

碩士學位論文

웹 환경에서의 개인화 북마크 에이전트 시스템



濟州大學校 大學院

컴퓨터工學科

秦 庸 碩

2006年 6月

웹 환경에서의 개인화 북마크 에이전트 시스템

指導教授 李 尙 俊

秦 庸 碩

이 論文을 工學 碩士學位 論文으로 提出함



秦庸碩의 工學 碩士學位 論文을 認准함

審査 委員長 郭 鎬 榮 印

委 員 李 尙 俊 印

委 員 金 度 縣 印

濟州大學校 大學院

2006 年 6 月

*Personalized Bookmark Agent Sytem
for the Web Environment*

Yong-Seok Jin

(Supervised by professor Sang-Joon Lee)

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE DEGREE OF MASTER OF
ENGINEERING

DEPARTMENT OF COMPUTER ENGINEERING
GRADUATE SCHOOL
CHEJU NATIONAL UNIVERSITY

2006. 6.

목 차

Abstract	1
I. 서 론	3
II. 관련 연구	5
1. 개인화	5
2. 에이전트	9
3. 웹 접근성 향상을 위한 관련 연구	12
III. 개인화 북마크 에이전트 시스템	16
1. 시스템 구현	16
2. 시스템 평가	29
IV. 결론 및 향후 연구	30
참고문헌	31

그림 목 차

Fig. 1 My 네이트 페이지의 초기 화면	5
Fig. 2 네이버의 북마크	16
Fig. 3 개인화 북마크 에이전트 시스템의 구조	17
Fig. 4 개인화 북마크 에이전트 시스템의 실행 흐름도	18
Fig. 5 개인화 북마크 에이전트 시스템의 실행 화면	20
Fig. 6 HTML 문서의 메타 태그 정보	21
Fig. 7 메타 태그 추출 소스의 일부	22
Fig. 8 도메인의 분류 체계	23
Fig. 9 개인화 북마크 에이전트 시스템의 자동 분류 지정 흐름도	26
Fig. 10 분류와 북마크 목록 화면	27
Fig. 11 Ajax가 적용된 소스의 일부	28

표 목 차

Table 1. 개인화 북마크 에이전트 시스템의 데이터베이스 구조	19
Table 2. 최상위 도메인(Top Level Domain)	24
Table 3. 국내 도메인(Second Level Domain)	25
Table 4. 북마크 시스템 비교	29

Abstract

Personalized Bookmark Agent System for the Web Environment

Yong-Seok Jin
Dept. of Computer Engineering
The Graduate School
Cheju National University

Many solutions have been presented to improve Web accessibilities. A bookmark system is one of those convenient solutions. The convenience and usefulness of the bookmark system is decreased when the registered contents are increased. In this study, I proposed the personalized bookmark agent system.

This system automatically keeps the URL of the web document which the user access, and provides the access-link which is discriminate. And this system automatically renovates the priority order of bookmark list which stands on the user's visit-record as well.

With this system, the bookmark contents are automatically updated by reflecting personal user's web preferences. And by managing the bookmark system with the server system, users can access the bookmark system at any place.



I. 서 론

인터넷은 가장 유용한 정보 전달의 수단으로 웹상에서 접할 수 있는 모든 정보는 웹 문서(Web Document)를 통해 제공되고 있다. 이러한 웹 문서는 현재 인터넷에 무수히 존재하고 있으며 지금도 계속적으로 증가하고 있다. 인터넷 사용자로서는 많은 양의 웹 문서 중에서 필요한 정보를 빠른 시간 안에 정확하게 찾는 것이 매우 중요하다. 이를 위해 인터넷 사용자는 웹 검색 엔진이나 웹브라우저에서 제공하는 즐겨찾기 기능 또는 북마크 기능 등을 사용하고 있다.

웹 검색 엔진의 사용은 초기 검색에 있어서 원하는 정보를 가진 웹 문서까지의 경로를 제공받는 데는 유리하지만, 재방문을 고려할 경우 웹 접근성(Web Accessibility)에 있어 효율성이 떨어지는 면이 있다. 따라서 대부분의 인터넷 사용자는 정확한 웹 사이트까지의 접속을 위해 웹브라우저가 제공하는 북마크 도구를 사용하고 있다. 인터넷의 모든 정보는 웹상의 정보 위치를 알려주는 URL(Uniform Resource Locator)을 가지고 있는데, 북마크 도구는 바로 이러한 URL을 관리하기 위하여 웹브라우저가 제공하는 기능이다. 많은 웹 문서로 인하여 정보 과적(Information Overload) 상황에 노출된 인터넷 사용자들은 필요한 정보를 빠르고 정확하게 검색하기 어렵다. 따라서 효율적인 웹 서핑(Web Surfing)을 위해 인터넷 사용자들은 북마크(Bookmark) 도구를 이용한다.

웹브라우저가 제공하는 북마크 도구는 정보 이용자가 방문하고자 하는 웹 사이트의 주소를 기억할 필요없이 URL을 기억하고 있지 않아도 재방문이 용이하도록 한다. 웹브라우저에 종속된 북마크 도구는 이러한 효율적인 웹 서핑에 도움을 주지만 몇 가지 문제점을 가지고 있다. 첫째, 인터넷 사용자가 웹 사이트를 북마킹(Bookmarking)할 것인지의 여부를 직접 판단하여야 하는 점

이 있다. 둘째, 북마킹도 인터넷 사용자가 수동적으로 하여야 한다는 것이다. 셋째, 북마크의 등록 개수가 늘어날수록 북마크의 분류와 정렬과 같은 관리가 지속적으로 이루어져야 하며, 이는 사용자가 직접 수행해야 하는 기능이라는 것이다. 넷째, 북마크 등록 개수의 증가와 사용자의 직접 관리 기능 수행은 북마크의 본래 목적인 웹 접근성 향상을 어렵게 한다라는 것이다.

본 논문에서는 현재 북마크 시스템의 문제점을 해결하기 위한 방안으로 사용자 개인의 선호도를 이용한 개인화 북마크 에이전트 시스템을 제안한다. 제안한 시스템은 크게 사용자가 방문하는 웹 문서의 URL을 자동 저장하여 차별화된 접근 링크를 제공하는 기능과 방문 횟수를 바탕으로 한 북마크 우선 순위 자동 갱신 기능을 제공할 수 있도록 하였다.

웹 사이트 주소는 도메인 체계를 이용하여 기본적인 분류를 통하여 자동 저장되고, 분류명의 변경이 있으면 웹 사이트 주소는 접속한 웹 사이트의 문서 제목과 메타 태그의 keywords 속성과 description 속성을 참조하여 자동으로 분류하여 저장할 수 있도록 하였다.

본 논문의 구성은 I장 서론에 이어 II장에서는 관련연구 분야를 살펴보고, III장에서는 개인화 북마크 에이전트 시스템의 구현에 대해서 설명한다. IV장에서는 결론을 내리고 향후 연구에 대해 기술한다.

II. 관련 연구

1. 개인화

웹 기반의 개인화의 예는 네이트닷컴(<http://www.nate.com>), 네이버(<http://www.naver.com>), 야후!코리아(<http://www.yahoo.co.kr>) 등 대부분의 포털 사이트에서 흔히 볼 수 있다. 즉, My 네이트(<http://my.nate.com>)와 같은 기능들이 그것이다. Fig. 1에서 볼 수 있듯이 My 네이트와 같은 기능들은 사용자의 선호에 맞게 편집하여 볼 수 있는 기능을 비롯해 사용자의 스타일에 맞는 정보를 선별하여 볼 수 있다. 또한 아마존(<http://www.amazon.com>)를 비롯한 전자상거래 사이트에서도 사용자의 개인적 선호도에 따라 자신의 페이지를 구성하고 사용자의 구매기록과 성별, 나이 등을 바탕으로 상품을 추천 받을 수 있는 기능을 제공한다.



Fig. 1 My 네이트 페이지의 초기 화면

1) 개인화의 개념

일반적으로 개인화라고 하면 웹 사이트와 연관지어서 웹 사이트 개인화라는 용어를 사용하는 것이 보통인데, 특히 개인화라는 개념을 설명하기 위해서는 인터넷 마케팅에 관련하여 정리할 필요가 있다.

E-business의 중요한 요소로 부상되고 있는 개인화 마케팅(Personalized Marketing)의 기본 골격은 오래 전부터 논의되어 왔다. 1993년 D. Peppers와 M. Rogers가 쓴 “The One-to-One Future: Building Relationships One Customer at a Time”이라는 책에서 처음으로 소개된 개인화 마케팅(1대1 마케팅)의 개념은 현재 우리들에게 전혀 생소한 것은 아니다. 개인화 마케팅은 관계 마케팅의 일종으로서, 매스마케팅(Mass Marketing)이 TV나 라디오와 같이 불특정 다수에게 같은 내용의 광고나 정보를 제공하는 방법을 취한다면 관계 마케팅(Relational Marketing)은 고객과 1대1의 관계를 맺고 표적화된 광고나 서비스를 제공하는 방법을 말한다.

그런데 개인화 마케팅을 실행함에 있어서 고객이 많아지면 많아질수록 개인의 정보를 관리하고 이를 마케팅에 활용하는 데 대한 제반 비용이 높아지는 문제가 발생한다. 이로 인해 현실적으로 고객을 군집화(Segmentation)하는 경우가 많아 진정한 의미에서의 개인화 마케팅은 실현되고 있다고 보기 어렵다.

개인화(1대1) 마케팅이 이전까지의 마케팅 방법과 다른 점은 고객을 동질화된 시장 개념으로 보고 시장 점유율을 높이는 것을 목표로 삼는 기존의 마케팅과는 달리 고객을 각각 다른 욕구를 가진 독특한 존재로 보고 각각의 고객을 만족 시키는 것을 목표로 삼고 있다는 것이다.

하지만 이러한 개인화 마케팅은 기존의 매체에 의해서 시도되기에는 기술적, 비용적 한계로 인하여 실현 불가능한 개념에 불과하였다. 그러나 오늘날 인터넷이라는 새로운 매체가 보편화 되면서, 그와 함께 데이터 마이닝 기법이 진전되고 컴퓨터 시스템이 발전되면서 개인화 마케팅의 실현이 가능하게 되

었다. 즉, 인터넷 쇼핑몰에서의 모든 거래행위는 실시간으로 데이터베이스에 축적되어 질 수 있고, 고객의 모든 정보를 이용, 분석하여 해당 고객에게 적합한 서비스를 선별적으로 제공하는 것이 가능해진 것이다.

이는 마케팅 이론에서 말하는 ‘시장의 세분화’를 극단적으로 한사람을 목표로 하는 “고객 한 명의 시장”이라는 차원으로까지 실현가능하게 된 것을 의미한다. 이러한 새로운 패러다임이 Don Peppers와 Martha Rogers가 말하는 진정한 “일대일 마케팅”의 개념이다.

인터넷은 일대일 마케팅이 가능한 환경을 조성해 주고 있으며 그것의 중요한 전제는 “개인화”이다. 즉 개인화는 정보통신 기술의 발전에 기반하여 사용자가 원하는 것을, 사용자가 원하는 시기에, 사용자가 원하는 방식으로 제공하고자 하는 지속적인 과정이자 전략이라고 정의할 수 있다[1].

2) 개인화의 유형

웹 사이트에서의 개인화는 일련의 가치교환 과정이다. 사용자는 자신의 선호, 관심, 구매경력과 같은 정보를 웹 사이트에 제공하며 웹 사이트는 사용자가 제공한 정보를 기초를 사용자에게 가장 알맞은 정보를 제공한다. 개인화를 통해서 웹 사이트 운영자는 사용자에게 관한 정보를 얻고 사용자의 지속적인 이용이나 구매를 얻어낼 수 있게 되며 사용자는 자신에게 가장 알맞은 정보를 보다 편리한 방법으로 얻을 수 있게 된다.

현재 웹 사이트에서 구현되고 있는 개인화는 다음과 같이 크게 4가지 유형으로 나누어 볼 수 있다.

- 가. 개인화 메뉴 (예: My네이트, MyYahoo! 등)
- 나. 추천 및 고객 서비스 센터 (예: 인터파크 등)
- 다. 검색 에이전트 (예: Coldwell Banker's Personal Retriever 등)
- 라. 개인화 광고

이 중 일부를 살펴보자면, 우선 ‘개인화 메뉴’는 국내 유수의 포털 사이트들에서 모두 제공하는 기능으로 예를 들어 Fig. 1의 My네이트 페이지와 같이 사이트 운영자가 제시하는 서비스들 중에서 이용자가 자신이 원하는 서비스 항목만을 선택하여 자신만의 페이지를 구성할 수 있도록 하는 기능이다. 이것은 웹상에서 개인이 원하는 정보를 복잡한 과정을 거치지 않고 빠른 시간 내에 획득할 수 있게 함으로써 개인의 해당 사이트에 대한 의존도 내지는 방문도를 유의적으로 증가시킬 것이다.

또한 ‘개인화 광고’는 일반대중을 대상으로 하는 기존의 광고(Advertising)와는 달리 특정개인을 대상으로 한 광고를 의미하는데, 상거래와 연계시킬 수 있는 단순하면서도 효과가 높은 구체적인 개인화 광고로는 기념일 통보(Tailroad email Alerts)나 선호 물품 리스트(Wish List) 등을 들 수 있다. 그 외 중·고급의 정보 분석 기법이 필요한 서비스로는 제품 추천(Product Recommendations), 타겟 마케팅/광고(Targeted Marketing Advertising), 개인화된 가격(Custom Pricing) 등을 들 수 있다.

지능형 검색 에이전트(Intelligent Searching Agent) 서비스와 관련하여, 그 개념에 대해서는 뒤에 이어지는 에이전트 관련 연구 부분에서 설명하기로 하고, 우선 그 종류에 대해서 살펴보자면 다음과 같다. 즉, 구매자의 상품 탐색 과정을 도와주고 추천해 주는 상품 추천 에이전트(Recommendation Agent), 구매자의 취향 및 특성에 따라 상품에 대한 카달로그 정보를 제공하는 통지 에이전트(Notification Agent), 구매자가 지정한 특정 상품에 대하여 여러 판매 사이트의 가격 및 조건 등을 비교하여 제시하는 비교 쇼핑 에이전트(Comparison Shopping Agent), 구매자와 판매자가 제시한 여러 조건을 만족시킬 수 있도록 협상하고 가장 적합한 구매자와 판매자를 중개해주는 협상 에이전트(Negotiation Agent) 등이 있다.

여기에서는 편의상 위의 네 가지 유형으로 구분하였지만, 개인화 광고는 광의적으로 추천(Recommendation)에 포함된다고 볼 수 있다. 또한 서비스 사용자 입장에서 보자면 개인화 메뉴가 검색 에이전트나 추천(Recommendation)

에 비하여 사용자의 참여와 관심이 필요하다고 볼 수 있으나, 커스텀마이징 (Customization)된 검색 에이전트나 추천 시스템 또한 ‘개인화’로 구분하는데 무리가 없다[1].

본 논문에서도 이와 맥을 같이 하여 개인화의 한 유형으로 개인화 메뉴 영역의 기능 확장으로서 개인화 북마크 에이전트 시스템을 제안하고자 한다.

2. 에이전트

에이전트 시스템 분야의 연구는 컴퓨터와 통신기술이 결합된 분산 컴퓨팅 환경이 출현하게 됨에 따라 새로운 영역으로 확장되고 있다. 즉, 초기 분산인공지능 분야의 한정된 연구에서 벗어나 정보 검색, 전자상거래, 원격교육, 이동 컴퓨팅 등의 분산 환경에 기초한 고도의 분산 복합 시스템을 분석하고 설계, 구현하는 효과적인 방법론으로 발전하고 있다.

에이전트에 관한 연구의 역사적인 배경을 보면 지난 반세기 동안 인공지능 분야의 사실상의 주된 연구 과제였다고 해도 과언이 아니다. 인공지능 분야의 초창기인 50년대와 60년대에 단순한 모델과 이론을 토대로 지능시스템을 구현하고자 했던 태도는 70년대와 80년대에 들어서 퇴색되면서 특정 기술 개발을 위한 구체적인 연구로 발전되었다. 그 결과로 지식표현, 탐색기법, 게임 방법론, 전문가시스템, 기계학습 자연어처리 등 인공지능 연구는 세분화 또는 전문화 되었다. 그 과정 중에서 70년대 후반에 분산인공지능(Distributed Artificial Intelligence)이라는 새로운 연구 분야가 형성되어 다양한 종류의 분산인공지능 문제들을 ‘에이전트’라는 추상적인 단위와 에이전트간의 상호작용을 토대로 해결하려는 방안이 시도되었다. 이러한 과정을 거쳐 90년대에 들어서는 인터넷과 웹의 급속한 성장과 더불어 웹 정보환경의 소프트웨어 에이전트의 출현을 맞게 된 것이다[2][3].

1) 에이전트의 정의

1980년대 초반부터 언급되기 시작한 에이전트(Agent)라는 용어는 인공지능, 데이터 통신, 병렬처리 시스템, 로봇 시스템, 사용자 인터페이스 설계 등의 여러 분야에서 사용하고 있다. Russell과 Norvig은 에이전트란 센서(Sensor)를 통하여 자신의 주변 환경을 인식(perceiving)하고 효과기(effectors)를 이용하여 환경에 대한 동작을 수행하는 것이라고 포괄적으로 정의하고 있다. 좀더 구체적으로 정의하면 에이전트란 지속적으로 환경에서 지각된 것과 내부 지식을 바탕으로 추론하고 그 추론된 결과를 행동하여 환경에 영향을 미치고 또한 사용자를 포함한 다른 에이전트와 의사 소통을 하는 지속적으로 존재하는 소프트웨어 요소라고 말할 수 있다. 따라서 에이전트는 일반적으로 주변 환경으로부터 감지기를 통하여 지식을 얻어 내부에서 추론하여 환경에 작용하는 구조를 가지고 있다.

하지만 최근 에이전트라는 용어를 사용하여 진행되는 연구 주제들과 접근 방법들은 너무 다양하여 에이전트란 용어를 학문적으로 정확히 정의하는 것은 힘들다. 따라서 에이전트가 일반적으로 가져야 하는 주요 특성들을 열거하여 간접적으로 에이전트라는 용어를 정의하고 있다. 일반적으로 에이전트가 갖추어야 할 특성들로는 다음과 같은 것들이 있다.

첫째, 자율성(autonomy)은 에이전트와 다른 일반 프로그램을 구별해주는 가장 핵심이 되는 특성으로 사용자의 직접적인 참여나 지시가 없이 스스로 어떤 작업이나 목표를 달성하기 위한 행동을 할 수 있는 능력을 말한다. 둘째, 사회성(social ability)은 에이전트 통신언어를 사용하여 다른 에이전트와 통신할 수 있는 능력이 있어서 서로간의 대화를 통한 상호작용이 가능함으로서 설정된 작업을 수행해 나가는 능력을 말한다. 셋째, 외부환경을 감지하고 환경의 변화에 맞게 내부에 미리 설정해놓은 행동을 취할 수 있는 능력을 나타내는 반응성(reactivity)과, 과거에 유사한 상황에서 성공했던 행동을 찾아내어 취할 수 있도록 하는 능력을 나타내는 적응성(adaptation)이 있다. 넷째, 학습

성(learning)은 에이전트가 사용자의 습관이나 패턴을 학습하여 다음에 그와 비슷한 상황이 주어지면 과거의 학습을 바탕으로 하여 사용자에게 결과를 돌려줄 수 있는 능력을 말하며 이러한 자신이 가지고 있는 능력이 시간의 흐름에 따라 성능이 향상되어야 한다. 다섯째, 에이전트가 어떤 상황에서 행동을 취하는데 있어서, 그 상황에 맞는 행동이 하나 이상 될 수 있으며 그 중 하나의 행동을 취해야 한다. 이처럼 상황에 맞는 행동을 선택하고 선택한 행동들 중에서 가장 적절한 행동이 무엇인가를 판단하여 그 행동을 취할 수 있는 능력, 반면 그 행동이 실패하였을 때 어떻게 해야 하는가를 결정하는 능력 등을 지능성(intelligence)이라 한다. 여섯째, 이동성(mobility)은 에이전트가 하나의 지역적인 컴퓨터 내에서만 작업하는 것이 아니라 분산 네트워크상의 여러 호스트간을 이동하면서 필요한 작업을 수행하는 능력을 말한다[2].

2) 에이전트의 응용분야

에이전트가 사용된 응용 시스템의 연구는 이론과 구조 등을 기반으로 실제 사용할 수 있는 응용 소프트웨어를 개발하는 것이다. 이러한 연구는 현재 실험 단계를 벗어나 실세계의 산업 및 상업용으로 급속히 발전해 나가고 있다. 에이전트 기법이 적용되는 대표적인 여러 응용 분야 중에서 특히 현재 에이전트 응용 시스템의 필요성이 가장 많이 요구되고 또한 많은 기술 연구와 제품 개발이 이루어지고 있는 분야를 산업용, 상업용, 오락용, 의료용으로 구분할 수 있다[2][4].

본 논문에서 제안하는 개인화 북마크 에이전트 시스템 또한 이러한 에이전트 이론을 응용함으로써 앞에서 설명한 바 있는 개인화를 가능케 하고자 한다.

3. 웹 접근성 향상을 위한 관련 연구

웹 접근성 향상을 위한 목적으로 많은 연구가 진행되어 왔다. 특히 인공지능과 웹에서의 정보 통신 기술에 대해 다루고 있는 WI(Web Intelligence) 분야에서는 웹 정보에 대한 마이닝, 검색, 웹 에이전트 등의 다양한 연구가 이루어지고 있다[9].

이들 연구는 기존의 웹 정보를 다루는 관점에서 벗어나 여러 관점에서 접근하였는데, 여러 웹페이지에 나뉘어 기술되어 있으나 논리적으로는 하나의 문서로 볼 수 있는 정보들에 대하여 정보를 묶어 하나의 페이지셋으로 보는 논리적인 정보 단위의 추출을 대한 연구[6]와 재방문 시의 웹 접근성 향상을 위한 연구[7], 현재의 웹 서핑의 문제점과 불편함을 개선하기 위한 WI(Web Intelligence)[8] 연구 등이 진행되고 있다.

1) 논리적인 정보 단위의 추출

Web Skimming은 사용자 질의를 입력받아 관련있는 웹페이지들의 흐름을 지정하는 컨텍스트 패스(context path)를 찾아 정보의 흐름 순으로 해당 페이지들을 보여준다. 이를 이용해 사용자는 웹사이트 내에서 원하는 정보들의 구성을 쉽게 찾을 수 있다.

Wen-syan L. 등은 관련된 웹페이지들의 묶음을 하나의 논리적인 정보의 단위로 보고, 이 연관된 정보를 담은 페이지들을 검색한다. 이를 통해 논리적으로 연관된 정보들을 함께 볼 수 있는 장점이 있다.

Xiaoli Li 등은 검색 시 하나의 웹페이지에 여러 주제의 정보가 함께 담길 수 있음을 고려해 페이지 내의 정보 단위를 검색에 활용하였다. 이는 사용자의 질의어 분포가 하나의 구역에 밀집할수록 페이지에 포함된 질의어들이 같은 주제의 정보를 나타낼 가능성이 높음을 고려한 것으로, 프레임과 테이블

구조를 이용해 한 페이지 안에 많은 정보를 담는 경향을 반영한 효율적인 검색 방법을 제시하였다[9].

2) 재방문시의 웹 접근성

북마크 기능의 가장 큰 한계는 사용자가 직접 페이지를 북마크해야 한다는 점에서 수동적이라는 것이다. 또한 북마크 저장의 단위가 하나의 페이지 단위로 고정되어 있기 때문에 사용자의 관심 영역을 구체적으로 반영하지 못하는 문제가 있다. 이와 같은 문제점을 개선한 연구가 진행되어왔다.

Shaun K. 등은 “back”, “history”, 그리고 “bookmark” 기능을 통합해 사용자의 별도의 요청이 없이도 자동으로 사용자가 자주 방문하는 페이지를 북마크한다.

Tsuyoshi E. 등은 페이지 내의 관심있는 문장에 북마크할 수 있도록 함으로써 재접근 시 해당 콘텐츠의 위치로 빠르게 이동할 수 있도록 한다. 따라서 텍스트가 많은 페이지 내에서 정보의 재접근 시간을 단축시켜 준다. 이를 통해 새롭게 갱신된 부분의 접근도 지원해 신문과 같이 내용 갱신이 잦은 페이지를 다시 읽어 내려가야 하는 불편함을 감소시키는 방법을 다루고 있다.

Web VCR은 웹에서의 정보가 정보의 논리적인 흐름을 따라 하이퍼링크와 프레임을 이용해 여러 웹페이지에 걸쳐 표현하고 있기 때문에 필요한 정보를 얻기 위해 여러 페이지를 링크를 따라가며 일일이 열어 보아야하는 불편함을 개선시키고자 하였다. 이를 위해 북마크 시 바로 서비스에 접근하지 않고 로그인 등의 과정을 거쳐야 도달할 수 있는 서비스에 바로 접근할 수 있도록 사용자의 접근 경로, 이동 경로와 로그인 폼, 검색 폼 등에 입력하는 정보 등을 저장해 재접근 시 사용자의 경로대로 반복하여 바로 접근할 수 있도록 한다[9].

3) 웹 탐색을 위한 WI

현재의 웹 서핑의 문제점과 불편함을 개선하기 위해 사용자의 웹 탐색을 도와주고 조언해주는 시스템에 대한 연구 또한 많이 이루어지고 있다.

링크된 페이지의 내용을 미리 제공하여 사용자의 웹 탐색을 돕는 연구로서 MS WebScout은 마우스 포인터가 링크 텍스트 위에 위했을 때 링크 페이지를 이미지 형태로 미리 가져와 보여준다. 현 브라우저의 사용자가 접근했던 링크에 대해 색을 달리하여 구분케 하는 방식에서 나아가 마우스를 링크에 대면 해당 링크에 대한 페이지 저자, 표현 언어, 그리고 접근 예상 시간 등의 정보를 보여주는 연구도 진행되었다[9].

4) 웹 정보의 표현

좀 더 이해하기 쉬운 정보 형태를 제공하기 위해 웹페이지를 분석하여 편리한 인터페이스를 제공하고자 하는 연구 또한 많이 진행되었다. 예를 들어 Hunter Gatherer는 여러 웹페이지로부터 문장, 이미지, 테이블 등의 정보 객체를 추출하여 하나의 페이지에 관리하는 기법을 제공한다.

웹페이지의 HTML 코드 변환을 통해 원래의 구성, 형식과 다른 형태로 웹페이지를 변형하는 연구도 진행되어왔다. 대부분 HTML 코드 변환 과정은 여러 단으로 구성된 페이지를 하나의 단으로 만들고, 주석을 이용하여 중요도 높은 부분부터 위치하도록 하는 등 페이지의 구조를 바꾸는 데 이용된다. 또한 중복되는 부분을 제외하여 페이지의 복잡도를 줄여주는 방법으로 시각 장애인을 위해 간략한 형태의 문서로 만들어주거나 기기(device)의 차이를 고려하여 사용자의 기기에 알맞은 형태로 변환해 사용자 맞춤 변환 서비스를 제공할 수 있도록 한다[9].

본 논문에서는 앞서 설명한 웹 접근성 향상을 위한 방안 중 재방문시의 웹 접근성 향상을 위한 방안으로 기존의 북마크 시스템을 개선한 자동 북마킹과

북마크가 저장될 분류의 자동 지정 기능 그리고 가장 최근, 가장 많이 접속한 사이트를 기준으로 한 북마크된 사이트의 순위 기능을 제공하는 개인화 북마크 에이전트 시스템을 제안하고자 한다.



Ⅲ. 개인화 북마크 에이전트 시스템

1. 시스템 구현

Fig. 2는 네이버(<http://www.naver.com>)에서 제공하는 웹 기반의 북마크 시스템이다. 이 시스템에는 북마크 추가, 폴더 추가, 수정, 삭제 등과 같은 북마크 시스템의 기본적인 기능이 포함되어 있다.



Fig. 2 네이버의 북마크

네이버의 북마크 시스템은 사용자가 직접 북마크를 등록하고 관리하여야 한다. 이는 인터넷 익스플로러 등의 웹브라우저에서 제공하는 북마크 시스템과 같다. 본 논문에서 제안하는 시스템은 위 북마크 시스템이 가지는 기본 기능에 자동 분류 등록 및 방문 우선순위 북마킹을 포함하고 있다.

제안 시스템의 전체적인 구조는 Fig. 3과 같다. 사용자는 웹브라우저를 통하여 개인화 북마크 에이전트 시스템에 접근하며, 실행은 아파치 웹 서버 하에서 이루어진다. 사용자의 요청은 PHP로 구현된 개인화 북마크 에이전트 시스템으로 전달되고, URL 정보는 URL 검색기와 통계기능수행을 통해 데이터 베이스에 우선순위와 함께 북마크 된다. 처리과정이 모두 끝나면 응답페이지

를 작성하고, 사용자가 북마크를 요청하면 웹 브라우저를 통해 응답페이지를 전달한다.

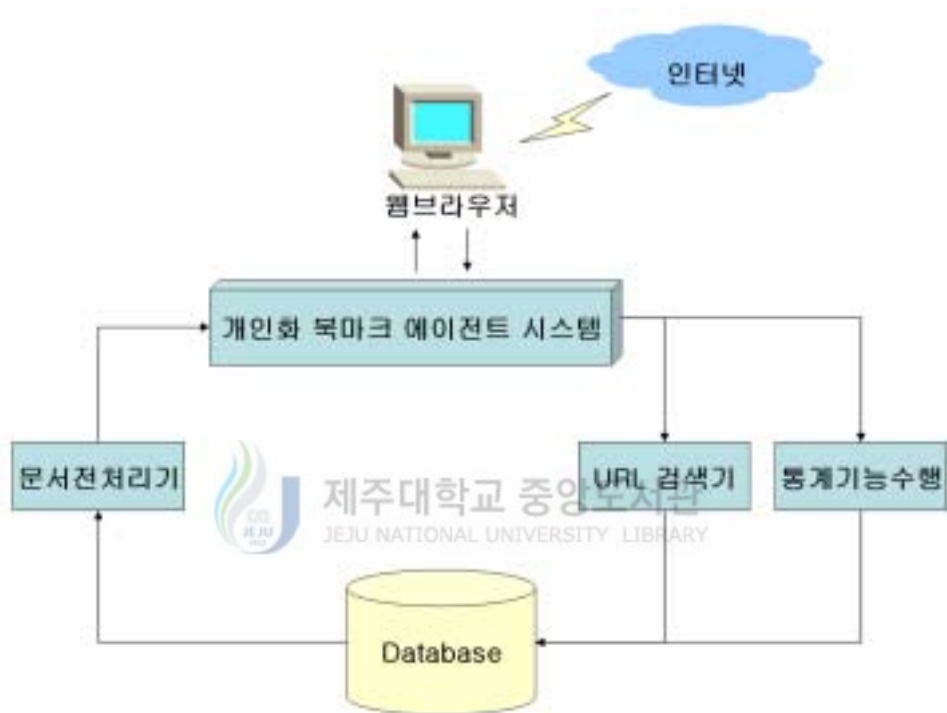


Fig. 3 개인화 북마크 에이전트 시스템의 구조

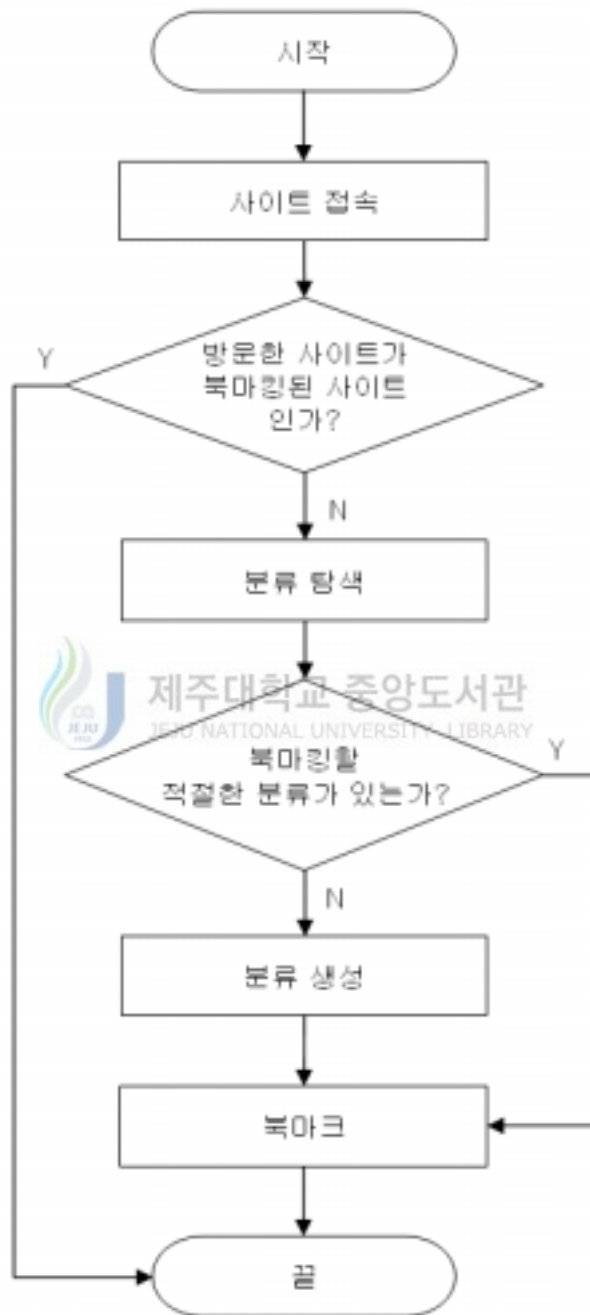


Fig. 4 개인화 북마크 에이전트 시스템의 실행 흐름도

북마크 구조는 기존의 북마크 구조에 가중치 정보와 자동 분류를 위한 태그 정보를 포함하는 확장된 북마크 구조를 가진다. 확장된 북마크 구조는 Table 1과 같다.

Table 1. 개인화 북마크 에이전트 시스템의 데이터베이스 구조

Table 이름	tb_bookmark	
필드명	타입	설명
num	int	auto_increment
uid	varchar	사이트 id
title	varchar	사이트 제목
url	text	사이트 URL
createDate	timestamp	만든 날짜
modDate	timestamp	수정한 날짜
accessDate	timestamp	액세스한 날짜
groupId	varchar	분류 id
weight	varchar	가중치

Table 이름	tb_group	
필드명	타입	설명
gid	varchar	분류 id
groupName	varchar	분류 제목
createDate	timestamp	만든 날짜
modDate	timestamp	수정한 날짜
accessDate	timestamp	액세스한 날짜
weight	varchar	가중치

tb_bookmark 테이블은 북마크 정보를 저장하는데 사용되며, 가중치 필드는 북마크의 우선순위를 결정하는데 사용된다. tb_group 테이블은 북마크한 사이트의 분류를 정하고 저장하는데 사용된다.



Fig. 5 개인화 북마크 에이전트 시스템의 실행 화면

개인화 북마크 에이전트 시스템의 실행 화면은 Fig. 5와 같다. 상단 프레임의 주소 입력창을 통해 사이트에 접속하면 개인화 북마크 에이전트 시스템은 사이트의 URL(Uniform Resource Locator)과 일치하는 북마크를 북마크 저장소(Bookmark Repository)에서 찾는다. 일치하는 북마크가 없으면 사용자가 처음으로 액세스한 사이트이다. 처음 액세스한 사이트이면 개인화 북마크 에이전트 시스템은 Fig. 6에서 처럼 사이트의 타이틀과 메타 태그 주변 정보를 분석하여 가장 유사한 분류에 사이트를 북마크한다.

```
<head>
  <title>PHPBuilder.com, the best resource for PHP tut
manuals, content management systems, scripts, classes ar
  <meta name="keywords" content="PHP, php, mysql, php
manual, server, pdf, database, flash, phpbuilder, conter
system, sql, script, oracle, string, xml, regular expres
webalizer, php tutorials, code, nusoap, classes, develop
  <meta name="description" content="PHPBuilder.com, th
tutorials, templates, PHP manuals, content management sy
classes and more for the PHP developer.">
```

Fig. 6 HTML 문서의 메타 태그 정보

아래 Fig 7. 은 HTML 문서의 타이틀 태그와 메타 태그의 keywords 속성 및 description 속성 정보를 추출하는 소스의 일부분이다.

```

fputs ($fp, "GET / HTTP/1.0\r\nHost:$url\r\n\r\n");

/**** <head> ~ </head> 추출 ****/

while(!feof($fp)){
    $head .= fgets($fp, 4096);
    if( strpos($head, "</head>") ) break;
}

$head = trim($head);

/**** <title> ~ </title> 추출 ****/

$start = strpos_f($head, "title", 0);
$start = strpos_f($head, ">", $start) + 1;
$end   = strpos_f($head, "<", $start);

$title = substr($head, $start, ($end-$start));

/**** <meta> 추출 ****/
/**** <meta name='' content=''> 으로 가정 ****/

$start = 0;
$meta = $head;
for($i = 0; $i < 5; $i++){
    if(!strpos_f($meta, "<meta", $start)) break;
    $start = strpos_f($meta, "name=", $start);
    $start += strlen("name=");
    $end   = strpos_f($meta, " ", $start);
    $name = substr($meta, $start, ($end-$start));

    if(meta_check($name)){
        $start = $end;
        $start = strpos_f($meta, "content=", $start);
        $start += strlen("content=");
        $end   = strpos_f($meta, ">", $start);

        $keyword .= addslashes(substr($meta, $start, $end-$start));
        $start = $end;
    }
}

$keyword .= $title;

$keyword = str_replace("'", "", $keyword);
$keyword = str_replace("W'", "", $keyword);

fclose($fp);

```

Fig. 7 메타 태그 추출 소스의 일부





 제주대학교 중앙도서관
 JEJU NATIONAL UNIVERSITY LIBRARY

Fig. 8 도메인의 분류체계

URL을 구성하는 도메인은 Fig. 8와 같은 분류체계를 가지고 있다. 인터넷 초창기에 생성된 COM, NET, ORG (등록 시 별다른 제한 없음), EDU, GOV, INT, MIL (등록 시 제한 있음) 등 7개의 TLD와 2001~2002년에 걸쳐 새롭게 생성된 BIZ, INFO, NAME, PRO (이상 Unsponsored TLD), AERO, COOP, MUSEUM(이상 Sponsored TLD) 등 7개 TLD를 합쳐 현재 총 14개의 gTLD가 있으며, Table 2에서 볼 수 있는 것처럼 각 국가별 최상위 도메인도 TLD에 속한다.

Table 3은 국내 도메인(Second Level Domain)을 보여주고 있는데, 국내 도메인에는 지역 KR 도메인도 포함되어 있다. 서울은 SEOUL.KR, 부산은 PUSAN.KR, 제주는 CHEJU.KR 등으로 등록되어 있다.

Table 2. 최상위 도메인(Top Level Domain)

도메인	기 관
COM	영리를 목적으로 하는 기업/기관
NET	네트워크를 관리하는 기관
ORG	일반적으로 비영리 단체
EDU	4년제 단과대학/종합대학
GOV	미국 연방정부기관
MIL	미국 연방군사기관
INT	국제적인 특성을 가진 기관(유엔 등)
KR	한국 도메인
JP	일본 도메인
CN	중국 도메인
UK	영국 도메인
FR	프랑스 도메인
IT	이탈리아 도메인
DE	독일 도메인
NL	네덜란드 도메인
NP	네팔 도메인
AU	오스트레일리아 도메인
NZ	뉴질랜드 도메인
TH	태국 도메인
NP	북한 도메인
기타	각 국가별 도메인 및 BIZ, INFO, NAME, PRO 등

Table 3. 국내 도메인(Second Level Domain)

도메인	기 관
CO.KR	영리를 목적으로 하는 기업이나 회사
PE.KR	개인 도메인
OR.KR	비영리 목적의 기관이나 단체
NE.KR	네트워크에 관련된 기관 (NM.KR)
GO.KR	대한민국정부 관련 기관
RE.KR	연구를 목적으로 하는 학술기관
AC.KR	대학교
HS.KR	고등학교
MS.KR	중학교
ES.KR	초등학교
SC.KR	기타학교
SEOUL.KR	서울 특별시
PUSAN.KR	부산 광역시
CHEJU.KR	제주특별자치도
TAEGU.KR	대구 광역시
INCHON.KR	인천 광역시
KWANGJU.KR	광주 광역시
ULSAN.KR	울산 광역시

개인화 북마크 에이전트 시스템의 초기 분류는 URL의 TLD(Top Level Domain)과 SLD(Second Level Domain)를 추출하여 기본 분류를 정하고 URL과 사이트의 타이틀 정보를 저장한다. 처음으로 정해지는 분류는 TLD와 SLD를 가지고 지정하므로 정부기관, 회사, 대학교 등과 같은 분류를 Fig. 8의

흐름도를 따라 기본 분류로 가지게 된다. 초기에 개인화 북마크 에이전트 시스템에 의해 지정된 분류는 사용자가 세분화하여 분류를 확장한다.

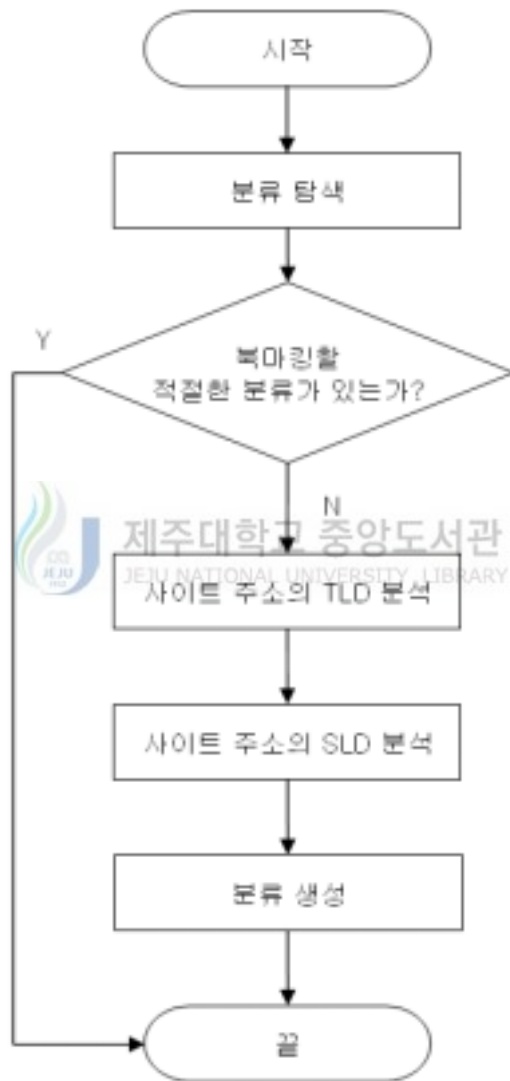


Fig. 9 개인화 북마크 에이전트 시스템의 자동 분류 지정 흐름도

Fig. 8은 분류와 북마크 목록 화면이다. 상단 프레임의 북마크 보기 단추를 누르면 분류와 북마크 목록 페이지가 호출된다. 분류는 수정과 삭제 기능으로 사용자 정의가 가능하고, 분류와 북마크는 가중치가 적용되어 값이 높은 순으로 나열된다. 가중치는 사이트 접근횟수로 정해지며, 가중치가 같은 경우에는 최근 접속일을 기준으로 정렬된다.

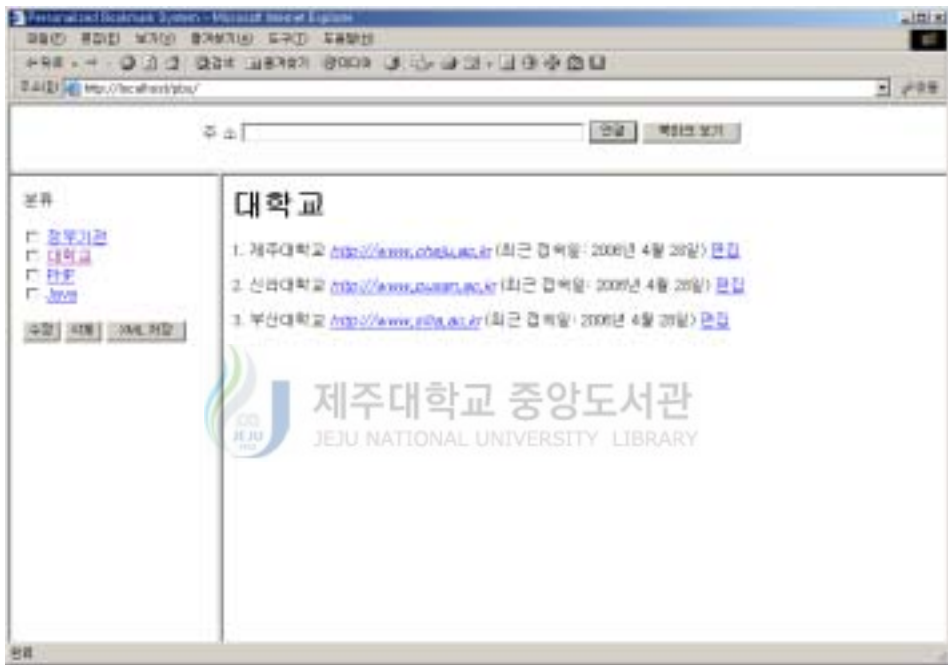


Fig. 10 분류와 북마크 목록 화면

북마크는 편집을 통해 개인화 북마크 에이전트 시스템에서 자동으로 지정된 사이트의 제목, URL을 변경이 가능하고, 사이트 설명을 추가할 수 있다. 북마크 저장소(Bookmark Repository)에 저장되어 있는 분류와 북마크 목록은 XML 문서로 변환되어 이종 시스템으로 자료 전달이 가능하다. 또한 Ajax[10]를 적용하여 사용자가 접속하고자 하는 사이트가 북마크된 사이트이면 사이트의 몇 문자의 입력만으로 사이트의 주소를 출력하도록 하였다.

Ajax는 Asynchronous JavaScript and XML의 약자로 ‘에이작스’라고 읽으며, XHTML과 CSS를 사용하는 표준 기반의 프레젠테이션, 브라우저의 문서 객체 모델(Document Object Model)를 사용하는 동적인 디스플레이 및 상호 작용, XML과 XSLT를 사용하는 데이터 교환 및 처리, XMLHttpRequest나 XMLHttpRequest를 사용하는 비동기 데이터 추출 등의 기술을 포함하고 있다 [11][12].

Fig. 11에서 Ajax를 적용한 개인화 북마크 에이전트 시스템의 일부 소스를 볼 수 있다.

```
$db = new db_que();
$db->dbconn();

$Result = $db->query("SELECT url FROM bookmark WHERE url LIKE '$k
int", 1);
$Count = mysql_num_rows($Result); // 결과물이 적은 경우 대비.
if( $Count == 0 ) { exit; }

$re = "<select name='test' size='2' id='search' style='width:100%
:)'>";
for($i = 0; $i < $Count; $i++){
    $Data = mysql_result($Result, $i, 0);
    $re .= "<option value='$Data'$Data</option>";
}
$re .= "</select>";
echo $re;
```

Fig. 11 Ajax가 적용된 소스의 일부

2. 시스템 평가

제안한 시스템은 웹 접근성 향상을 위하여 사용자의 관리가 필요없는 자동 북마킹 및 분류 자동 지정 기능을 가지고 있다. 웹 서핑시 가장 많이 사용되고 있는 브라우저인 마이크로소프트사의 인터넷 익스플로러에 내장된 북마크 시스템과 네이버에서 제공하는 북마크 시스템, 그리고 제안한 시스템과의 비교를 통하여 평가를 해보았다. 결과는 Table 4와 같다.

Table 4. 북마크 시스템 비교

구분	인터넷 익스플로러 내장 북마크 시스템	네이버의 북마크 시스템	제안한 시스템
자동 북마킹	X	X	○
분류 자동 지정	X	X	○
순위 제공	X	X	○
분류 수정	○	○	○
분류 추가	○	○	○
분류 삭제	○	○	○
하위 분류 추가	○	○	X

Table 4에서 볼 수 있듯이 제안한 시스템은 하위 분류 추가를 제외한 나머지 기능을 모두 만족하고 있다. 하위 분류 추가는 분류의 계층 문제 해결을 포함한 성능 향상에 관한 사항으로 향후 개선이 필요한 부분이다.

IV. 결론 및 향후 연구

본 논문에서는 웹 접근성 향상을 위한 방안으로 웹브라우저에 종속되지 않는 분류 자동 지정과 자동 북마크를 할 수 있는 개인화 북마크 에이전트 시스템을 제시하였다. 구현 결과 기존 웹브라우저에 종속된 북마크 도구 보다 웹 접근성에 있어 향상이 있음을 확인 할 수 있었다. 본 시스템의 성능 향상과 기능 개선을 위해 고려되어야 할 향후 연구 과제로는 분류의 중복 문제 해결, 대·중·소분류로 나눌 수 있는 계층 문제 해결, 사용자들간 북마크 공유 지정을 통해 협업을 지원할 수 있는 시스템의 구현 등이 있다. 또한 사용자가 URL이 아닌 키워드를 입력 가능하게 하여 북마크 목록에서 사용자의 기대에 가장 근접하는 URL을 추천하는 시스템으로의 확장을 검토 중이다.



[참고문헌]

- [1] 배희호, “웹 마이닝을 이용한 개인화(Personalized) 시스템 설계에 관한 연구”, 경북 논총, 제6권, 2002, pp.11-14.
- [2] 최정민, “북마크 자동 분류를 위한 학습 에이전트에 관한 연구”, 경기대학교 대학원, 2000, pp.4-16.
- [3] 이재호, “에이전트 시스템의 연구 및 개발 동향”, 한국정보과학회지 제18권 제5호, 2000, pp.4-9.
- [4] 장병탁, “학습 에이전트”, 한국정보과학회지 제18권 제5호, 2000, pp.26-35.
- [5] Ning Zhong, Jiming Liu, Yao, Y.Y., Ohsuga S., "Web Intelligence(WI)", Computer Software and Applications Conference, 2000.
- [6] Wen-Syan L., Quoc V., Divyakant A., "Retrieving and Organizing Web Pages by InformatinUnit", WWW10, May 1-5 2001, Hong Kong., ACM 1-58113-348-0/010005.
- [7] Shaun K, et al, "Integrating Back, History and Bookmarks in Web Browsers", Proceedings of CHI'01, ACM Press, 2001, pp.379-380.
- [8] Liebrman H., "An Agent the Assists Web Browsing", 14th Int'l joint conference on Artificial Intelligence IJCAI95, 1995.
- [9] 이수철, 이시은, 황인준, 웹 접근성 향상을 위한 웹 서핑 도우미, 정보과학회 제31권 제9호, 2004, pp. 1181-1183.
- [10] <http://en.wikipedia.org/wiki/AJAX>

[11] Justin Gethland, Ben Galbraith, Dion Almaer, 「Pragmatic Ajax」,
2006, pp18

[12] <http://www.ajaxian.com/>



감사의 글

먼저 이 논문이 나오기 까지 부족한 저를 제자로 삼아 많은 관심과 격려로 보살핌과 가르침을 주신 이상준 지도교수님께 진심으로 감사드립니다.

그리고 이 논문을 면밀히 검토하시고 조언을 해주신 곽호영 교수님, 김도현 교수님에게도 감사를 드립니다. 또한 대학원에서 진정한 학문의 길을 열어주신 김장형 교수님, 송왕철 교수님, 안기중 교수님, 변상용 교수님, 변영철 교수님께도 감사 드립니다.

우리 연구실의 선후배님에게도 감사 드립니다. 대학원 생활에 대하여 아낌 없는 조언을 해주신 김영민 선생님, 그리고 박충희 선생님, 송재경씨, 특히 연구실의 살림꾼이었던 정은경 선생님과 친구 문남원에게 깊은 감사를 드립니다. 그리고 한경복 선생님, 권훈, 허지완 선배님, 이정하 선생님에게도 이 지면을 빌어 감사의 마음을 전합니다. 이밖에 대학원 생활을 하면서 알게 된 동기와 선후배님들에게도 감사를 드립니다.

또한 제주한라대학 컴퓨터정보활용과의 김형수 교수님, 연구실 선배님이시기도 한 제주한라대학 인터넷전자상거래과의 김휴찬 교수님, 주식회사 지노의 용상우 팀장, 김미희씨, 박만표씨에게도 깊은 감사를 드립니다.

마지막으로 사랑하는 아버지, 어머니, 동생 민정, 성영 그리고 아침마다 날 깨워주는 조카 서진에게 진심으로 고마움을 전하면서 감사의 글을 마칠까 합니다.