

웹의 접근성 향상을 위한 자동 북마크 시스템

이시은⁰ 황인준
아주대학교 정보통신전문대학원 정보통신공학과
{sfmovie, ehwang}@ajou.ac.kr

Automatic Bookmarking System to Improve Web Accessibility

Sieun Lee⁰ Eenjun Hwang
Graduate School of Information and Communication, Ajou University

요약

웹 환경이 복잡해짐에 따라 시각 장애인이 웹 페이지에서 원하는 정보와 서비스를 찾는 데 많은 시간이 소요된다. 본 논문에서는 웹의 접근성 향상을 위해 웹 페이지를 분석해 사용자가 접근하는 객체와 관련된 객체 주변의 정보를 추출하여 자동으로 북마크를 생성하고, 이를 이용하여 재접근 시에 저장된 북마크가 반영된 페이지로 변환시키는 시스템을 제안한다. 본 시스템의 특징은 전체 페이지가 아닌 사용자의 접근 구역을 자동 추출함으로써 재접근 시의 소요시간을 단축시키고, 페이지에 변화가 있을 경우에도 저장된 북마크 정보와 가장 상응되는 객체나 구역을 추출하여 유용적으로 제공한다. 또한 이전에 접근한 객체나 구역만으로 구성된 변환 페이지를 생성하여 사용자에게 원본(original) 페이지와 변환 페이지 사이를 필요에 따라 전환할 수 있게 한다.

1. 서론

인터넷은 정보전달과 서비스 이용의 수단으로써 시각 장애인에게 새로운 장을 열어놓고 있다. 웹 상의 정보는 신문이나 책 등의 매체에 비해 손쉽게 접할 수 있으며, 스크린 리더(screen reader)/보이스 브라우저(voice browser)와 같은 음성 인터페이스 응용프로그램을 통해 다른 사람의 도움 없이도 접근할 수 있다. 그러나 웹 환경이 복잡해짐에 따라 페이지 상의 정보와 서비스에 접근하기 위해 많은 시간과 노력이 요구되고 있고 재접근 시에도 역시 많은 시간을 소비하게 된다. 음성 인터페이스의 특성상 원하는 정보에 접근하기까지 많은 광고와 웹 페이지의 다른 요소들을 거쳐야 하기 때문이다.

본 논문에서는 시각 장애인이 갖는 웹 상의 정보 접근의 어려움을 해결하기 위한 방안으로 자동 북마크(bookmarking) 기법을 제안한다. 자동 북마크 시스템은 크게 사용자가 접근하는 객체와 구역을 추출하여 마크를 생성하는 과정과 재접근 시에 마크된 객체와 구역들의 접근 링크를 제공하는 과정 그리고 북마크의 갱신과정으로 나뉜다.

본 논문의 구성은 다음과 같다. 1장 서론에 이어 2장에서는 자동 북마크 생성 시스템의 배경 기술과 관련 연구에 대해 소개하고, 3장에서는 접근 객체와 구역을 추출하는 과정에 대해 설명한다. 4장에서는 북마크 생성과 이를 반영한 사용자 페이지 변환 방법 그리고 이를 위한 전체적인 시스템 구조에 대해 논하며, 5장에서는 향후 활용 방안 및 계획 등을 제시하며 결론을 맺는다.

2. 관련 연구

시각 장애인의 웹 접근성을 향상시키기 위해 빠른 시간 내에 페이지의 전체 구조와 내용을 파악할 수 있는 기법과 정보의 재이용 시에 필요한 시간을 단축시키는 기법 등의 연구가 진행되어왔다.

2.1. 시각 장애인의 웹 이용을 위한 코드 변환 연구

코드 변환(transcoding)을 통해 웹 페이지의 구성을 바꾸어 여러 단으로 구성된 페이지를 하나의 단으로 만들고, 주석을 이용하여 중요도가 높은 부분부터 위치하도록 하는 등 페이지의 구조를 바꾼다. 또한 중복 되는 부분을 제외하여 페이지의 복잡도를 줄여준다. 이를 통해 사용자는 음성 인터페이스가 읽어주는 정보를 더 쉽게 이해할 수 있고, 접근성을 높일 수 있다 [1, 2].

2.2. 문장중심의 북마크/갱신정보 추출 시스템

페이지 내의 관심 있는 문장에 북마크 함으로써 재접근 시 해당 콘텐츠(contents)의 위치를 빠르게 이동할 수 있도록 한다. 이를 통해 텍스트가 많은 페이지 내에서 정보의 재접근 시간을 단축시킬 수 있다 [3]. 그리고 새롭게 갱신된 부분의 접근이 가능하여 신문과 같은 내용 갱신이 잦은 페이지를 다시 읽어야 하는 불편함을 감소시킬 수 있는 방법을 다루고 있다 [1, 3].

2.3. page map 시스템

이 시스템은 PDA 같은 작은 화면 기기를 위해 페이지를 의미구역(같은 유형의 정보가 포함된 부분)으로 나누어 각 구역의 정보를 제공하는 시스템이다. 이 시스템을 시각 장애인을 위한 시스템으로 확장 시킨다면 사용자가 웹 전체의 내용을 쉽게 파악하고 원하는 구역으로 빠르게 이동할 수 있는 방법을 제공하게 될 것이다 [4, 5, 6].

3. 자동 북마크 생성

시각 장애인들은 단지 음성 인터페이스가 읽어 내려가는 순서대로 정보를 파악할 수 밖에 없기 때문에 원하는 정보를 찾기 위해 많은 시간이 소요된다. 이러한 어려움은 웹 페이지의 구성이 점차 복잡해지면서 더욱 심해지고 있다.

일반 사용자가 웹 페이지를 볼 때는, 한 눈에 페이지의 구성과 내용을 훑어 볼 수 있다. 이는 웹 페이지를 구성하는 정보가 시각적으로 구역을 이루기 때문이다. 즉 같은 주제, 분류의 정보들을 하나로 묶어진 구역 안에 두어 정보들을 정리해 놓음으로써 페이지 내에서 원하는 정보와 서비스를 빠르게 찾을 수 있다 [5].

3.1. 접근 객체 추출

사용자는 웹 페이지 내에서 여러 작업(클릭, 텍스트 입력 등)을 수행한다. 현재 브라우저가 제공하는 기능은 이러한 행동 중에서 특정 주소의 페이지로 이동하거나 특정 주소로 정보를 보내는 이벤트(event)를 발생 하는 작업을 통해 생성되는 접근 링크 주소를 모아 즐겨 찾기 (favorite-MS Internet Explorer, bookmark-Netscape navigator), 목록 보기(history list)와 같은 기능을 제공하는데 그치고 있다. 이러한 접근 방법은 구체적이지 못하며, 해당 페이지 내에서 사용자의 관심이 있는 부분을 반영하지 못한다. 따라서 관심 있는 영역을 북마크하는 방법이 요구된다. 또 로그인(login)이 필요한 페이지를 북마크하는 경우 저장된 북마크 주소로 이동할 때 해당 페이지로 바로 접근하지 못하는 경우가 많다. 이는 사용자 확인을 요구한 뒤 해당 페이지로 리다이렉트(로그인 요구 후 해당 페이지로 바로 접근할 수 있게 하는 기능)되는 기능이 사이트에서 제공하는데 의존하고 있기 때문이다. 이런 리다이렉트가 필요한 페이지의 경우, 본 시스템의 기능을 이용하면 로그인 후 사용자의 경로가 추적되므로 재접근 시에 리다이렉트 되는 효과를 제공한다.

자주 사용하는 페이지, 특히 그 중에서도 특정 서비스를 제공하는 웹 사이트의 경우 사용자는 자주 이용하는 구역, 링크 등을 갖는다. 따라서 이러한 부분들을 하나의 객체로 보고 이들을 자동으로 북마크 해 사용자에게 제공한다면, 재접근 할 때 빠르게 원하는 부분으로 접근 할 수 있다.

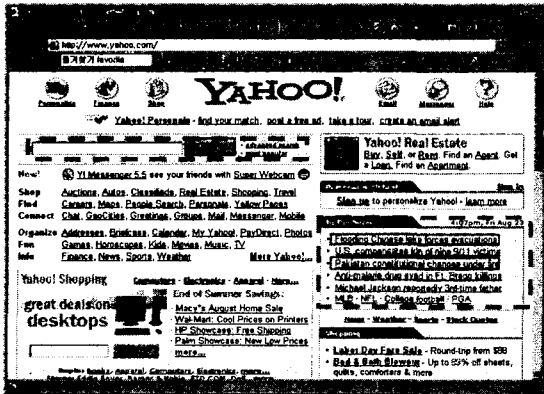


그림 1. 추출 객체와 구역의 예

사용자의 접근 객체를 추출하는 과정은 크게 접근한 링크(단일객체)의 위치를 파악하는 과정과 단일객체가 위치한 구역을 추출하는 과정으로 나눌 수 있다. 단일객체만으로 의미를 제공하기 어려운 경우 구역을 추출할 필요가 있다. 예를 들어 웹 페이지 내의 로그인을 위한 폼과 검색을 위한 폼의 경우 텍스트를 입력하고 버튼을 누르면 입력한 정보가 정해진 주소로 전송되게 된다. 따라서 접근한 버튼 객체를 사용자가 관심을 가진 객체라고 판단하여, 추출하게 된다. 그러나 이 단일 객체만으로는 의미를 제공하기 어려우므로 버튼 객체가 위치한 로그인(또는 검색 폼) 구역을 추출하여 북마크 한다.

또한 접근한 단일 객체들이 한 구역 내에 비율-구역의 전체 콘텐츠 대비 마크 된 콘텐츠의 비율-이 높은 경우에도 구역을 추출한다. 그림 1은 웹 페이지의 태그를 분석하여 추출되는 객체와 구역을 보여준다. 그림 내의 실 선은 객체이고 점선은 이 단일 객체들이 위치한 구역이다. 사용자가 객체에 이벤트를 발생시키면 이 객체를 마크하고 객체가 위치한 구역을 추출한다. 표 1은 접근 객체를 추출하고 이를 구역추출로 확장시키는 과정을 나타낸다.

표 1. 접근 객체,구역 추출 과정

1.	접근 링크 주소를 바탕으로 HTML 파일을 분석하여 태그 내의 위치를 파악한다.
2.	객체가 속하는 구역 내에 로그인, 검색을 위한 폼이 있는 경우 구역을 북마크한다.
3.	객체가 이미 북마크 되어 있다면 중요도를 높여주고 위치한 구역의 중요도도 높여준다.
4.	구역이 북마크 되어 있지 않고 마크 된 단일 객체 중 같은 구역에 포함된 객체가 있다면 구역의 비율을 높여준다. 일정 수준 이상의 비율이 되면 구역을 마크한다.

3.2. 접근 객체의 관리

추출된 접근 객체들의 정보는 XML형태로 구성되어 클라이언트에 저장된다. 객체의 정보는 링크 URL주소와 링크와 관련된 텍스트나 이미지 등의 콘텐츠 내용, 그리고 객체가 페이지에서 위치한 태그정보 등을 담게 되고, 구역에 대한 정보는 구역의 태그정보(구조정보)와 구역의 첫 콘텐츠, 구역 내에 있는 접근 객체 정보를 담는다. 이 XML파일에는 객체와 구역의 정보가 사이트 주소, 사이트 이름, 해당 페이지 이름과 주소, 객체와 구역의 중요도, 구역 내의 접근 객체의 비율 등의 정보와 함께 저장된다. 비율은 객체들이 한 구역 내에 나타나는 분포 정도로 비율이 높으면 추출 범위를 구역으로 확장하여 북마크한다. 중요도는 객체나 구역의 중요도로 사용자가 북마크 된 객체, 구역을 재접근 시에도 자주 이용했는지를 나타낸다. 이 정보들은 계속 유지, 관리되어 재접근이 발생할 때 마다 갱신되어 비율, 중요도 등을 계산하게 된다.

현재 브라우저에서 “즐거찾기”와 쿠키 정보를 다른 응용프로그램이나 파일로 주고 받을 수 있도록 하고 이 자료를 공유할 수 있는 서비스를 제공하는 웹 사이트도 여럿이다. 이와 유사하게 본 시스템의 객체와 구역에 대한 북마크도 공유하고 주고 받아 재활용할 수 있다.

4. 추출된 북마크를 이용한 페이지 변환 시스템

4.1. 시스템 구조

그림 2는 북마크 생성 시스템의 전체적인 구조를 보여준다. 전체 시스템은 크게 북마크 관리자(bookmark manager), 북마크 정보를 XML형태로 저장해 두는 XML 저장소(XML repository), 페이지 변환 관리자(transcoding manager)로 구성된다. 사용자가 웹에 접속 하면 페이지 변환 관리자 내의 파서(parser)는 웹 서버(web server)로부터 요청한 페이지를 받아 브라우저에 전달한다. 그리고 현재 페이지의 태그를 분석하여 구역 정보를 추출한다. 북마크 관리자는 이 페이지의 URL(uniform resource locator)주소와 일치하는 페이지의 북마크를 XML 저장소에서 찾는다. 일치되는 북마크가 없다면

사용자가 처음 접근한 페이지이다. 페이지 내에서 사용자가 링크에 대한 클릭 등으로 접근 객체를 발생시키면 해당 객체의 URL 정보가 발생하게 된다. 북마크 관리자 내의 객체 추출기(accessed object extractor)가 URL 정보를 얻어와 페이지 내에서 상응되는 URL을 갖는 객체를 찾고 해당 객체가 페이지에 존재하면 파서가 분석한 구역정보를 이용하여 접근 객체가 위치한 구역을 찾는다.

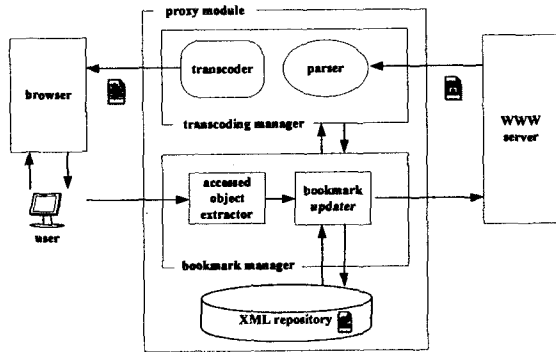


그림 2. 자동 북마크생성 시스템

해당 객체가 페이지에 존재하지 않는다면 이는 페이지 내의 접근에 의해 발생한 것이 아니라는 것을 의미한다. 추출된 객체와 구역은 북마크 관리자에 의해 XML 형태로 XML 저장소에 저장된다. 과거에 접근하여 그 기록이 XML 저장소에 있는 페이지에 접근 하여 페이지 내의 객체에 접근 시 북마크의 갱신이 발생된다. 이전 북마크 객체, 구역 정보와 비교하여 새로운 객체일 경우 저장하고, 이전에 저장되어 있던 객체일 경우 중요도를 높인다. 또한 해당 구역의 비율도 높아지게 되고 비율과 중요도의 갱신이 있으므로 다음 접속 시 갱신된 정보가 반영되어 페이지가 구성된다. 페이지에 다시 접근할 때 사용자가 북마크 된 객체로 구성된 페이지를 요구하면 북마크 관리자가 XML 저장소로부터 북마크 정보를 페이지 변환 관리자의 페이지 변환기(transcoder)에게 넘겨준다.

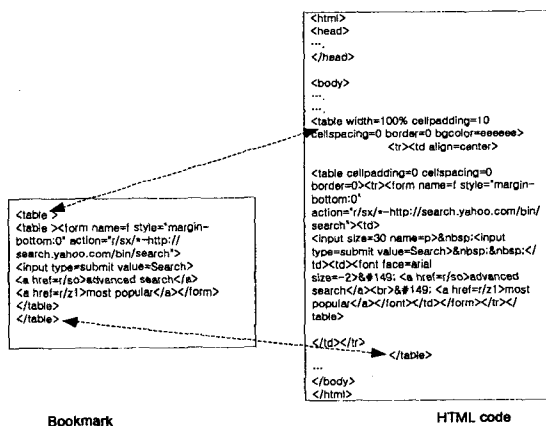


그림 3. 북마크와 페이지의 구조 비교

4.2. 페이지 구성

페이지 변환기(transcoder)는 북마크 정보를 이용하여 사용자가 접근했던 객체, 구역에 빠르게 접근 할 수 있도록 페이지를 구성한다. 사용자는 북마크가 반영 되지 않은 페이지를 볼 수 있고 북마크 된 객체와 구역들만 이루어진 변환 페이지로 전환할 수 있다. 페이지 변환기가 이러한 전환과 북마크가 반영된 변환 페이지를 구성한다. 변환 페이지로의 전환 할 때, 북마크 된 객체와 구역의 구조정보와 의미정보를 바탕으로 가장 상응되는 객체, 구역을 찾아 페이지 내에 중요도가 높은 객체, 구역 순서로 위치시키고, 이미지(image)와 같이 텍스트가 아닌 객체는 ALT 태그 정보를 이용하여 텍스트 정보를 대신 제공한다. 이렇게 북마크 된 객체와 구역으로 페이지를 구성함으로써 기존 브라우저의 북마크 기능의 단점인 로그인 필요한 페이지의 북마크 후 접근 시 리다이렉트(로그인 요구 후 해당 페이지로 바로 접근할 수 있게 하는 기능)되는 기능이 웹 서버에서의 제공에 의존한다는 점을 개선시킬 수 있다. 접근한 객체를 제공함으로써 로그인 후 사용자가 이동한 경로를 쉽게 찾아 나갈 수 있기 때문이다. 그림 3은 북마크의 구역 정보와 상응되는 HTML 코드 내의 구역을 보여준다. 상응되는 객체, 구역이 발견되면 이들을 추출하여 변환된 페이지를 생성한다.

5. 결론 및 향후 연구 과제

본 논문에서는 사용자의 접근 경로를 추적, 페이지 내의 구역을 자동 북마크하여 접근시간을 단축할 수 있도록 페이지를 변환하는 시스템을 제안하였다. 그리고 접근 구역을 추출하는 과정과 페이지를 재구성 하는 과정을 살펴보았다. 본 시스템은 기존의 북마크 시스템의 단점인 전체 페이지를 북마크하기 때문에 사용자의 관심영역을 구체화 시켜주지 못하는 점을 객체와 구역을 북마크하여 사용자의 접근을 반영함으로써 개선하고, 웹 서버에 맡겨진 리다이렉트의 기능을 간접적으로 제공한다. 하지만 HTML 페이지 구조의 다양성 때문에 객체가 속한 구역을 일반 사용자가 시각적으로 구분하는 구역과 일치시키는 데 대한 연구가 좀 더 진행되어야 할 필요가 있다. 한편 본 논문의 기법은 시각 장애인들 염두에 두고 제안하였으나, 작은 화면기거나 저 사양 시스템에서의 연구에도 활용이 가능하고 향후 북마크 된 객체, 구역 정보를 이용하여 컬렉션(collection) 페이지를 구성하는 시스템으로의 확장을 검토 중이다.

6. 참고 문헌

- [1] Hironobu T, and Chieko A., "Transcoding Proxy for Nonvisual Web Access", The 4th Int'l ACM conference on Assistive technologies 2000.
- [2] Masahiro Hori, et al., "Annotation-Based Web Content Transcoding", The 9th Int'l. WWW Conference, 2000.
- [3] Tsuyoshi E, Seiji I, and Teruhisa M., "Fast Web by Using Updated Content Extraction and a Bookmark Facility", The 4th Int'l ACM conference on Assistive technologies 2000.
- [4] Hironobu T, and et al., "Web Content Transcoding For Voice OUTPUT", Center On Disabilities Technology And Persons With Disabilities Conference 2002.
- [5] D. Song, and E.Hwang, "PageMap: Summarizing Web Pages for small Display Devices", The International Conference on Internet Computing 2002.
- [6] F.G.-C. L. Anido, et al., "hierarchical Atomic Navigation for Small Display Devices", 10th Int'l. WWW Conference, 2001.