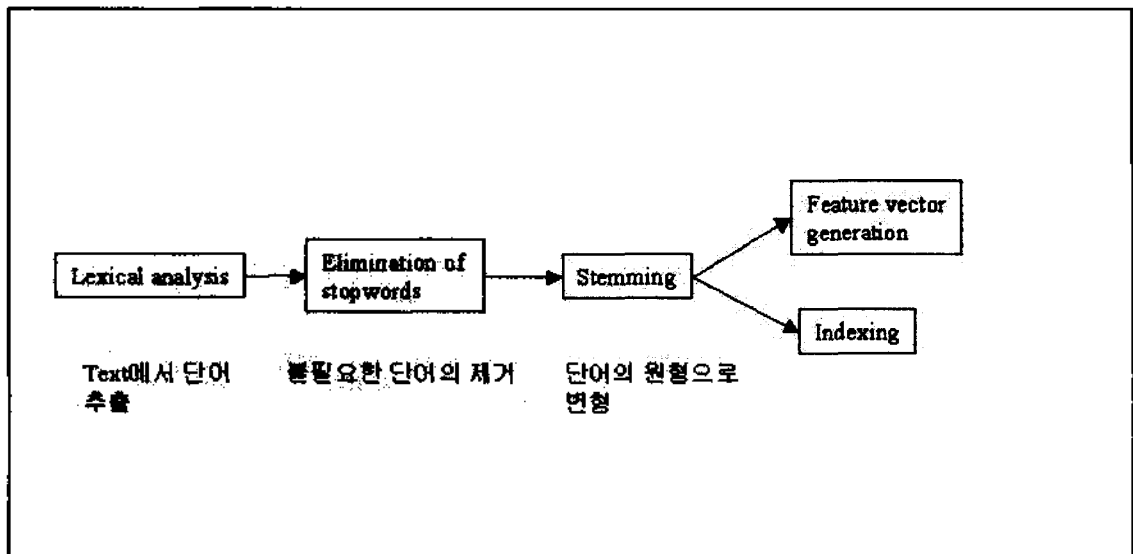
##### 1) 로봇

로봇은 시스템에 지정된 URL 리스트로부터 해당 사이트 주소를 획득하여 웹 문서를 가져오는 역할을 담당한다. 본 연구에서는 테스트를 위하여 디지털 타임스[16]와 전자신문[17] 두 곳의 사이트를 대상으로 하였다.

로봇은 사용자로부터 해당 날짜를 입력받아서 각 사이트에서 해당하는 날짜의 기사 URL 을 가져 오는 것과 주어진 URL 에서 신문 기사를 가져오는 두 가지 기능으로 구성되어 있다.



<그림 2> 웹 마이닝 시스템 구조도



<그림 3> 용어 빈도수 계산 절차

## 2) 용어 빈도수 생성

용어 빈도수 계산은 로봇이 가지고 온 웹 문서에 대하여 각 용어 별로 빈도수를 계산하기 위한 모듈이다. 즉, 용어 빈도수는 문서 내에 존재하는 특정 용어의 발생 빈도를 측정하는 것으로, 이 때 측정하고자 하는 용어는 관리자가 별도로 입력 및 수정할 수 있다. 본 연구에서는 용어 인덱스를 정보 통신과 관련된 용어를 선정하여 사용하였다. <그림 3>은 본 연구에서 사용된 용어 빈도수 계산 절차를 도식화 한 것이다.

## 3) 문서 유사성 계수 계산

본 연구에서는 문서 간의 유사성을 계산하기 위하여 Salton 의 vector space 모델 [13]을 이용한다. 기본적으로 문서의 의미는 사용된 용어에 의한다고 가정하여 각 문서를 나타내는 특성 벡터를 구한다. 각 문서를 벡터로 표현하는 방법은 일반적으로 tf-idf가 있으며, 정규화한 term vector 를 그대로 이용할 수도 있다. Keyword 에 해당하는 단어,  $k_i$  를 index 로 구성하였다고 하고 다음을 정의한다[13].

$n$ : 총 문서 수

$n_i$ :  $i$ 번 째 term 인  $k_i$ 가 존재하는 문서의 수 ( $i=1,2,\dots,t$ )

$freq_{ij}$ : 문서  $d_j$  에 나타난  $k_i$  의 횟수 (Term frequency: TF)

그러면,  $freq_{ij}$  를 정규화하기 위하여 다음식을 사용한다.

$$f_{ij} = \frac{freq_{ij}}{\max_i \{freq_{ij}\}}$$

또한, 다음 식을 이용하여 Inverse document frequency (IDF)를 계산한다.

$$idf_i = \log \frac{N}{n_i}$$

그러면, 최종적으로 각 문서의 벡터 성분은 다음식으로 계산된다.

$$w_{ij} = f_{ij} \times idf_i$$

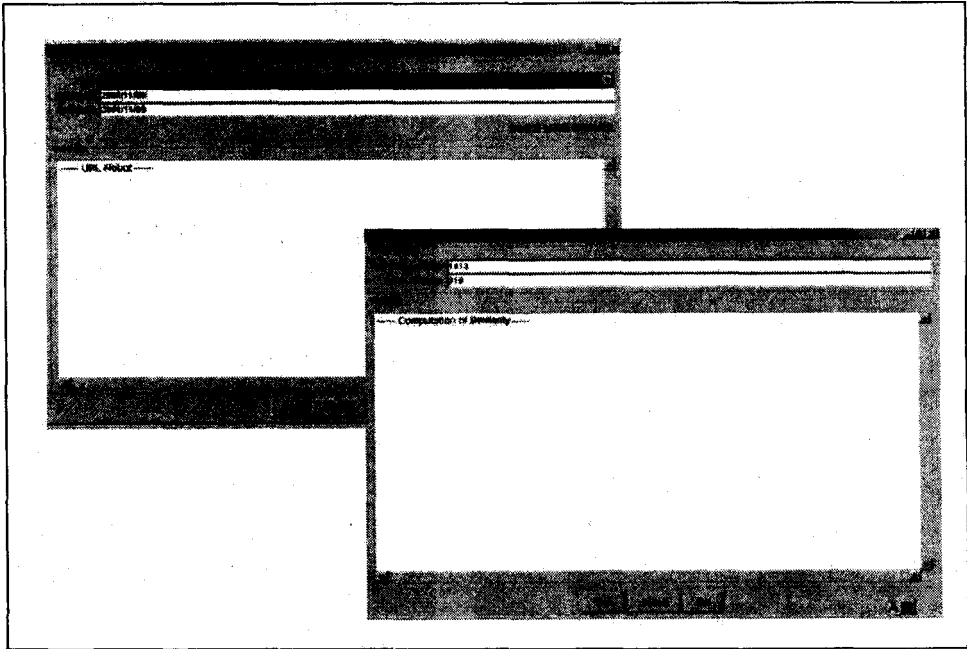
만약, 두 문서,  $d_1, d_2$  를 표현한 벡터를  $\vec{d}_1, \vec{d}_2$  라고 하면 이 벡터 간의 각도에 의한 두 문서간의 유사성은 다음과 같다.

$$sim(d_1, d_2) = \frac{\vec{d}_1 \cdot \vec{d}_2}{|\vec{d}_1| \times |\vec{d}_2|} = \frac{\sum_i w_{i1} w_{i2}}{\sqrt{\sum_i w_{i1}^2} \sqrt{\sum_i w_{i2}^2}}$$

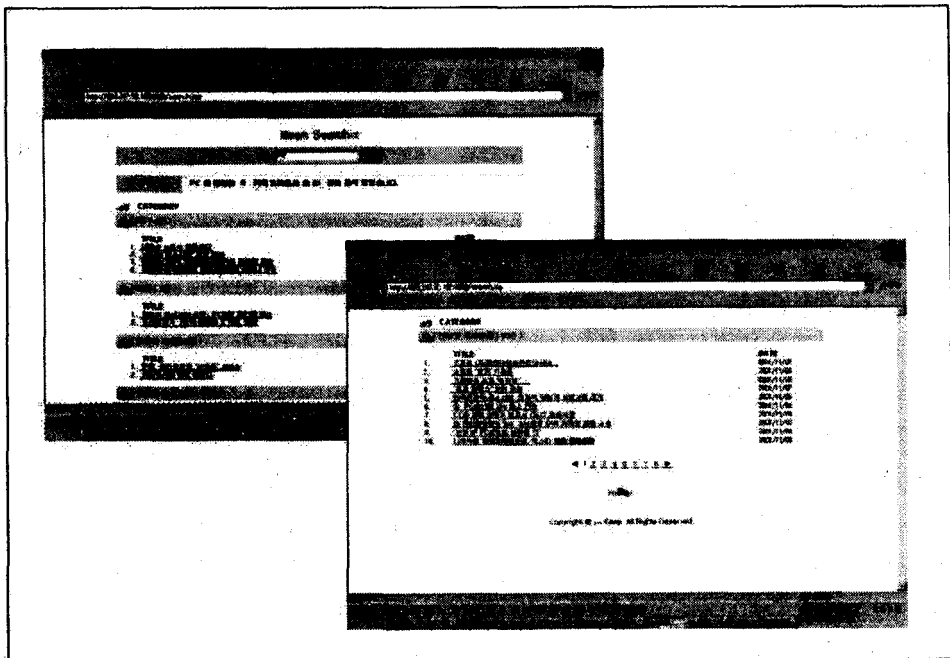
이 유사성계수의 값은 0 부터 1 사이에 존재하여 값이 클수록 두 문서는 유사하다고 할 수 있다.

<그림 4>과 <그림 5>는 본 연구의 결과 화면으로 <그림 4>는 서버 측면의 결과 화면으로 URL 로봇 화면과 유사도 계산 화면을 나타낸다. 또한 <그림 5>는 일반 사용자가 직접 대면하는 클라이언트 화면으로 사용자 입력 용어에 대하여 문서 유사도를 중심으로 카테고리 별 결과가 제시되었다.





<그림 4> 웹 마이닝 시스템 서버 화면



<그림 5> 웹 마이닝 시스템 클라이언트 화면

## 5. 결 론

인터넷의 폭발적인 이용으로 수 없이 많은 정보가 온라인 상에 존재한다. 이러한 정보를 효과적으로 인덱싱하고 클러스터링하여 사용자가 원하는 정보를 즉각적으로 지식 베이스의 형태로 제공하는 웹 문서 마이닝은 지식 정보의 통합을 위해 매우 중요하다.

지식 베이스는 웹과 같이 정보가 대량으로 유지되고 있는 상황에서, 이러한 보유 데이터 중 유용한 지식이 될 가능성이 높은 것을 추출하는 것에 초점이 맞추어져 있다. 즉, 효과적인 지식 베이스의 구축을 위해서는 단순한 질의어로는 추출이 불가능한 숨

겨진 형태의 정보를 찾아내어 이를 바탕으로 데이터에 대한 해석 및 의사 결정을 지원하는 웹 문서 마이닝이 필수적인 것이다.

본 연구에서는 웹의 정보를 지식 베이스로 변환하기 위한 웹 마이닝 시스템의 설계와 구조에 대하여 제안하였으며, 실제적인 구현으로 웹 마이닝 시스템을 개발하였다. 제안된 웹 마이닝 시스템은 서치 에이전트를 이용하여 정보를 검색하고 검색된 결과를 유사도 측정을 이용하여 서로 연관된 웹 문서 간의 주요 특징을 추출하여 체계화한다. 개발된 시스템은 뉴스 자동 분류, 지식 관리에서의 인덱스 용어 사전 생성과 같은 도메인에 효과적인 사용이 기대된다.

## 참고문헌

- [1] Chakrabarti S., "Data mining for hypertext: A tutorial survey", SIGKDD Explorations, Vol. 1, Issue 2, pp. 1-11, 2000.
- [2] Chawathe S., Garcia-Molina H., Hammer J., Irland K., Papakonstantinou Y., Ulman J., Widom J., "The Tsimmos project: Integration of heterogeneous information sources", Proceedings IPSJ Conference, Tokyo, 1994.
- [3] Doorenbos R., Etzioni O., Weld D., "A scalable comparison shopping agent for the world wide web". Technical Report 96-01-03, University of Washington, Dept. of Computer Science and Engineering, 1996.
- [4] Fowler, M., Scott K., UML Distilled, 1997, Addison Wesley.
- [5] King R, Novak M., "Supporting information infrastructure for distributed, heterogeneous knowledge discovery", Proceedings of SIGMOD 96 Workshop on Research Issues on Data Mining and Knowledge Discovery, 1996.
- [6] Konopnicki D., Shmueli O., "W3QL: A query system for the world wide web", Proceedings of the 21th VLDB Conference, pp. 54-65, Zurich, 1995.

- 
- [7] Lakshmanan L., Sadri F., Subramanian, I., "A declarative language for querying and restructuring the Web", Proceedings of 6th International Workshop on Research Issues in Data Engineering: Interoperability of Nontraditional Database Systems (RIDE-NDS'96), 1996.
- [8] Lee, R. C., Teppenhart, W., UML & C++, 1997, Prentice Hall.
- [9] Maarek Y., Ben Shaul I., "Automatically organizing bookmarks per content", Proceedings of 5th International World Wide Web Conference, 1996.
- [10] Merialdo P, Atzeni P., Mecca G., "Semistructured and structured data in the Web: Going back and forth", Proceedings of the Workshop on the Management of Semistructured Data, 1997.
- [11] Perkowitz M., Etzioni, O. "Category translation: Learning to understand information on the Internet", Proceedings of 15th International Joint Conference on AI, pp. 930-936, Canada, 1995
- [12] Quass D., Rajaraman A., Sagiv Y., Ullman J., Widom J., "Querying semistructured heterogeneous information", International Conference on Deductive and Object Oriented Databases, 1995.
- [13] Salton, G., Yang, C., and Wong, A., "A vector-space model for automatic indexing", Communications of the ACM, Vol. 18, No. 11, pp. 613- 620, 1975
- [14] Weiss R., Velez B., Sheldon M., Namprempre C., Szilagy P., Duda A., Giord D., "Hypursuit: A hierarchical network search engine that exploits content-link hypertext clustering", Hypertext'96: The Seventh ACM Conference on Hypertext, 1996..
- [15] Zaiane O., Han J., "Resource and knowledge discovery in global information systems: A preliminary design and experiment", Proceedings of the First International Conference on Knowledge Discovery and Data Mining, pp. 331 - 336, 1995.
- [16] <http://www.dt.co.kr>
- [17] <http://www.etimesi.com>

## 저자 소개

**이강찬(chan@etri.re.kr)**

충남대학교 컴퓨터공학과 졸업 (학사, 석사, 박사)

현재 한국전자통신연구원 표준연구센터 표준기반연구팀 선임연구원

관심 분야 : 데이터베이스, 정보통합, XML, 미디어어터, 시맨틱 웹

**민재홍(jhmin@etri.re.kr)**

고려대학교 산업공학과 졸업 (학사)

고려대학교 경영학과 졸업 (석사)

고려대학교 전산학과 수료 (박사)

현재 한국전자통신연구원 표준연구센터 표준기반연구팀장

관심 분야 : 정보 검색, 데이터 마이닝, 시맨틱 웹

**박기식(kipark@etri.re.kr)**

서울대학교 영어교육학과 졸업(학사)

서울대학교 행정학과 졸업 (석사)

충남대학교 행정학과 졸업 (박사)

현재 한국전자통신연구원 표준연구센터장

관심 분야 : 정보통신 표준화, 정보통신 기술 정책, MIS 분야

**임동순(dsyim@hannam.ac.kr)**

한양대학교 산업공학과 졸업 (학사)

한국과학기술원, 산업공학과 졸업 (석사)

Iowa State University, 산업공학과 졸업 (박사)

현재 한남대학교 산업공학과 교수

관심 분야 : 웹 마이닝, 시뮬레이션, FMS, Petri-Net

**우훈식(hswoo@dju.ac.kr)**

한양대학교 산업공학과 졸업 (학사)

Iowa State University, 산업공학과 졸업 (석사)

Iowa State University, 산업공학과 졸업 (박사)

한국전자통신연구원 전자상거래연구부 선임연구원

현재 대전대학교 컴퓨터정보통신공학부 조교수

관심 분야 : 전자상거래, 웹 마이닝, 시맨틱 웹