

웹 통합문서의 저작과 검색을 지원하는

자동링크지원 시스템의 구축

김운미*, 정창원*, 성재석**, 주수종*

*원광대학교 컴퓨터공학과

** (주)정성 C.S.D

e-mail: woonmi@gaebyok.wonkwang.ac.kr

The Construction of The Automatic Linking Support System(ALS II) for Authoring and Retrieving Web Integrated Documents

Woon-mi Kim*, Chang-Won Jeong*, Su-Chong Joo*
Jae-Seok Seong**

*Dept. of Computer Engineering, WonKwang Univ.

** Jeong Seong C.S.D co., LTD

요 약

인터넷 기술의 발전으로 인하여 다양한 멀티미디어 서비스를 제공하고 있으나, 최근 들어 사용자들의 수가 증가함에 따라 정보를 제공하는 쪽에서의 단순한 서비스가 아닌 사용자 위주의 맞춤 정보 서비스를 제공하는 쪽으로 변화되고 있는 추세이다. 따라서, 본 논문에서는 차세대 웹 서비스에 초점을 맞추어, 기존의 키워드 검색방법과 수 작업에 의한 하이퍼링크 기법의 제한점을 해결한 사용자 위주의 맞춤 정보 서비스를 제공하는 자동 링크 지원 시스템(ALS II)을 제안한다. 이는 마우스 이벤트에 의한 쉬운 검색 방법과 기존 문서상의 키워드에 관련된 다양한 멀티미디어 정보를 자동으로 연결시켜 하이퍼미디어 웹 문서 작성에 편리하며, 새로운 멀티미디어 통합 정보를 구축하는데 용이하다.

1. 서론

최근 인터넷 기술을 바탕으로 기하급수적인 이용 증가율을 보인 웹은 가장 광범위한 사용자 계층을 확보하고 있으며, 기존의 다양한 컴퓨팅 기술과의 접목이 시도되고 있는 추세이다. 기존의 웹은 텍스트 기반의 구조를 통한 비동기적인 형태의 서비스 유형으로 문서검색과 단순한 표현 방식을 사용하였으나 컴퓨팅 환경과 웹 기술의 발전으로 인터넷상의 데이터들이 보다 복잡해지고 구조화되면서, 다양한 멀티미디어 정보를 동기화하여 표현하고자 하는 요구가 나타나게 되었다[1,2]. 이러한 요구는 웹의 새로운 구조 및 표현방식을 필요로 한다. 따라서 차세대 웹은 기존의 구조형식을 토대로 새로운 표현 기술과 웹서비스 기술을 이용하여 확장하고, 이를 통한 통합적인 웹서비스 기술 개발로 변화됨을 예측하게 한다.

이러한 변화는 최근 들어, 인터넷을 사용하는 사용자들의 수가 증가함에 따라 정보를 제공하는 쪽에서의 단순한 웹서비스가 아닌 사용자 위주의 맞춤 정보 서비스를 제공

하는 쪽으로 변화되고 있다.

그러나, 웹 상의 문서나 멀티미디어 정보를 효율적으로 검색하거나 서비스하기에는 한계가 있으며, 이를 해결하기 위해서는 사람의 도움이 많이 필요로 하다[3].

따라서, 본 논문은 차세대 웹 서비스에 초점을 맞추고, 기존의 문서상의 관련된 정보를 자동으로 링크시켜, 하이퍼미디어 정보로 변환시킴으로 저작하는데 편리성을 제공하도록 한다. 또한 일반 사용자로 하여금 효율적인 검색과 사용자 요구에 따라 멀티미디어 정보를 제공하는 맞춤 서비스가 가능한 자동 링크 지원 시스템(ALS II)을 개발하고자 한다.

본 논문의 구성은 2장에서는 ALS II에 대해 기술한다. 그리고 3장에서는 구현 부분으로 ALS II의 프로토타입의 주요화면을 보인다. 4장에서 관련 연구 및 기존의 시스템과 비교하고, 끝으로 5장에서는 향후 연구내용에 대해서 기술한다.

2. 자동 링크 지원 시스템(ALSII)

을 요약하였다.

ALSII는 웹 상에서 CD-ROM형태와 같은 동기화된 형식으로 재생하거나, 전문 콘텐츠, 용어사전 및 관련 웹사이트(URLs)에 대한 정보를 실시간 또는 일괄처리 정보검색 서비스를 제공한다. 또한 작성된 문서상에 관련된 정보를 자동으로 연결시켜, 다양한 하이퍼미디어 정보를 제공할 수 있는 통합 문서 생성을 지원한다. 다음은 본 논문에서 제시하는 시스템의 기능 및 구성요소에 대해 기술한다.

2.1 기능

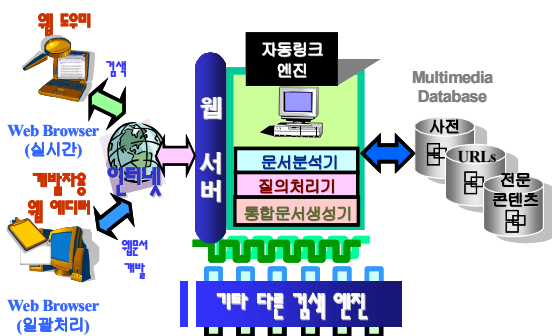
기능을 크게 3가지로 나누어 보면 다음과 같다.

첫째, 필요한 정보를 찾기 위해 텍스트 기입형 검색 방법에서 웹 브라우저 상에서 마우스의 커서를 해당 단어에 위치시킴에 따라 지정된 단어에 대한 정보를 실시간으로 검색할 수 있도록 한다.

둘째, 사용자가 하이퍼미디어 문서를 작성할 때, 수 작업을 통해서 하이퍼미디어 자료들을 링크 시켜왔다. 이러한 번거로움을 개선하도록 사용자가 작성한 텍스트 문서에 포함된 단어들에 대한 정보(전문 콘텐츠, 용어사전 및 관련 웹사이트의 URLs)를 데이터베이스 시스템으로부터 추출하여 자동으로 작성한 문서에 하이퍼링크 시켜 통합된 하이퍼미디어 문서를 생성할 수 있도록 한다.

마지막으로 통합된 하이퍼미디어 문서를 자유로이 사용자의 의도에 맞도록 재편집기능을 제공한다.

다음 (그림 1)은 ALSII의 전체 구성도 이다.



(그림 1) 전체 시스템 구조

2.2 구성 요소

시스템의 구성요소는 (그림 1)에서 나타난 바와 같이 전문 콘텐츠, 용어사전 및 관련 웹사이트(URLs)를 제공하기 위한 멀티미디어 데이터베이스 시스템, 웹 문서 작성을 위한 웹 에디터와 연결되어 통합된 하이퍼미디어 생성 및 편집을 지원하는 자동링크 엔진, 그리고 웹 브라우저와 같은 기능 외에 마우스 이벤트를 통해 실시간으로 검색기능을 제공할 수 있는 웹 도우미로 구성된다. 특히, 웹 도우미나 웹 에디터는 클라이언트 시스템에서 수행되고, 멀티미디어 데이터베이스 시스템과 자동링크 엔진은 서버 시스템 상에서 여러 클라이언트들을 지원한다.

다음 (표 1)은 본 시스템을 구성하는 각 요소의 기능들

구 성 요 소	기 능
웹 에디터	웹 문서 작성을 위하여 서버 측에서 제공하는 형태의 에디터(Web editor)
자동링크 엔진	웹 서버와 상호작용을 하며, 에디터를 통하여 작성된 문서의 문장을 해석하고, 데이터베이스와 연동을 통하여, 전문용어 DB(컴퓨터, 전자, 정보 통신)의 정보를 자동으로 연결시키는 기능 및 생성된 통합 문서를 작성자에게 전송하는 기능
멀티미디어 데이터베이스	분류된 정보별 자료를 데이터베이스화하여 자료의 삽입, 삭제 갱신 기능
웹 도우미	일반 사용자에게 마우스 이벤트에 의한 정보 검색

(표 1) 구성 요소의 기능

다음은 이에 대한 세부적인 구성 요소의 기능에 대해 기술한다.

2.2.1 웹 에디터 기능

웹 에디터 부분은 문서 작성하는 사용자에게 편리한 인터페이스를 제공한다. 별도의 에디터를 사용하지 않고, 서버에서 제공하는 형태로 자동링크 엔진과 상호작용한다.

- ◆멀티미디어 정보를 포함한 멀티미디어 데이터베이스와 연동 기능
- ◆사용자 인터페이스를 갖는 에디터
 - ▶입력 형태에 따른 구분 : 작성할 문서의 텍스트와 HTML 사용
 - ▶출력 형태에 따른 구분 : 이미지와 텍스트 사용, 텍스트 또는 이미지만 사용
 - ▶에디터 박스
 - ▶에디터 박스의 내용을 클리어하는 부분
 - ▶파일 덧붙이기 위한 부분 : 기존의 작성된 텍스트 또는 HTML문서를 에디터상에 붙일 수 있는 기능
 - ▶미리보기 : 작성된 문서를 미리보는 기능(DB와 연동하여 관련된 키워드 부분에 연결된 상태)
 - ▶자동 링크 지원 시스템과 상호작용 부분 : 원하는 형태의 문서를 생성시키기 위해 전송

2.2.2 자동링크 엔진

웹 에디터 부분과 웹 도우미와 상호작용하며, 문서 작성자에게는 텍스트 위주의 문서를 자동으로 하이퍼미디어 문서를 작성케 하며, 생성된 문서에 관련된 정보를 획득하게 한다. 또한, 일반 사용자에게는 키워드 형태의 검색방법에서 마우스 이벤트에 의해 인터넷상의 다양한 정보에 접근하는데 보다 쉬운 검색 방법을 지원한다.

◆ 문서분석기

문서분석기는 사용자가 작성한 문서를 분석하여 자동링크 엔진에서 연결할 수 있는 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 자동링크 엔진의 처리 단위인 단어로 분리하는 과정이 필요하게 되는데 이를 어휘분석(형태소분석[4,5])이라 한다. 문서분석기에서 이 어휘분석 과정을 처리한다. 또한 어휘분석 전에 HTML문서가 입력된 경우 태그를 제거한 후 어휘분석을 한다.

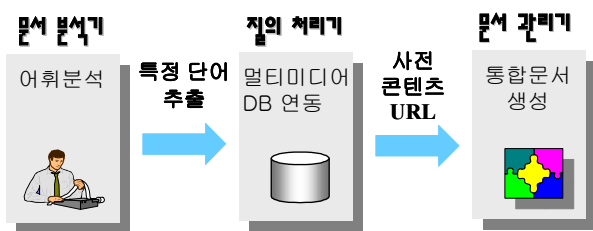
◆ 질의 처리기

추출된 키워드에 대해 멀티미디어 데이터베이스와 연동하여 관련 정보를 획득하는 과정으로 추출된 키워드에 ASP형태의 질의 처리기를 링크 시켜, 키워드를 클릭하였을 경우, 이에 해당되는 정보(텍스트, 멀티미디어, 관련 링크 등)를 제공한다.

◆ 통합문서 생성기

문서분석기에서 추출된 키워드에 링크가 만들어지고, 링크를 클릭하는 경우, 나타나게 되는 관련된 문서들의 집합을 통합문서라 한다. 이 통합문서는 텍스트 정보뿐만 아니라 관련 이미지, 사운드 등의 멀티미디어 정보를 포함하는 HTML문서를 말하며, 통합문서 생성기에서는 질의 처리기를 통해서 얻어진 정보를 가공하여 통합문서를 생성한다.

다음 (그림 2)는 자동링크 엔진을 통하여 통합 문서 생성하는 과정을 나타낸다.



(그림 2) 통합문서 생성과정

2.2.3 멀티미디어 데이터베이스

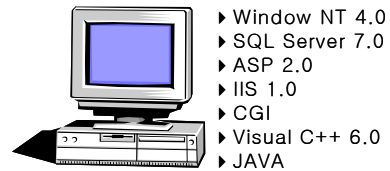
데이터베이스는 프로토타입으로 개발하며, 자동 링크 지원 시스템 구성 요소인 질의 처리기와 상호작용한다. 데이터베이스 구조는 방대한 사전 정보에 관련된 DB와 이미지에 관련된 DB, 그리고 URL에 관련된 DB로 구분하였다. 그리고 관련 분야에 따라 컴퓨터, 전자, 정보통신으로 분류하여 각각의 DB별 콘텐츠를 추가/삭제/갱신 가능하도록 한다. 또한, 그 밖의 다른 데이터베이스의 내용을 활용하여 확장 가능하다.

2.2.4 웹 도우미 프로그램

웹 도우미는 자동링크 엔진과 데이터베이스와 상호작용 [6]하며, 특히, 일반 사용자의 편리성을 고려한 프로그램으로써 자동링크 엔진으로부터 생성된 통합 문서 또는 그 밖의 문서(Web browser를 통한 문서, 워드 프로세스상의 작성 문서 포함)를 읽다가 필요한 용어에 대한 정보 또는 이와 관련된 그 밖의 멀티미디어 정보를 얻고자 할 때 사용한다.

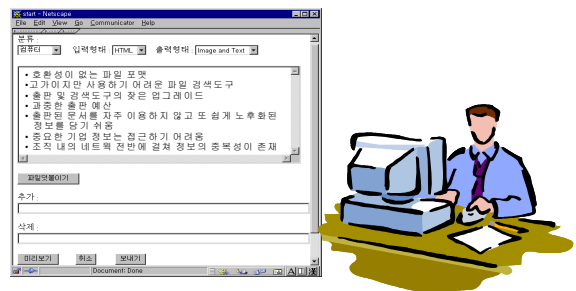
3. 구현

구현 환경은 다음 (그림 3)과 같이 Windows NT 서버에서 운영하고, 웹서버는 MS사의 IIS(Internet Information Server)를 이용하며, 데이터베이스는 SQL Server 7.0으로 구축한다. 자동링크 엔진 부분은 CGI 프로그램을 사용하며, 문서 관리기 부분과 웹 도우미 부분은 Visual C++ 6.0과 JAVA를 이용한 프로그램으로 구현한다. 웹 에디터 부분은 ASP 프로그램으로 구현한다.



(그림 3) 구현 환경

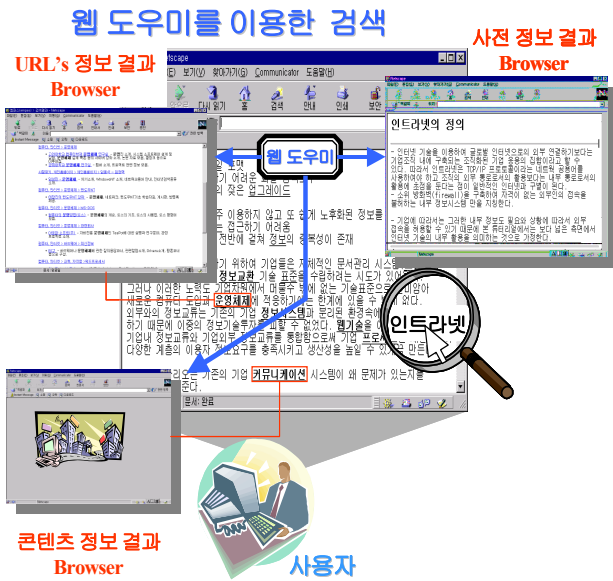
다음 (그림 4)는 문서 작성자가 웹 문서 작성하는 웹 에디터 부분으로 먼저 해당 분야를 선택하고, 입력 및 출력 형태를 선택한 후 문서를 작성하거나 기존의 문서를 덧붙이고 필수적으로 입력을 원하는 키워드는 추가 부분에 기입한다. 또한 작성된 문서의 미리보기 기능을 통하여 자동링크된 부분을 미리 볼 수 있다. 최종 작성된 문서를 보내기 버튼을 이용하여 자동링크엔진으로 보냄으로써 웹 문서의 작성을 마친다.



(그림4) 웹 에디터에서 문서 작성

위와 같이 자동링크 엔진에 보내진 문서는 앞서 언급했던 통합문서 생성과정을 통하여 만들어진 문서를 다시 받게 되며, 이를 웹 서버가 있는 호스트 컴퓨터로 옮겨 이용하게 된다. 일반 사용자가 별도의 하이퍼링크되지 않은 부분에 마우스 커서를 위치시킴으로써 웹 도우미를 이용해서 관련된 사전 정보를 비롯하여 URL 그리고 콘텐츠 정보를 다양하게 받아 볼 수 있다.

다음 (그림 4)는 웹 에디터에서 작성된 문서를 ALSII에 실행시킨 통합문서 상에서 직접 웹 브라우저 창을 이용하여 마우스 이벤트에 의한 검색 과정을 나타낸다.



(그림 4) 웹 도우미를 이용한 검색

4. 관련 연구

문서에서 관련된 정보를 자동으로 링크시켜, 하이퍼미디어 정보로 변환시키는 본 자동 링크 지원 시스템의 처리를 위해서는 가장 크게 부각되는 부분이 한글처리 문제일 것이다.

한국어는 띄어쓰기에 의해 어절 구분이 되는 형태가 아니기 때문에 단어의 의미를 가지는 형태를 구분하는 방법이 중요시되고, 형태소분석 결과의 완전성과 신속한 분석 속도를 필요로 한다. 이를 위해서 조사와 어미 분석사전을 이용하거나 어절 구조규칙을 이용하여 잘못된 분석결과를 제거하고, 어절타입 예측을 통하여 미등록어를 처리하거나, 어절 타입을 평가하는 등 여러가지 방법을 사용한다.

한국어 형태소 분석 방법인 형태소의 분리와 결합 제약, 그리고 인식이라는 문제가 복합되는 형태소 분석 기법은 형태소 분석 방향, 형태소 인식방법, 문법형태소 기술방법, 형태소 분리방법, 형태소 결합조건 기술방법, 불규칙 어절의 원형 복원법, 입력 단위에 따른 분류로 나뉘어진다.

기존의 서비스에서는 제한된 정보를 바탕으로 사용자의 수작업으로 입력된 정보에 의해서만 서비스 가능한 형태이거나, 입력되어 있던 정보와 일치하지 않은 내용에 대해서는 서비스되지 않는 제한을 가지고 있다.

자동링크 엔진의 문서 분석기에서는 형태소 분석 방향에 따른 분류방법 중 좌-우 분석법을 사용한다. 또한, Win32 Hooks 기술 중 마우스 이벤트에 의한 키워드 추출 방법을 이용하여 웹 도우미 모듈 부분을 구현함으로써 사용자가 마우스의 이동만으로 원하는 정보를 검색할 수 있

도록 하였다.

기존의 서비스와 본 논문에서 제시하고 있는 시스템에서 제공하는 정보 검색 서비스를 비교한 내용은 다음 (표 2)와 같다.

분류	내용	비고
기존의 서비스	-CD-ROM 형태의 서비스(제한된 정보) -웹 상의 검색방법의 제한점 -웹 문서 작성 기술의 한계	전자사전 키워드 검색 수작업 링크
자동 링크 지원 시스템	-풍부한 웹 문서 개발 환경 제공 -검색 방법의 다양화 -관련된 정보의 자동 링크	멀티미디어 정보 마우스 검색 정보선택

(표 2) 기존 서비스와의 차이점

5. 결론

본 논문의 자동 링크 지원 시스템은 웹 문서 작성자와 일반 사용자들에게 인터넷상의 수많은 정보들을 쉽게 접근하며 빠른 검색 효과와 텍스트 위주 콘텐츠들의 재활용과 새로운 멀티미디어 통합 정보 구축에 매우 효과가 크다. 또한 초급자의 웹 수준을 끌어올릴 수 있는 계기 뿐 아니라 사이버대학, 재택 교육을 위해서 가상 교육에서 관련 자료를 구축하는데 이점이 있으며, 전자 문서 및 신문에서 텍스트 위주의 기사 내용을 하이퍼텍스트 문서로 재구성하여 부가가치가 높은 기사를 만드는데 효과가 있다.

추후 연구 내용으로는 웹 에디터에 WYSIWYG 기능을 추가해야 하며, 전용 클라이언트 프론트 엔드 전용 에디터를 구현 할 예정이다. 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.

※ 본 연구는 중소기업 기술혁신개발사업에 의해 지원된 결과임.

참고 문헌

[1] 신 명기, 김 용진 “W3C에서의 차세대 웹 표준 활동 동향” 정보처리학회지, vol.6 No. 3 1999. 5. p 7~17.
 [2] 김 태석, 박 철제, 임 환섭 “차세대 웹 서비스 기술 개발 동향” 정보처리학회지, vol.6 No. 3 1999. 5. p 18~24.
 [3] 나 홍석, 채 진석, 김 창화, 백 두권 “차세대 웹 상에서의 문서 교환 및 검색을 위한 프레임워크” 정보처리학회지, vol.6 No. 3 1999. 5. p 52~61.
 [4] 효율적인 한국어 형태소분석 방법
<http://lactt.taegu.ac.kr/data/internet/indexing/efficient.htm>
 [5] 음절 특성을 이용한 한국어 불규칙 활용
<http://lactt.taegu.ac.kr/data/internet/indexing/illegal.htm>
 [6] G. Riessen, H.A. Jacobsen & O. Gunther, MMM, Web Engineering '99 International Workshop, 1999.