

# 웹 통합문서의 효율적 생성과 검색을 위한 자동링크지원 시스템의 설계 및 구축

이 원 중<sup>†</sup> · 정 은 재<sup>††</sup> · 주 수 중<sup>†††</sup> · 이 승 용<sup>††††</sup>

## 요 약

분산 컴퓨팅과 웹 서비스 기술의 발달과 함께, 급증하는 인터넷 사용자는 웹 기반의 맞춤형 정보를 편리하게 작성하고 제공받을 수 있는 서비스들을 요구하고 있다. 이를 위해, 본 논문에서는 맞춤형 정보로서 웹 기반의 통합문서를 생성하고, 사용자 요구에 따라 다양한 검색을 지원할 수 있는 자동링크지원 시스템(ALSS: Automatic Linking Support System)을 구축하고자 한다. 본 시스템의 구성은 클라이언트/서버 환경을 기반으로, 서버는 어휘분석, 질의처리 및 통합문서생성 기능들을 제공하는 자동링크엔진과 사전, 이미지 콘텐츠 및 URL.s로 이루어진 데이터베이스를 지원하도록 구축하였다. 클라이언트 측은 서버 측의 자동링크엔진과 데이터베이스를 접근하여 웹 기반의 통합문서를 생성하는 웹 에디터와 검색 서비스를 지원하는 웹 도우미로 구축하였다. 웹 에디터나 웹 도우미 프로그램은 클라이언트 측에 별도의 설치 없이 서버로부터 다운로드하여 실행할 수 있으며, 서버의 실행기능들의 일부를 클라이언트 측에 분산시키므로써 서버의 부하를 감소시켰다. 본 시스템의 구현으로서, 사용자 인터페이스는 JDK 1.3 기반의 SWING을 이용하고, 클라이언트와 서버간의 연동을 위한 자바 RMI 기법을 적용하였으며, SQL Server 7.0을 사용하여 데이터베이스를 구축하였다. 마지막으로 웹 에디터와 웹 도우미에 의해 자동링크엔진과 데이터베이스를 접근하는 과정과 그들의 실행결과를 보였다.

## Design and Implementation of Automatic Linking Support System for Efficient Generating and Retrieving Integrated Documents Based on Web

Won Jung Lee<sup>†</sup> · Eun Jae Jung<sup>††</sup> · Su Chong Joo<sup>†††</sup> · Seung Yong Lee<sup>††††</sup>

## ABSTRACT

With the advent of distributed computing and Web service technologies, lots of users have been requiring services that can conveniently obtain and/or support well-assembled information based on Web. For this reason, we are to construct Automatic Linking Support Systems for generating Web-based integrated information and supporting retrieval information according to user's various requirements. Our system organization is based on client/server system. A server environment consisted of automatic linking engine that can provide lexical analyzing, query processing and integrated document generating functions, and databases that are made of dictionaries, image and URL contents. Also, client environments consisted of Web editor that can generate integrated documents and Web helper that can retrieve them via automatic linking engine and databases. For client's user-friendly interfaces, web editor and helper programs can directly execute by down loading from a server without setup them before inside clients. For reducing server's overheads, parts of server's executing modules are distributed to clients on which they can be executing. As an implementation of our system, we use the JDK 1.3, SWING for user interfaces like Web editor and helper, RMI mechanism for interaction between clients and a server, and SQL server 7.0 for database development, respectively. Finally, we showed the access procedures of automatic document linking engine and databases from Web editor or Web helper, and results appearing on their screens.

키워드 : 자동 링크(Automatic Linkin), 웹 에디터(Web Editor), 통합 문서(Integrated Paper), 웹 도우미(Web Helper)

### 1. 서 론

최근 인터넷 기술을 바탕으로 기하급수적인 이용 증가율을 보인 웹은 가장 광범위한 사용자 계층을 확보하고 있다. 기존의 웹은 텍스트 기반의 구조를 통한 비동기 적인

형태의 서비스 유형으로 문서검색과 단순한 표현 방식을 사용하고 있다. 그러나 컴퓨팅 환경과 웹 기술의 발전으로 인터넷상의 데이터들이 보다 복잡해지고 구조화되면서, 다양한 멀티미디어 정보를 표현하고자 하는 요구가 증대되고 있다[1, 2]. 이러한 요구는 웹의 새로운 구조 및 표현방식을 필요로 한다. 따라서 차세대 웹은 기존의 구조형식을 토대로 새로운 표현 기술과 웹 서비스 기술을 이용하여 확장되고 있다.

기존의 웹 서비스에서 텍스트는 물론, 음성, 영상, 이미

† 정 회 원 : 군산군장대학 교수  
 †† 정 회 원 : 원광대학교 대학원 컴퓨터공학과  
 ††† 정 회 원 : 원광대학교 컴퓨터공학과 교수  
 †††† 종신회원 : 원광보건대학 컴퓨터응용개발과 교수  
 논문접수 : 2002년 6월 17일, 심사완료 : 2003년 3월 14일

지와 같은 멀티미디어 데이터를 동적인 서비스로 제공하기 위해 노력으로 발전을 해왔다[4,5,12]. 이러한 발전은 최근 들어, 인터넷을 사용하는 사용자들의 수가 증가함에 따라 정보를 제공하는 쪽에서의 단순한 웹 서비스가 아닌 사용자 위주의 맞춤 정보 서비스를 제공하는 쪽으로 변화되고 있다. 즉, 기존의 전문 콘텐츠 및 사전(Dictionary) 개발이 CD-ROM 형태에서 인터넷 기술의 발전으로 웹을 기반으로 한 개발에 초점을 맞추어지고 있다. 그러나 기존의 텍스트 문서를 재사용하거나 하이퍼텍스트로 변환시키는데 문서 개발자의 수작업에 의해 이루어진다[13,14]. 또한, 웹 서비스의 대부분이 서버측에서 제공하는 서비스가 주류를 이루고 있어 클라이언트의 수가 증가할 경우, 서버측의 부하로 인하여 처리 시간이 지연된다. 이를 해결하기 위해 본 논문에서는 서버의 부하를 클라이언트 측으로 둔 3계층 구조의 클라이언트/서버 모델을 기반으로 기존의 텍스트 문서에서 연계된 정보를 자동으로 링크시켜 하이퍼텍스트로 변환시켜 웹 정보활용의 편리성을 제공한다. 또한 사용자로 하여금 전문 콘텐츠, 전자 사전 및 관련 웹 사이트 정보를 사용자 요구의 맞춤형 정보로 검색할 수 있는 자동링크지원시스템(ALSS : Automatic Linking Support System)을 제안한다.

자동링크지원 시스템은 웹 기반에서 전문 콘텐츠, 용어사전 및 관련 웹 사이트(URLs)에 대한 실시간 또는 일괄처리 정보검색 서비스를 제공할 뿐 아니라, 작성된 문서상에 관련된 정보를 자동적으로 연결시켜 다양한 하이퍼미디어 정보를 통합문서로 생성한다. 사용자 측면에서는 새로운 하이퍼미디어 문서 작성기술 없이도 다양한 관련정보(전문 콘텐츠, 용어사전 및 관련 URL들)를 자신의 텍스트 문서에 담을 수 있어 웹 문서작성이 용이하다. 또한 상기 관련정보를 웹 브라우저 상에서 실시간으로 검색이 가능하다. 시스템 측면에서는 기존의 웹 문서를 비롯하여, 텍스트 문서의 재활용이 가능하며, 텍스트를 비롯하여, 이미지, 멀티미디어 데이터들을 이용한 복합문서의 활용이 가능하다. 또한 이미 구축된 데이터베이스의 정보를 사용할 수 있어 확장 가능하다. 그리고 기존의 웹 서버들과 상호운용이 가능하다[15].

본 논문의 구성은 2장에서는 제시하는 ALSS의 전체 시스템 구성에 대해 소개한다. 그리고 3장에서는 ALSS의 구성요소들에 대한 수행과정과 설계내용을 기술하고, 4장에서 개발환경 및 ALSS의 실행과정과 생성된 통합 문서의 결과를 화면을 통해 보인다. 끝으로 5장에서는 향후 연구내용에 대해 기술한다.

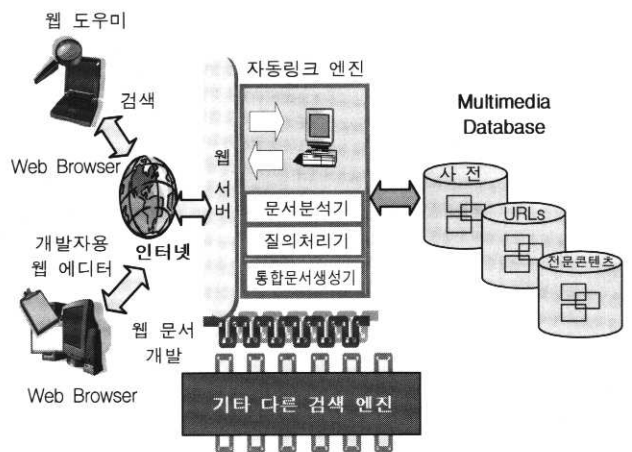
## 2. 자동링크지원 시스템(ALSS : Automatic Linking Support System)

ALSS는 웹 서비스 부분으로, 기존의 문서를 하이퍼텍스트

로 변환하여 연계된 정보를 자동으로 링크시켜 정보 재사용성을 증대시키고, 기존 정보의 재활용 및 통합문서 작성에 용이하다. 본 장에서는 웹 사용자로 하여금 전문 콘텐츠 및 전자 사전 및 관련 웹 사이트에 대한 효율적인 검색과 사용자를 위한 맞춤 서비스를 제공할 수 있는 자동 링크 지원 시스템의 전체 구조와 구성요소의 기능에 대해 기술한다.

### 2.1 ALSS 전체 구조

ALSS의 구성요소는 다음과 같은 기능들 갖는다. 첫째, 필요한 정보를 검색하기 위해 기존의 텍스트 기입형 검색 방법에서 포인팅 디바이스를 해당 단어에 위치시킴에 따라 지정된 단어에 대한 정보를 실시간으로 검색할 수 있도록 한다. 둘째, 사용자가 하이퍼미디어 문서를 작성할 때, 사용자가 작성한 텍스트 문서에 포함된 단어들에 대한 정보(전문 콘텐츠, 용어사전 및 관련 웹 사이트의 URLs)를 데이터베이스로부터 추출하여 텍스트 문서에 자동링크시켜 통합된 하이퍼미디어 문서를 생성할 수 있도록 한다. 마지막으로 통합된 하이퍼미디어 문서를 자유로이 사용자의 의도에 맞도록 재편집 및 실시간 검색을 제공할 수 있도록 한다. 이를 위한 ALSS는 클라이언트/서버 구조로서 인터넷기반에서 수행된다. 서버 측은 전문 콘텐츠, 용어사전 및 관련 웹 사이트(URLs)를 제공하기 위한 데이터베이스 시스템과 웹 문서작성을 웹 에디터와 연결되어 통합된 하이퍼미디어 생성 및 편집을 지원하는 자동링크엔진으로 구성된다. 그리고 클라이언트 측은 웹 에디터와 웹 도우미로서 서버측과 연동하여 하이퍼링크된 통합문서 생성과 검색 서비스를 요청한다. (그림 1)은 ALSS의 전체구조를 나타낸다.



(그림 1) ALSS의 전체구조

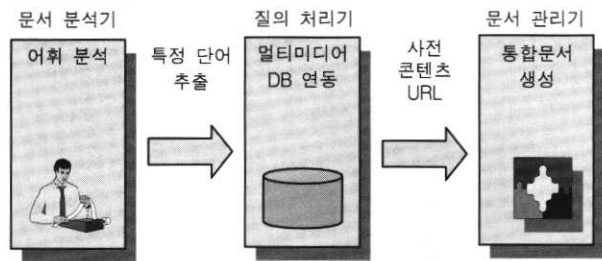
(그림 1)에서 나타난 바와 같이 ALSS의 구성요소는 클라이언트인 웹 에디터, 웹 도우미와 서버인 자동링크엔진, 그리고 데이터베이스 시스템으로 3계층(3-Tier) 구조로 구

성된다. 시스템의 구성요소와 기능들을 다음과 같다.

### 2.2 ALSS의 구성요소와 기능의 설계

웹 에디터는 웹 문서작성 에디터로서 별도의 클라이언트에 제공되는 에디터 프로그램 달리 서버로부터 다운로드하여 사용하는 에디터 프로그램이다. 즉, 웹 에디터는 기존의 에디터 기능 외에 추가로 자동링크엔진 내의 문서분석기와 통합문서 생성기와 연동하여 하이퍼링크된 통합문서를 생성하는 기능과 생성된 통합문서를 재편집기능이 가능하다. 이들 기능은 일반적인 윈도우 프로그램과 같이 기본 문서 작성을 위한 일반 메뉴, 실행 메뉴, 선택 메뉴 그리고 화면 전환 메뉴로 구성된다.

자동링크엔진의 기능은 웹 에디터 부분과 웹 도우미, 그리고 데이터베이스 시스템과 상호작용을 지원하고, 웹 에디터 상에서 문서 작성자가 만든 텍스트 위주의 문서에 데이터베이스의 정보를 하이퍼링크 시키고, 하이퍼미디어 문서로 변환된 하나의 통합문서를 생성시킨다. 또한, 일반 사용자에게는 키워드 형태의 검색 방법에서 마우스 이벤트에 의한 검색으로 인터넷 기반의 정보 검색을 지원한다. 자동링크엔진은 크게 문서분석기, 질의처리기 그리고 통합문서 생성기로 구성된다. 아래 (그림 2)는 자동링크엔진의 각 구성요소와 그들의 기능을 나타낸다.



(그림 2) 자동링크엔진의 구성과 기능

문서분석기는 사용자가 작성한 문서를 분석하여 자동링크엔진에서 데이터베이스와 하이퍼링크 시킬 명사형 단어, 즉 키워드를 추출한다. 영어로 작성된 문서인 경우는 공백으로 분리되는 단위가 바로 링크 할 수 있는 키워드가 되지만, 한글의 경우는 공백으로 구분되어지는 단위가 여러 형태소(명사, 조사 등)로 이루어져 있기 때문에 형태소를 분리하여 자동링크엔진의 처리 단위인 단어로 분리하는 과정이 필요하게 되는데 이를 어휘분석(형태소분석[6-8])이라 한다. 문서분석기에서 이 어휘분석 과정[9-11]을 수행하여 데이터베이스의 인덱스와 매칭하게 될 명사형 단어인 키워드들을 생성한다. 그리고 데이터베이스 시스템과 연동하여 문서 분석기로부터 추출된 키워드에 해당하는 정보를 데이터베이스로부터 얻는 과정을 지원한다. 질의 처리단계에서 추출된 키워드와 데이터베이스의 인덱스 매칭을 통해 해당

된 키워드에 대한 정보를 검색하게 된다.

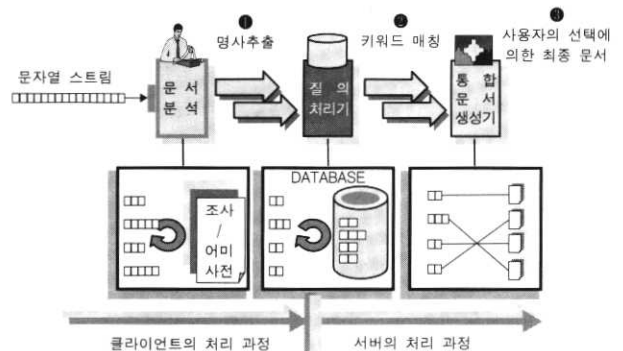
문서분석기에서 추출된 각각의 키워드에 링크된 정보는 질의처리기를 통해 얻어진 데이터베이스의 정보이다. 연결된 정보들 중에 사용자가 선택한 키워드에 해당하는 정보를 하나의 문서로 생성한 문서를 통합문서라 한다. 즉 통합문서 생성기에서는 질의 처리기를 통해서 얻어진 정보를 사용자의 선택에 따라 필요한 다양한 통합문서가 생성된다. 원 문서에서 추출된 명사(키워드들)의 세부 정보를 데이터베이스로부터 획득하여 원 문서에 하이퍼링크 시킨 문서를 생성하고 관리한다. 링크되는 정보는 사전정보, 이미지 콘텐츠 정보 및 웹 사이트의 URL 정보들로서 통합문서는 HTML 문서로 보여진다.

웹 도우미의 기능은 자동링크엔진의 질의 처리기 부분과 상호작용하며, 특히, 일반 사용자의 편의성을 고려한 프로그램으로써 자동링크엔진으로부터 생성된 통합 문서 또는 그 밖의 일반문서(Web browser를 통한 문서, 워드 프로세스상의 작성 문서 등)를 검색하여 도우미 화면에 보일 수 있다. 또한 내용 검색시 용어에 대한 자세한 정보(사진, 콘텐츠, URLs)를 자동링크엔진을 통해 데이터베이스를 연동하여 실시간으로 참조가 가능하다.

데이터베이스는 ALSS의 구성요소들 중에 하나인 질의처리기와 상호작용 한다. 데이터베이스 구조는 전문용어 사전 DB와 이미지 DB, 그리고 URL정보 DB로 구분하여 구축된다. 이들 정보들은 각 DB별로 추가, 삭제, 갱신 등이 가능하다.

### 3. ALSS의 수행과정

ALSS의 전반 수행과정은 (그림 3)과 같이 3단계로 진행된다. 각 단계의 수행을 위해 각 구성요소들은 클라이언트와 서버 측에서 각각 나뉘어 실행하게 된다. 클라이언트 측 은 웹 도우미와 웹 에디터로 이루어지며, 서버 측은 자동링크엔진이 된다. 클라이언트/서버 모델의 관점에서 처리과정을 살펴본다.



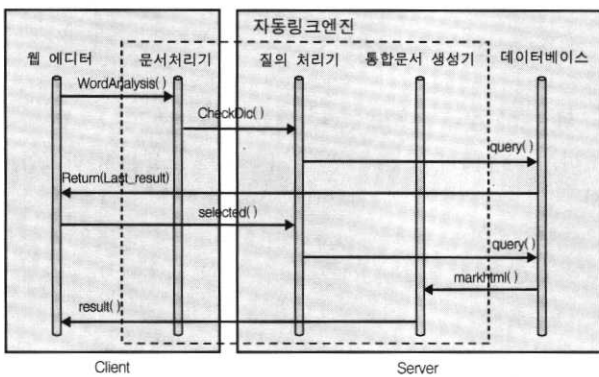
(그림 3) ALSS의 전반 수행과정

3.1 클라이언트 측 처리과정

클라이언트 측에서 웹 에디터의 주요 처리는 문서작성과 명사 추출과정을 들 수 있으며, 웹 도우미 부분은 키워드 추출을 위한 처리 과정이다.

자동링크엔진을 사용하기 위한 전용 웹 에디터로 자바 RMI 기법을 통하여 서버 측에 있는 자동링크엔진의 주요 메소드들을 웹 에디터에서 호출하는 방식으로 처리된다. 그 기능으로는 크게 문서작성, 사용자 정의, 최종문서 확인을 갖는다. 처리 과정은 다음과 같다.

- 웹 브라우저를 통하여 서버에 접속하여 웹 에디터를 실행한다.
- 새로운 문서를 직접 작성하거나 기존의 작성된 문서를 불러온다.
- 작성하거나 읽어 들인 문서를 문서 분석기를 통해 명사 후보군을 추출한다.
- 웹 에디터는 명사 후보군 리스트를 출력하고, 사용자는 각 명사 후보군에 대한 키(key)값을 서버 측에 보내면, 서버 측의 질의처리를 통하여 데이터베이스에서 키값에 해당되는 사전, 이미지, 관련 사이트들의 정보를 웹 에디터에 출력한다.
- 사용자는 웹 에디터를 통하여 필요한 용어를 선택하여, 서버 측의 통합문서 생성기로 보내어 HTML 형식의 통합문서를 되돌려 받는다.
- 최종적으로 통합문서를 확인 후 저장한다.



(그림 4) 웹 에디터에서 통합 문서 생성까지 ETD

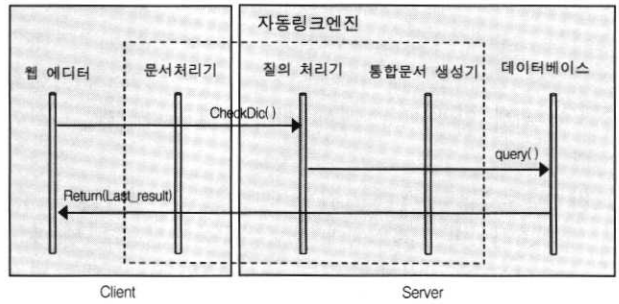
(그림 4)는 클라이언트 측에서 웹 에디터를 다운로드받은 후 웹 에디터에서 통합 문서 생성까지의 ETD를 보이고 있다. 웹 도우미의 처리과정은 웹 에디터의 처리과정에서는 명사 추출 과정 없이 사용자가 마우스 이벤트를 사용하여 키워드를 지정하는 방법 외에는 웹 에디터의 수행과정과 동일하다.

- 웹 문서상의 특정 키워드를 마우스 이벤트에 의해 추출한다.
- 추출된 키워드에 대해 필요로 하는 참조할 정보(사전,

이미지콘텐츠, URLs)를 선택하여 자동링크엔진에 전송한다.

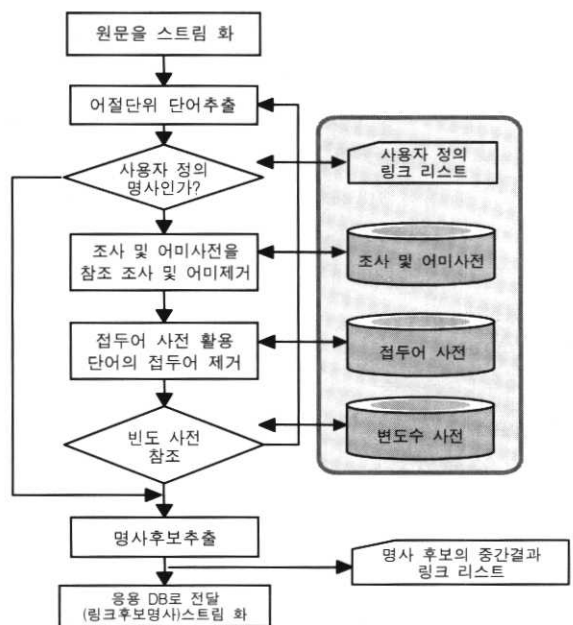
- 키워드에 관련된 검색된 정보를 웹 도우미 창에 출력한다.

(그림 5)는 웹 도우미에서 키워드 검색을 하여 최종 결과를 보이기까지의 ETD를 보이고 있다.



(그림 5) 웹 도우미의 ETD

웹 에디터에서 새로운 문서를 작성하거나 불러오기를 통한 텍스트 문서로부터 해당 키워드인 명사를 추출하기 위한 흐름도를 (그림 6)에서 나타낸다. 흐름도를 간단하게 설명하면, 웹 클라이언트로부터 입력된 문자 스트림(문서)을 공백 단위의 어절을 추출한다. 클래스 패키지 형태의 사전(조사, 어미, 접두어, 감탄사, 자주 사용되는 부사, 형용사, 동사)을 이용하여 어절에서 명사 링크후보가 될 수 있는 키워드들을 추출한다. 이 때, 중복된 명사들과 특수 문자들은 제거한다. 마지막으로 웹 에디터의 사용자 정의 부분에서 링크를 원하는 키워드들의 후보 명사들을 서버 측으로 전송한다.



(그림 6) 키워드 명사의 추출과정

### 3.2 서버 측 처리 과정

서버 측 처리는 웹 에디터에서 작성한 문서를 문자열 스트림으로 받아 데이터베이스와의 연동을 통해 데이터베이스 인덱스 키 값과 원 문서에서 추출된 키워드 명사들과의 매칭하는 질의처리과정과 매칭된 정보를 원 문서에 하이퍼링크를 통해 묶고 통합문서를 생성하는 과정을 수행한다. 웹 에디터의 지원을 위한 자세한 서버 측의 수행과정은 다음과 같이 이루어진다. 즉, 서버 측의 자동링크엔진에서는 클라이언트 측으로부터 추출된 키워드 명사들의 링크후보와 사용자가 재 정의한 명사를 스트림 형태로 받아 JDBC를 통한 데이터베이스의 키매칭을 통해 매칭된 결과 값을 스트림 형태로 접속한 클라이언트에게 반환되며, 반환된 링크결과를 통합문서 형태(태그 삽입된 형태)로 반환된다. 또한 웹 도우미에서 요청한 키워드에 대한 관련 정보 매핑 및 제공을 위하여 서버측의 과정은 웹 에디터 상에서 실행하는 과정 중 명사 추출과정이 필요 없고, 단순히 키워드에 대한 매칭과정만 이루어지고 매칭된 정보는 클라이언트 측의 웹 도우미에게 반환되도록 한다.

### 3.3 클라이언트/서버기반의 ALSS 환경

클라이언트 상에서 웹 에디터와 웹 도우미 실행 프로그램을 설치하여 실행하는 경우, 클라이언트 시스템의 메모리의 확장은 물론, 실행 프로그램의 기능 추가로 인한 버전업 관리의 어려움이나 재 설치와 같은 번거로움의 문제점들을 가진다. 이를 해결하기 위해서 클라이언트용 실행 프로그램들을 자동링크지원 서버 시스템에 설치하고 필요시 클라이언트에서 실행환경을 내려받기 하여 웹 에디터와 웹 도우미를 실행할 수 있는 환경을 구축한다.

ALSS의 자동링크엔진이 있는 서버 측에서 원 문서로부터 키워드 명사의 추출, 추출된 키워드와 데이터베이스의 인덱스와의 매칭, 그리고 통합문서를 생성하기 위해 원 문서 내에 해당 키워드 명사들에 매칭된 데이터베이스 정보들을 자동으로 하이퍼링크 시키는 처리 과정들을 실행된다. 이러한 실행은 서버 측에 과중한 부하를 자동링크엔진을 접속하는 클라이언트들의 요청 수를 제한하는 요인이 된다. 따라서 서버의 부하를 클라이언트 측에 전가할 수 있는 환경을 구축하였다. 자동링크지원 시스템의 부하를 감소시키기 위해서 서버 측의 수행과정들을 클라이언트와 서버 환경에서 분산 수행한다. 클라이언트 측에서는 키워드(명사)추출과정을 수행하여 추출된 키워드(명사)와 사용자가 지정한 매칭 정보(사전, 컨텐츠, URL)를 서버 측의 자동링크엔진에 전송한다. 서버 측에서는 이들 정보를 받아 데이터베이스로부터 해당 키워드에 대한 정보를 매칭시키고, 관련된 정보(키워드+데이터베이스 인덱스)를 다시 클라이언트에게 전송한다. 클라이언트에서 서버로부터 받은 매핑된 정보를 기존문서상에

자동으로 하이퍼링크 시켜 통합문서를 생성한다. 위와 같이 자동링크지원 서버는 키워드(명사)추출과정과 통합문서 실행을 클라이언트 측에 부하를 위탁시키므로써, 서버는 줄어든 부하에 대한 잉여처리여력만큼 더 많은 클라이언트들의 요청을 지원 할 수 있도록 하였다.

## 4. ALSS의 개발환경 및 구현

### 4.1 개발환경

ALSS 환경은 <표 1>과 같이 3계층 구조의 클라이언트/서버 모델을 기반으로 구축했다. 구현환경은 Window NT 4.0 서버에서 운영하고, 웹 서버는 MS사의 IIS(Internet Information Server 4.0)을 이용하며, 데이터베이스는 SQL Server 7.0로 구축했다. 또한 사용자 인터페이스 부분은 JDK 1.3버전의 SWING 컴포넌트(Component)를 사용하여 구현하였다. 서버 측과 클라이언트 측의 연동 부분은 Java RMI 방식을 택하였다.

<표 1> 개발 환경

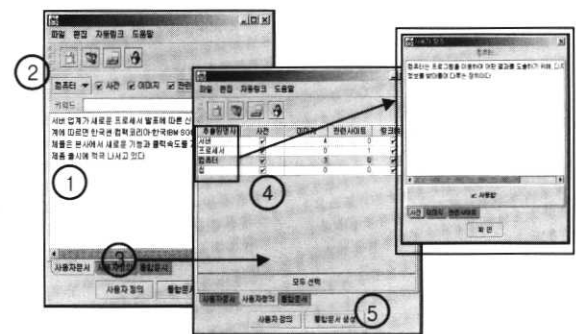
구분	사용 환경	비 고
운영체제	Window NT Server 4.0	IE4.0 이상
DB 엔진	M/S SQL Server 7.0	데이터베이스
웹 서버	IIS 4.0(Internet Information Server)	자동링크엔진
개발도구	Forte Java 1.0	자동링크엔진
개발언어	JAVA	문서 관리기, 웹도우미

### 4.2 ALSS의 구현 및 실행화면

ALSS의 클라이언트 측의 웹 에디터와 웹 도우미는 서버 측의 부하를 줄이기 위해 애플릿 형태로 구현하였으며, 특히 웹 에디터는 어휘분석기 부분을 포함하고 있다. 본 절에서는 시스템을 이루는 각 구성요소의 인터페이스 부분을 위주로 실행 화면을 보인다.

#### 4.2.1 자동링크엔진 지원 웹 에디터 구현 및 실행

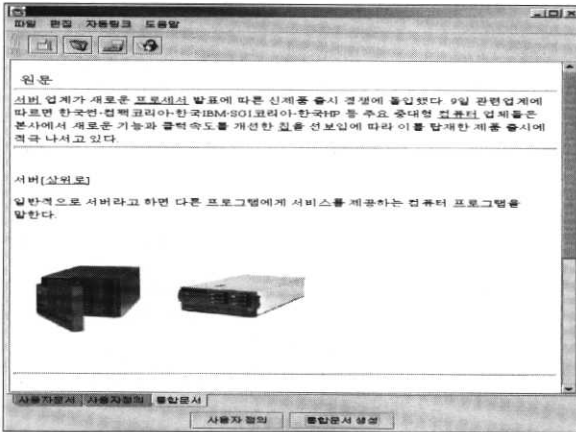
웹 에디터는 서버 측에서 자바 애플릿 형태로 제공한다. Java2의 강력한 컴포넌트인 SWING을 사용하여 사용자 인터



(그림 7) 웹 에디터 입력화면

터페이스를 구성했으며, Java RMI 기법을 이용하여 서버 측 ALS 2의 메소드를 호출하여 사용할 수 있도록 하였다. 다음 (그림 7)은 웹 에디터 실행화면이다.

웹 에디터를 사용하여 필요한 정보를 자동링크하여 통합 문서를 생성하기 위해서는 총 5단계를 거치게 된다. 우선 1 단계로 기존에 만들어 놓은 텍스트 형식의 문서를 읽거나 새로운 문서를 웹 에디터 상에서 작성한다. 문서 읽기 또는 문서 작성이 끝나면 2번째 단계로 자신이 원하는 키워드를 넣거나 콘텐츠(정보)를 선택한다. 3단계로 '사용자정의' 버튼을 클릭하면 내부적으로 어휘 분석 작업이 이루어지며 그 결과로 키워드가 추출되고 만약에 2단계에서 문서에 필요한 키워드를 직접 선택한 경우 키워드에 해당하는 정보만 사용자 정의 화면에 출력된다. 4단계에서는 키워드에 해당하는 정보, 즉 사전, 이미지, 관련 사이트에 대한 원하는 정보를 선택하게 된다. 각 키워드에 해당하는 정보를 반복적으로 모두 선택한 후 5단계인 '통합문서생성' 버튼을 클릭하면 사용자가 정의한 내용으로 정보들이 통합문서 생성된 후 화면에 출력된다. 다음 (그림 8)은 통합문서 화면을 나타낸다.



(그림 8) 통합 문서 생성 화면

4.2.2 자동링크엔진

자동링크엔진은 많은 클라이언트의 접근에 대한 서버 측 부하를 고려하여, 원문(사용자 작성 웹 문서)의 최소한의 명사 후보 링크어절을 추출하여, 데이터베이스에 등록된 키워드를 비교/매칭함으로써 사용자가 원하는 전문용어에 대한 정보를 연결한다. 명사후보의 추출은 다중 사용자를 고려하여, 웹 클라이언트에 경량화된 사전(조사, 접두사, 접미사)을 전송하여 링크될 명사 후보의 1차 추출이 수행되고, 수행의 결과로서 생성된 명사 후보를 사용자가 정의하여 자동링크엔진에게 스트림 형태로 전송한다. 이를 데이터베이스의 키워드와 매칭하여 통합문서 형태로 클라이언트에 전송한다.

```
public class AlsObjectImpl extends UnicastRemoteObject
    implements AlsObject {
    .....
    public String WordAnalysis (String UserDocument) {
    .....
    try { // 받은 문자열 토큰처리(공백으로)
        eomi = new Eomi( ); // 어미
        josa = new Josa( ); // 조사
        jubmisa = new Jubmisa( ); // 접미사
        jubdusa = new Jubdusa( ); // 접두사
        .....
        for (int i = 0; i < josa_string.length; i++) { // 조사처리 시작
            .....
            if (word_length >= char_num ) {
                eogan_start = word_length - char_num;
                word_eomi = new String( word.substring
                    (eogan_start, word_length ));
                if (word_eomi.equals( josa_string [i] )) {
                    if (Rplica_Tag.equals("NO")) {
                        word_eogan = new String(word.substring(0, eogan_start));
                        Noun_string = word_eogan;
                        Rplica_Tag = "YES";
                    } else {word_eogan = new String( word.substring
                        (0, eogan_start));
                        if (Noun_string.length( ) > word_eogan.length( )){
                            Noun_string = word_eogan; }
                    } ... } ... } ...
                }
            }
        }
    }
}
```

(그림 9) 명사 추출을 위한 조사 처리 부분

(그림 9)는 자동링크엔진 부분중 어휘분석기 부분으로 웹 에디터로부터 사용자가 작성한 문서를 스트링으로 넘겨받아 토큰처리 후 어휘분석을 하는 단계로, 어미, 조사, 접미사, 접두사를 생성한 후 그 중에서 조사를 비교 분석하기 위한 코드이다.

(그림 10)은 자동링크엔진 부분중 질의 처리기 부분으로 어휘분석기에서 추출된 명사와 서버 측의 데이터베이스의 키워드를 비교하기 위한 코드이다.

```
private void ConnectToDB( ) throws SQLException {
    try {
        Class.forName ("sun.jdbc.odbc.JdbcOdbcDriver");
        .....
        Connection conn = java.sql.DriverManager.getConnection
            ("jdbc : odbc : woon, " "sa, " " ");
        Statement stmt = conn.createStatement( );
        .....
        String sqlString = "select keyword from db_computer where
            keyword = key";
        ResultSet rs = stmt.executeQuery (sqlString);
        .....
    }
}
```

(그림 10) 키워드 매칭 부분

4.2.3 자동링크 엔진 및 웹 도우미 실행 결과

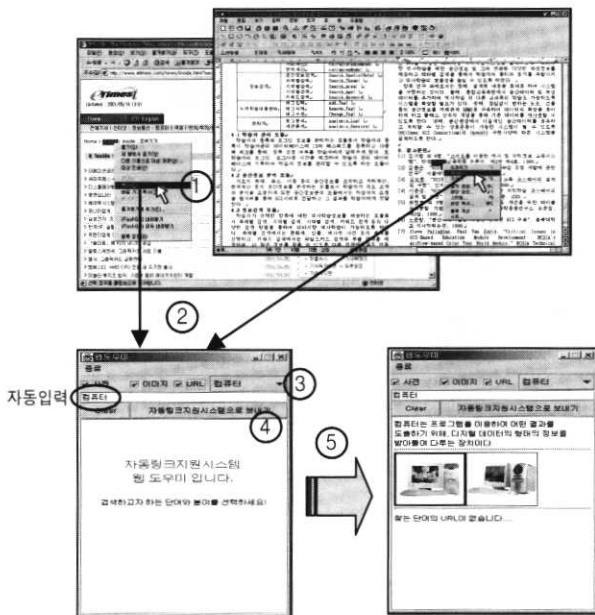
웹 도우미를 웹 검색이나 문서 작성 중에 사용하기 위해 위 그림의 다섯 단계를 거치게 된다. 1단계는 문서 내에서 웹 도우미를 이용하고자 하는 단어를 복사한다. 2단계에서 웹 도우미 GUI에 마우스 커서를 이동시키면 클립보드의 단어가 자동으로 웹 도우미 입력창에 표시된다. 3단계에서 검색하고자하는 분류 및 분야를 선택하고, 4단계에서 자동링크지원 시스템으로 정보를 스트링 형태로 전송한다. 5단계에서는 4단계에서 전송한 정보에 대한 결과를 스트링으로 수신받아 출력한다. 정보의 전송과 수신은 RMI 인터페이스를 통해 이루어지며 사용자는 자동링크지원 시스템에 메소드와 스트링 정보만을 전송하며 원하는 검색 작업을 할 수 있다. 또한, 사용자는 웹 도우미에 검색하고자하는 단어를 직접 입력하여 각각의 단계를 거치며 원하는 정보를 획득할 수 있다.



(그림 12) 데이터베이스

4.2.5 통합문서 생성의 결과 화면

(그림 13)은 웹 에디터 상에서 작성한 문서와 데이터베이스 내용이 하이퍼링크된 통합문서의 생성결과를 보인 결과 화면이다.

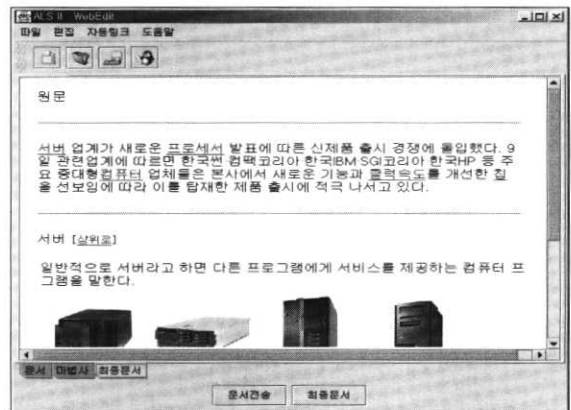


(그림 11) 웹 도우미 사용을 위한 단어 자동입력 사용화면

(그림 11)에서 나타난 바와 같이 웹 에디터의 최종 결과 화면은 에디터 상의 최종 문서 탭에서 확인한 화면으로 웹 도우미의 단일 키워드에 대한 결과화면과 유사하다. 차이점은 전자의 경우, 여러 키워드에 대한 각각의 콘텐츠를 순차적으로 보여지며, 후자의 경우는 하나의 키워드에 대해서만 사용자가 지정한 콘텐츠만을 보인다.

4.2.4 데이터베이스

데이터베이스는 사진 정보에 관련된 DB(컴퓨터, 전기, 전자)와 이미지에 관련된 DB, 그리고 URL에 관련된 DB로 구축하였다. 다음 (그림 12)는 SQL Server 7.0으로 구축된 각 데이터베이스의 화면을 나타낸다.



(그림 13) 통합 문서 결과 화면

5. 결론

본 논문의 자동링크지원 시스템(ALSS)은 맞춤형 정보로서 웹 기반의 통합문서를 생성하고, 사용자 요구에 따라 다양한 정보 검색을 지원한다. 본 시스템의 구성은 클라이언트/서버 환경을 기반으로, 서버는 문서분석, 질의처리 및 통합문서생성 기능들을 제공하는 자동링크엔진과 사전, 이미지 콘텐츠 및 URLs로 이루어진 데이터베이스를 구축하였다. 클라이언트 측은 서버 측의 자동링크엔진과 데이터베이스를 접근하여 웹 기반의 통합문서를 생성하는 웹 에디터와 검색 서비스를 지원하는 웹 도우미로 구축하였다. 웹 에디터나 웹 도우미 프로그램은 클라이언트 측에 별도의 설치 없이 서버로부터 다운로드하여 실행할 수 있으며, 서버의 실행기능들의 일부를 클라이언트 측에 분산시키므로써 서버의 부하를 감소시켰다. 본 시스템의 구현으로서, 사용자 인터페이스는 JDK 1.3 기반의 SWING을 이용하고, 클라이언트와 서버간의 연동을 위한 자바 RMI 기법을 적용하였으며, SQL

Server 7.0을 사용하여 데이터베이스를 구축하였다.

추후 연구 내용으로는 문서 분석기에서 처리되는 키워드 추출부분은 전문 용어 사전에서 사용되는 출현 빈도 수만을 따르고 있기 때문에, 한국어 특성을 반영할 수 있는 표준화된 사전이 필요로 하다. 또한, 다양한 멀티미디어 정보를 제공하기 위해 멀티미디어 데이터의 표현 방법 및 효율적인 정보 관리에 대한 연구가 필요로 하다. 그리고 현재 구축된 자동 링크 지원 시스템의 플랫폼을 객체지향 미들웨어를 이용하여 확장할 예정이다.

### 참 고 문 헌

- [1] 신명기, 김용진, "W3C에서의 차세대 웹 표준 활동 동향", 정보처리학회지, 제6권 제3호, pp.7-17, 1999.
- [2] 김태석, 박철제, 임환섭, "차세대 웹 서비스 기술 개발 동향", 정보처리학회지, 제6권 제3호, pp.18-24, 1999.
- [3] 나홍석, 채진석, 김창화, 백두권, "차세대 웹 상에서의 문서 교환 및 검색을 위한 프레임워크", 정보처리학회지, 제6권 제3호, pp.52-61, 1999.
- [4] 김홍배, 류근호, "HTML 기반 용어/약자 하이퍼링크 생성기", 산·한·연 멀티미디어 산업기술 학술대회 논문집, 1999.
- [5] 김홍식, 권기태, "멀티미디어 저작도와 데이터베이스를 이용한 웹 기반 형성평가 방안에 관한 연구", 한국컴퓨터교육학회논문지, 제2권 제4호, 1998.
- [6] 효율적인 한국어 형태소분석 방법, <http://lactt.taegu.ac.kr/data/internet/indexing/efficient.htm>.
- [7] 음절 특성을 이용한 한국어 불규칙 활용, <http://lactt.taegu.ac.kr/data/internet/indexing/illegal.htm>.
- [8] G. Riessen, H. A. Jacobsen & O. Gunther, MMM, Web Engineering '99 International Workshop, 1999.
- [9] 김영택, "자연언어처리", 교학사, 1994.
- [10] 배정일, "하이퍼텍스트 자동 생성에 관한 연구", 석사학위논문, 1996.
- [11] 김문석, "통계적 유사도와 의미적 유사도를 이용한 하이퍼텍스트 자동생성에 관한 연구", 석사학위논문, 1996.
- [12] Allan, "Automatic HyperText Construction," A Dissertation Presented to the Graduate School of Cornell University, January, 1995.
- [13] 김운미, 정창원, 성재석, 주수중, "웹 통합문서의 저작과 검색을 지원하는 자동링크 지원 시스템의 구축", 정보처리학회 춘계 학술발표논문집, 제7권 제1호, 2000.
- [14] 전병택, 정창원, 이원중, 성재석, 주수중, "웹 기반에서 하이퍼미디어 저작 및 검색이 용이한 자동링크지원 시스템의 개발", 한국인터넷정보과학회 추계 학술발표논문집, 제1권 제2호, pp.217-221, 2000.
- [15] 중소기업기술 혁신 개발 사업, "웹 기반에서 하이퍼미디어 저작 및 검색이 용이한 자동 링크 지원 시스템의 개발", 최종 보고서, 원광대학교, 2001.



### 이 원 중

e-mail : wjlee@kunjang.ac.kr

1991년 전북대학교 전산·통계학과(학사)

1998년 원광대학교 대학원 컴퓨터공학과(석사)

2000년 원광대학교 대학원 컴퓨터공학과 박사과정 수료

2001년~현재 군장대학 전임교수

2001년~현재 전산정보센터장

관심분야 : 분산객체모델, 멀티미디어데이터베이스, 지리정보시스템(GIS)



### 정 은 재

e-mail : hipster@wonkwang.ac.kr

2003년 원광대학교 컴퓨터, 정보통신공학부(학사)

2003년~현재 원광대학교 대학원 컴퓨터공학과 석사과정

관심분야 : 분산 컴퓨팅, 웹 데이터베이스, 모바일 데이터베이스, 실시간 객체 모델, 실시간 시뮬레이션



### 주 수 중

e-mail : scjoo@wonkwang.ac.kr

1986년 원광대학교 전자계산공학과(학사)

1988년 중앙대학교 대학원 컴퓨터공학과(석사)

1992년 중앙대학교 대학원 컴퓨터공학과(박사)

1993년~1994년 미국 Univ. of Massachusetts at Amherst, 전기 및 컴퓨터공학과, Post-Doc.

1990년~현재 원광대학교 컴퓨터 공학과 교수

2002년~현재 미국 Univ. of California at Irvine 전기공학 및 컴퓨터과학과 연구교수

관심분야 : 분산실시간 컴퓨팅, 분산 객체 모델, 시스템 최적화, 멀티미디어 데이터베이스



### 이 승 용

e-mail : dragon@wkhc.ac.kr

1986년 원광대학교 졸업(학사)

1990년 조선대학교 대학원 전산기공학과(석사)

2002년 원광대학교 대학원 컴퓨터공학과(박사)

1991년~현재 원광보건대학 컴퓨터응용개발과 교수

관심분야 : 멀티 데이터베이스, 분산시스템, 데이터통신, 인터넷 통신