

학습시스템에 기반한 개인화 정보 서비스에 관한 연구

A study on the personalization information service based on learning system

남 궁 황(Nam-Goong Hwang)*

초 록

본 연구에서는 도서관과 정보센터에서 전통적으로 제공되어진 SDI 서비스와 함께 일부 기관에서 제공하고 있는 맞춤정보 서비스를 기반으로 한 개인화 정보 서비스 시스템의 기술요소 및 전체 시스템의 구조를 제안하였다. 제안된 개인화 정보 서비스 시스템은 이용자의 개인 프로파일 정보를 바탕으로 이용자의 정보탐색행태 등을 추가 입력 값으로 해서 학습과정을 거쳐 이용자에게 가장 적절한 정보를 출력 값으로 제공할 수 있다. 이를 위해 개인화 정보 서비스 시스템에서 가장 중요한 기능을 수행하는 학습시스템과 추천시스템을 구축하는데 필요한 요소에 대해 살펴보았다.

ABSTRACTS

With SDI service provided in libraries and information centers traditionally, this paper studies component technologies and structure of system platform in PIS(personalization information service based on the customized information service served currently in some institutions. The PIS system should provide relevant information as an output through the learning system analyzing user information searching behavior as an input value with personal profile information. To do it, this paper studies requirements and algorithms to develop PIS, and proposes learning system and recommendation system as core components in PIS.

키워드: 맞춤정보 서비스, 개인화 정보 서비스, 선택적 정보배포서비스, 푸시기술, 이용자프로파일, 고객관계관리, PIS, SDI, CIS, CRM, learning system, personalization information service

* 합동참모본부 사서사무관/중앙대학교 문헌정보학과 강사(nam0181@hanmail.net)

- 논문접수일 : 2003. 11. 25
- 게재 확정일 : 2003. 12. 3

1 서 론

인터넷과 정보기술의 발전은 정보의 저장 및 가공과 함께 이를 이용자에게 제공해야 하는 도서관이나 정보유통기관에 새로운 도전과 발전의 기회를 제공하고 있다. 정보환경의 변화는 1960년대 이후 도서관 및 정보센터에 컴퓨터가 도입된 이래로 목록정보의 전산화, 온라인 열람용 목록(OPAC) 등이 구축됨으로서 정보제공의 시간적, 공간적인 제한점을 어느 정도 해결 할 수 있었으며, 이와 더불어 최근에는 정보자원의 형태도 인쇄형 매체나 목록정보 중심에서 전자형태 혹은 전문데이터 중심으로 그 범위가 확대되어 가고 있다.

그러나 정보 및 정보유통체계를 둘러싸고 있는 급격한 환경의 변화는 정보욕구 충족의 행태까지도 변화시키고 있다. 초기 인터넷 출현 시 대부분의 도서관 및 정보센터를 포함한 정보제공기관들은 전자도서관과 같은 정보저장소를 구축하여 자관의 서비스 이용자나 모든 정보이용자들의 요구를 충족시킬 수 있을 것이라고 기대하였지만, 정보생산을 위한 다양한 저작도구의 발달과 인터넷 활용의 확산에 따른 정보의 폭발적인 증가로 인해서 정보과잉의 문제가 대두되었다. 이로 인해 정보이용자가 자신이 원하는 정보를 찾는 데 많은 시간과 노력이 요구될 뿐만 아니라 상당수 부적합한 정보가 검색되어 지고 있다. 그래서 이러한 문제를 해결하기 위한 다양한 방법들이 도서관과 정보센터

를 중심으로 나타나게 되었는데, 그 대표적인 방법이 검색엔진의 등장이라고 할 수 있다. 그러나 단순히 검색엔진의 등장만으로도 근본적으로 문제를 해결 할 수 없었기 때문에 전통적인 도서관에서 제공되고 있는 선택적 정보배포(SDI: Selective Dissemination of Information)서비스가 중요한 대안으로 제시되었다. 예컨대, 선택적 정보배포 서비스의 개념을 기반으로 이용자의 프로파일 정보에 푸시기술을 적용시킨 보다 능동적이고 발전된 맞춤형 정보 서비스가 등장하게 되었다. 이러한 푸시기술을 적용한 맞춤형 서비스는 전통적인 도서관 및 정보센터에서 이용자의 요구에 맞추어 주기적, 정기적으로 정보를 제공하는 최신정보배포 서비스와 그 기능이 동일하다고 할 수 있다. 그러나 맞춤형 서비스는 단순히 이용자의 프로파일 정보에 기반을 두고 프로파일과 일치되는 모든 정보를 제공함으로써 여전히 정보과잉의 한계점은 극복할 수 없었으며, 특히, 이용자의 정보요구 행태의 변화에도 적절하게 대응 할 수 없다는 제한점이 있다.

따라서 본고에서는 개인화 정보 서비스 시스템의 개발을 위한 알고리즘 및 적용 가능한 기술에 대해 알아보고, 개인화 정보 서비스 시스템의 전체적인 구조와 더불어 이 시스템에서 가장 중요한 구성요소에 해당되는 이용자의 정보요구 성향을 분석하는데 필요한 학습시스템의 구조를 제안하고자 한다.

2 최신정보제공과 SDI서비스

정보 생산량의 급증과 더불어 내용의 전문화 및 세분화 그리고 정보 유통방법의 복잡화 등으로 인하여 이용자가 직접 정보시스템에 접근하여 적합한 정보를 적시에 탐색하거나 관련 주제 분야의 발전 동향을 파악하고 최신성을 유지하기가 점차 어려워지고 있다. 이러한 정보검색이나 최신정보와 관련된 문제에 대처하기 위한 방안으로 도서관 및 정보센터에서는 최신정보주지(Current awareness) 서비스를 제공하고 있다. 일반적으로 SDI 서비스는 선택적 정보배포서비스, 최신정보주지 서비스(Current awareness service) 또는 경고 서비스(alert service) 등으로 불리어지고 있는데, 이는 이용자가 탐색을 원하는 데이터베이스를 선택하여 탐색 전략을 입력하면, 데이터베이스가 갱신될 때마다 또는 이용자가 입력한 기간동안 자동으로 탐색이 실행되어 탐색결과가 이용자에게 전자메일을 통해서 제공되는 서비스를 의미한다.

최신정보주지 서비스는 전통적인 도서관이나 정보센터에서 새로이 입수되는 최신 자료를 이용자에게 알리는 속보서비스로써 목차속보, 속보지의 발간 배포, 잡지 회람, 신간전시, 게시판의 활용 등 여러 가지 방법으로 제공되고 있다. 특히, 이 서비스는 특정 분야나 특정인을 대상으로 하는 것이 아닌 불특정 인을 대상으로 하는 막연한 주지 서비스이다. 최근에는 인

터넷이 상용서비스에 개방됨에 따라 전자저널, 목차 속보 등 최신의 정보를 제공함으로써 인터넷은 최신정보주지 서비스를 위한 중요한 수단이 되고 있다.

선택적 정보배포 서비스는 최신정보주지 서비스와 혼용할 정도로 대표적인 최신정보제공의 수단으로써 다른 여러 가지의 최신정보 서비스와는 구분된다. 기계화된 시스템이든 수작업에 의한 시스템이든 간에 이용자의 요구주체에 맞는 최신정보를 개인에게 선택적으로 제공하는 기능을 모두 선택적 정보배포 서비스에 포함시킬 수는 있으나, 보통은 컴퓨터를 이용하여 이용자의 정보요구를 프로파일 에 등록시켜 놓고 이 프로파일과 일치하는 문헌을 이용자에게 제공하는 서비스를 의미한다.

이러한 선택적 정보배포 서비스는 도서관이나 정보센터에 입수되는 문헌에 한정해서 적용됨으로 현재와 같은 전자형식의 정보환경에서 이 서비스 업무를 효율적으로 수행하는 데 많은 한계가 있다. 즉, 선택적 정보배포 서비스의 경우에는 주로 도서관이나 정보센터에 입수되는 신착자료에 대한 목록이나 문헌에 대한 기본정보만을 제공하여 왔다. 그러나 정보의 형식이 전자화 되어지고 문헌의 유형이 신착자료목록 이외의 뉴스, 웹문서 등과 같이 다양화 되고 또한 정보의 생산주기가 짧아짐으로서 단순한 선택적 정보배포 서비스로는 이용자의 요구를 충분히 만족시킬 수가 없게 되었다. 그래서 이러한

한계점을 극복하기 위해서 일부 도서관 및 정보센터에서는 선택적 정보배포 서비스의 개념을 기반으로 하는 맞춤형 정보 서비스 등이 새롭게 등장하게 되었다.

3 SDI의 모델 분석

3.1 맞춤형 정보 서비스 (Customized Information Service)

3.1.1 맞춤형 정보 서비스의 개념

최근 인터넷과 웹(WWW)의 확산으로 인해 수많은 정보가 범람하게 되었고, 이에 따라 know-what, know-how에서 know-where가 증시되는 시대가 되었으며, 많은 이용자들이 자신이 원하는 정보만을 선택적으로 획득하고자 하는 맞춤화(customization) 욕구가 증가하게 되었다. 맞춤형 정보 서비스는 인터넷 웹 환경을 기반으로, 이용자로부터 자신이 원하는 정보만을 추출할 수 있도록 프로파일을 설정할 수 있게 하고 그에 따라 정보를 검색하여 일정 주기로 웹 페이지 또는 전자우편방식으로 자동 배달하는 서비스 시스템이다.

정보에 대한 맞춤화는 최근 다양한 분야에서 응용되고 있는 정보검색 기술과 정보 푸시 기술(Push Technology) 그리고 웹 에이전트(Agent) 등의 기술이 결합되어 구현될 수 있다. 이용자는 맞춤형 정보 서비스를 활용함으로써 정보를 찾기 위한 네비게이션(navigation) 시간과 노력을 절약

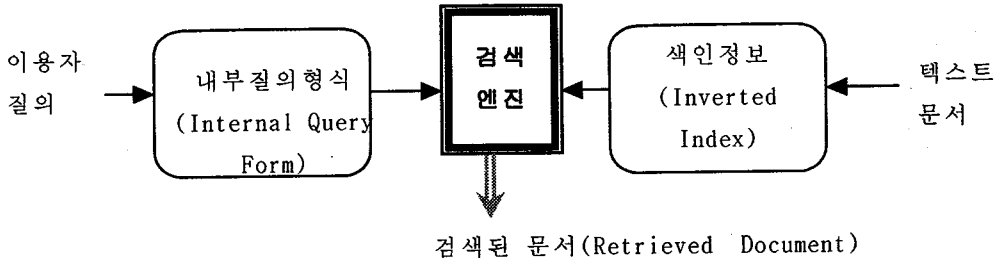
하고, 불필요한 정보의 이동으로 인한 네트워크 트래픽의 증가를 방지하며, 이용자 개개인이 양질의 정보를 쉽게 유지, 관리할 수 있다.

3.1.2 맞춤형 정보 서비스 시스템의 특징

일반적으로 맞춤형 정보 서비스를 제공하기 위해서는 다음과 같은 요구사항을 만족시켜야 한다.

- 맞춤형 정보 서비스를 위한 범용 플랫폼
- 표준화되고 사용이 편리한 웹 사용자 인터페이스
- 뛰어난 검색기능
- 맞춤형 정보 배달서비스
- 독립적인 시스템 구조

이러한 요구사항과 함께 맞춤형 정보 서비스 시스템은 향후 서비스 콘텐츠를 변경할 경우 부분적인 구성(configuration) 요소만 수정하여도 사용이 가능할 수 있도록 범용성을 극대화시켜야 한다. 최근에는 이 서비스 시스템의 기능이 대부분 웹 환경에서 구현됨으로 별도의 클라이언트 소프트웨어의 존재여부와 관계없이 웹 브라우저(메일 클라이언트 포함) 환경에서도 쉽게 서비스를 이용할 수 있어야 하고 정보검색엔진과의 연동도 완벽히 이루어져야 한다. 또한 맞춤형 정보 서비스 시스템의 각 모듈은 확장성을 높이기 위해 서로 독립적이며, 독자적으로 일정 주기에 따라 실행되도록 개발되어야 한다.



〈그림 1〉 정보검색 시스템의 실행과정

3.1.3 맞춤형 정보 서비스의 기술요소와 구조

3.1.3.1 기술요소

맞춤 정보 서비스는 이용자를 대신하는 일종의 에이전트 개념을 도입한 것으로, 정보검색 기술과 전자우편을 이용한 서버 푸시 기술 그리고 정보 필터링 기술 등을 결합하여 개발되며, 각 기술요소에 대한 구체적인 내용은 다음과 같다.

가. 정보검색기술

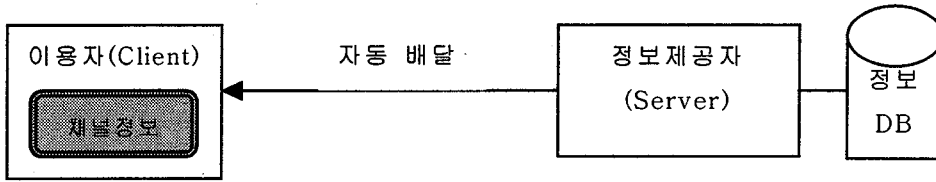
정보검색기술은 디지털 도서관 구현 등 주로 텍스트 정보를 대상으로 발전되었는데, 전형적인 정보검색시스템의 실행과정은 〈그림 1〉과 같다. 먼저, 색인(indexing)과정에서는 문서의 내용을 분석하여 색인정보를 생성하며, 보통 도치색인(inverted index)으로 구성된다. 검색과정은 이용자의 질의에 대하여 내부의 질의형식을 형성한 후 도치색인 정보와 검색엔진을 통해서 매칭되는 정보를 추출하고 가중치에 따라 순서화하여 출력한다.

정보검색 분야는 색인 & 요약화일, 도치 색인 파일 등의 색인구조와 검색모델, 질의확장 등 텍스트 검색기법 위주로 연구되어 왔으나(강현규 1998) 최근에는 인터넷과 멀티미디어 압축, 저장, 전송기능의 발전으로 멀티미디어 정보에 대한 요구가 증대되면서 내용 검색(Content-based Retrieval)분야에 대한 연구도 활발히 이루어지고 있다. 내용검색은 멀티미디어 정보의 특징을 추출하여 저장한 후, 그것에 기초하여 질의하는 검색방법으로서, 개발된 시스템의 예로는 QBIC(Query By Image Content), STAR 등이 있다.

나. 푸시기술

푸시는 이용자가 매번 요청하지 않더라도 원하는 특정정보를 이용자의 컴퓨터로 전송 받을 수 있도록 자동화한 기술로서 중앙의 메인컴퓨터에서 이용자의 컴퓨터로 원하는 정보를 밀어내는 원리이며, 그 구조는 〈그림 2〉와 같다.

푸시기술의 장점은 일일이 정보를 찾는 풀(Pull) 방식과는 달리, 이용자가 원하는



〈그림 2〉 푸시 서비스 구조

정보를 서버가 자동으로 제공함으로써 정보검색의 시간을 절약할 수 있고, 이용자의 취향에 따른 정보 제공이 가능하다는 것이다. 또한 웹이나 조직 내의 데이터베이스에 저장되어 있는 텍스트, 멀티미디어 데이터, 주식정보, 상품 카탈로그, 날씨 등의 다양한 정보를 원하는 형태로 가공하여 제공할 수 있다. 그리고 필요한 사람에게만 브로드 캐스팅함으로써 이용자가 일일이 체크해야 하는 불편을 해소할 수 있다(Pointcast network 2001). 현재 마이크로소프트, 넷스케이프, 마림바, 포인트캐스트 등에서 푸시기술을 선도하고 있는데, 개발된 서비스로는 마이크로소프트의 웹 캐스팅, 넷스케이프의 넷캐스터, 마림바의 캐스터넷(CasterNet) 등이 있다.

푸시기술의 적용분야로는 도서관이나 정보센터에서 회원제 또는 클럽식으로 운영하는 고품질 정보서비스나 기업간 정보교류를 위한 엑스트라넷 도구의 일부로 사용될 수 있다. 또한, 상업적 수단으로써 개별 이용자의 기호와 취미에 부합되는 유용한 정보의 지속적 제공으로 타겟 마케팅을 실현할 수 있다. 이외에도 행정기관, 교육기관, 각종 단체에서 관련된 사람

에게 공지사항, 홍보물, 교육 자료 등을 자동 전송하는데 사용함으로써 담당인력과 작업시간을 절약할 수 있을 것이다. 푸시기술은 인터넷 웹 표준문서 포맷인 HTML의 새로운 대안인 XML(Extensible Markup Language)과 함께 향후 인터넷 기술을 선도해 나갈 것으로 예상되며, 마이크로소프트의 CDF(Channel Definition Format), 넷스케이프의 MCF(Meta Content Framework Using XML), 자바 등의 기반 기술을 이용한 응용분야에도 확대되어 가는 추세이다.

다. 에이전트(Agent)

에이전트는 특정목적을 갖고 이용자를 대신하여 작업을 수행하는 자율적인 프로세스(Autonomous Process)로 정의할 수 있으며, 일반적으로 지식베이스와 추론기능을 보유하고 이용자나 자원 또는 다른 에이전트와의 정보교환과 통신을 통해 문제해결을 도모한다. 에이전트는 원래 인공지능의 한 분야로서 연구되어 오던 중, 80년대 말부터 분산협동 처리(Distributed Cooperative Processing), 에이전트간 통신(Inter-agent Communication) 등의 독자적인 에이전트 관련 연구가 시작되었고,

인터넷의 확산에 따라 응용분야가 무한히 확대되고 있다. 최근에는 초고속통신망과 연계하여 정보검색 뿐 아니라 온라인 쇼핑 등의 전자상거래 분야, 메시징과 같은 이동 컴퓨팅 분야에서도 에이전트의 이용이 활성화되는 추세이다. 에이전트가 다른 여타의 소프트웨어와 구별되는 점은 자율성(Autonomy), 이동성(Mobility), 지능성(Intelligence), 협동성 등의 특징을 갖는다는 것이다.

에이전트의 구성요소는 에이전트의 생성, 수행, 종료 등 일련의 작업을 조정하기 위한 제어지식 및 추론능력을 갖는 에이전트 엔진과 특정 응용분야의 작업수행에 필요한 영역지식(Domain Knowledge) 그리고 다른 에이전트에게 자신의 능력과 존재를 알리고 협력작업을 수행하기 위한 에이전트 통신 등 3가지로 구분할 수 있다.

에이전트와 관련된 언어에는 에이전트의 목적과 기능, 계획, 지식 등을 표현하는 에이전트 기술 언어(Agent Specification Language)와 에이전트를 프로그래밍하기 위한 언어인 에이전트 개발 언어(Agent Programming Language)가 있는데, 전자의 예로는 AOP, PLACA 등이 있고 후자의 경우는 대부분이 해석언어로서 객체지향 기법을 사용하여 시스템에 독립적인 에이전트를 개발할 수 있는 Java, Tcl/Tk, Telescript 등이 있다.

이러한 에이전트에 대한 연구는 정보검색 분야에서 가장 활발히 전개되고 있으며, 최근에는 전자상거래 분야에서 대리쇼핑,

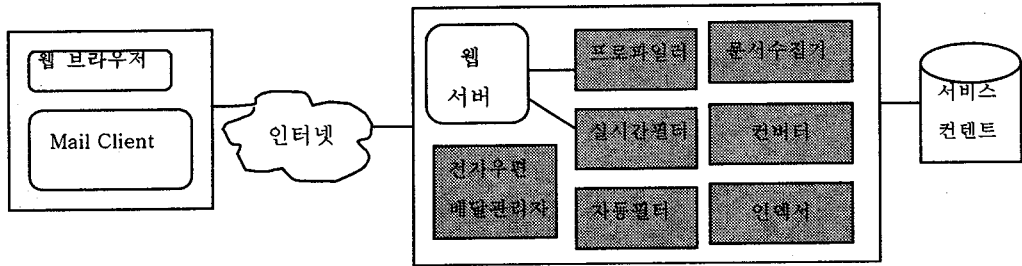
가격 비교, 계약, 교섭 등의 기능을 수행하는 에이전트의 개발에 대한 관심이 증가하고 있고, 인터넷 정보검색 분야에서는 WAIBA, WebAnts 등의 제품이 개발되었다. 또한 이용자에게 변경사항을 통지해주는 Mind-it과 같은 알람기능의 에이전트와 이용자 개인의 취향과 패턴을 반영하고 반복적인 작업을 대신해 주는 New Wave, Open Sesame 등과 같은 이용자 인터페이스 에이전트 등이 개발되어 있다(한국전자통신연구원 1998). 현재까지는 인터넷과 관련된 에이전트의 개발이 많은 비중을 차지하고 있으나, 향후 이동전화와 PDA 등의 발달로 이동 컴퓨팅을 지원하는 이동 에이전트의 중요성이 커질 것이며, 지식이 보강되어 보다 인텔리전트하고 보안성이 강화된 에이전트의 개발이 예상된다.

3.1.3.2 맞춤형 정보 서비스의 구조

일반적으로 맞춤형 정보 서비스의 시스템은 크게 7개의 서브시스템으로 구성되며, 각각 독립적으로 실행된다. 시스템의 구성도는 <그림 3>과 같고 각 모듈의 개별 기능은 다음과 같다.

가. 문서수집기(Web Robot)

정보 이용자의 기본적인 프로파일 정보와 함께 키워드 정보 및 문서의 형식 등의 검색조건이 설정되어진 환경파일을 참조하여 일정 주기로 지정된 웹 사이트로부터 HTML 파일을 가져와 로컬 파일

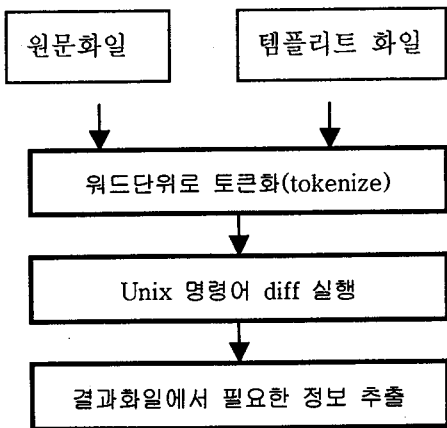


〈그림 3〉 맞춤형정보서비스 시스템 구성도

시스템에 원시화일 형태로 분류, 저장 기능을 수행하는 웹 로봇 시스템이다.

나. 컨버터(Converter)

문서수집기가 주기적으로 수집해 온 원시화일로부터 광고용 이미지, 메뉴 프레임과 같은 불필요한 정보를 제거하고 서비스에 필요한 정보만 추출하는 모듈로서, 색인의 효율을 높이고 저장 공간도 절약하며 결과 출력도 용이하도록 한다.



〈그림 4〉 컨버터 실행과정

구현원리는 〈그림 4〉에서 보는 바와 같이 각 문서타입에 따라 원문정보의 특징

을 분석해 템플리트 파일을 만들고, 원문과 템플리트 파일을 입력하면, 시스템 중에서 단어에 대한 비교작업을 수행하는 Unix의 비교명령어인 diff 명령어를 사용하여 해당 조건과의 차이를 보이는 정보를 걸러낸다. 즉, 컨버터 프로그램과 문서 자체의 일정한 형식을 담고 있는 템플리트 파일을 이용하여 원문정보로부터 불필요한 부분을 제거하고 필요한 정보만을 걸러낼 수 있다.

다. 인덱서(Indexer)

문서수집 시스템이 가져온 문서를 검색엔진을 이용해 주기적으로 자동 인덱싱하여 인덱스 파일 생성한다.

라. 프로파일러(Profiler)

프로파일러는 이용자 등록 및 탈퇴와 프로파일(이용자 ID, 패스워드, 전자우편 주소, 채널선택, 검색어, 구독방법, 최대 메일 개수) 작성 및 수정 기능을 제공하는 프로그램이다.

마. 실시간 필터(Real-time Filter)

이용자에게 개인용 맞춤형정보를 제공하

기 위해 실시간으로 검색해서 그 결과를 출력하는 기능을 수행한다.

바. 자동필터(AutoFilter)

주기적으로 모든 이용자의 프로파일을 참조해서 채널, 키워드 정보를 이용하여 검색을 수행한 후 그 검색결과에 대해 Document ID와 가중치를 이용자별로 지정된 결과 파일에 기록한다. 추후, 전자우편 배달 관리자에 의해 이 결과가 전자우편의 형태로 배달된다.

사. 전자우편 배달 관리자
(Email Delivery Manager)

자동필터가 이용자별로 필터링해 놓은 정보를 텍스트 메일의 형태로 전달하는 모듈이며 메일시스템과의 연동이 필요하다. 실행 시에는 이용자의 프로파일 정보 중 메일 주소, 배달방법, 최대 배달기사의 수, 검색유효기간 등을 참조한다.

이상과 같이 살펴 본 맞춤형 정보 서비스는 이용자가 자신의 프로파일 정보에 자신의 관심분야에 대한 키워드를 입력하면, 해당 에이전트는 키워드와 일치되는 단어가 포함된 문서를 찾아서 이용자의 전자우편주소로 전송하게 된다. 그러나 맞춤형 정보 서비스는 단순히 해당 키워드와 일치되는 모든 문서를 이용자에게 전송함으로써 이용자의 요구에 불필요한 문서도 상당수 포함되어 전송되어 질 수 있다. 따라서 여전히 정보과잉의 문제는 존재한다. 또한 단순히 이용자가 입력한 키워드

를 기준으로 적합성을 판별하기 때문에 해당 키워드가 분야별로 다른 의미를 가지는 경우에 이용자가 원하는 것과는 전혀 다른 문서가 검색될 수 있다는 한계점도 가지고 있다. 그래서 이러한 한계점을 극복하고 이용자의 정보요구에 보다 적합한 정보를 탐색하고 이를 제공하기 위해서는 보다 지능화된 서비스 시스템이 요구된다. 이를 위해 이용자의 프로파일 정보를 기반으로 한 정보탐색행위 등을 분석하여 정보의 적합성 여부를 판별할 수 있는 기능이 첨가된다면, 이용자의 요구에 보다 적합한 정보를 쉽고 편리하게 제공할 수 있을 것이다. 이러한 서비스는 현재 많은 분야에서 연구되고 있는 고객관계관리(CRM : Customer Relationship Management)의 개념을 기반으로 한 개인화 서비스 시스템을 출현하게 하는 계기가 되었다.

3.2 개인화 정보 서비스
(personalization information service)

3.2.1 출현배경 및 개념

CRM은 마케팅 기법의 하나로 기업중심의 매스 마케팅적인 접근이 아니라 고객 개인의 기호나 구매이력 등에 기초하여 그 요구를 이해하고 고도의 전문지식을 바탕으로 개인의 요구에 맞는 상품 또는 서비스를 제공한다는 고객 중심의 직접적인 마케팅적 접근법이다(사와노보리 2000). 인터넷의 출현은 이러한 마케팅

개념에 있어서도 새로운 변화를 요구하는 직접적인 계기가 되었다. 실제 인터넷은 쌍방향 커뮤니케이션이 가능한 상호작용 매체로써 세분화된 이용자 중심의 마케팅 관점이 반영된 일대일 마케팅(One to One Marketing) 전략으로 환경을 변화시켰다. CRM 기법의 응용에 따른 마케팅 관점의 변화는 개인화(Personalization)라는 새로운 요구들을 표출시키고 있으며, 이와 같은 요구사항은 도서관이나 정보센터와 같은 정보제공기관에서도 동일하게 적용되고 있다.

1997년 D. Peppers와 M. Rogers에 의해 처음으로 소개된 개인화 서비스는 고객을 동질화 된 시장개념으로 보고 시장점유율을 목표로 삼는 기존의 마케팅과는 달리 고객을 각각 다른 욕구를 가진 독특한 존재로 보고 있다는 점이다. 이러한 관점의 변화는 목표시장 세분화 과정에서 개인화를 통해서 한명의 고객을 하나의 시장으로 인식하는 고객 점유(Customer Share) 개념을 이끌어 내었다. <그림 5>는 개인화 서비스의 시장 세분화 과정을 보여주고 있다.

본고에서 사용하는 개인화(Personalization)라는 용어는 웹사이트 개인화를 의미한다. 웹 사이트에서의 개인화는 일련의 가치교환 과정이다. 즉, 이용자가 자신의 선호, 관심, 이용내역, 구매이력, 접속내역 및 빈도 등과 같은 정보를 웹에 적, 간접적으로 제공하면 웹사이트는 이용자가 제공한 정보를 토대로 해서 적합

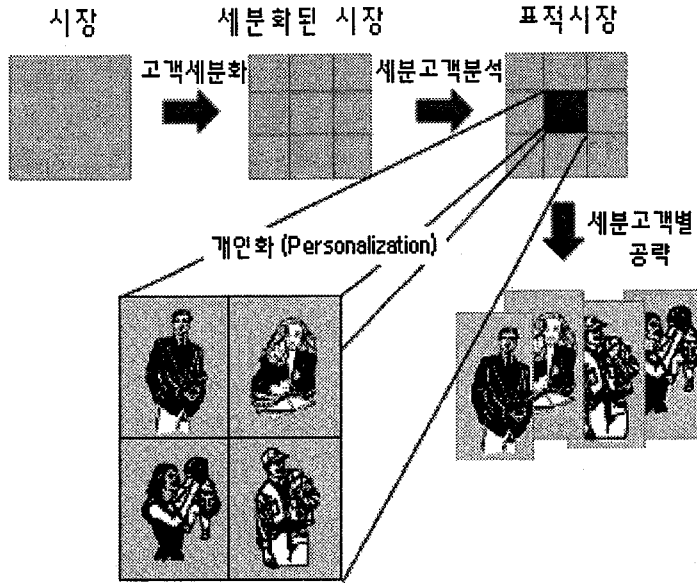
한 정보를 제공한다. 일반적인 관점에서 개인화라는 용어가 갖는 의미는 이용자의 요구에 적합한 정보를 추출하고 이를 신속하게 제공하는 일련의 결합된 방법을 개인화라고 정의할 수 있다(Alexander Pretschner, Susan Gauch 1999).

이상 설명한 개인화 정보 서비스의 개념을 도서관 관점에서 설명한다면, 기존 맞춤정보 서비스는 제한된 이용자 프로파일 정보를 토대로 개인의 취향이나 수준 등에 관계없이 단순히 이용자의 관심분야와 일치하는 정보를 제공하는 것에 중점을 두었으나 개인화 정보 서비스는 맞춤정보 서비스 보다 발전된 형태로 이용자의 정보탐색행위나 유사한 기호를 가진 이용자 그룹의 선호도 등을 다면적으로 추적, 분석하여 개인별로 성향에 맞는 최적의 정보서비스를 제공해 주는 것이라고 할 수 있다.

3.1.2 필요성

개인화는 정보 제공자와 정보 이용자 양측의 필요에 의해 발생된다고 할 수 있다. 정보 이용자 측면에서 인터넷 이용자들은 무수히 많은 정보 가운데 자신이 필요로 하는 정보만을 단시간에 얻을 수 있다는 점에서 개인화의 필요성이 요구되는 첫 번째 조건이다. 정보 제공자 측면에서는 정보 이용자가 필요로 하는 효과적인 마케팅을 구사할 수 있는 기반을 마련하기 위해서 개인화가 필요하다.

도서관이나 정보센터에서 기존의 정보



〈그림 5〉 개인화 서비스의 시장 세분화 전략(자료출처 : I-biznet)

접근 방식은 주로 ‘이용자에게 무엇을 해 줄 것인가’에 대한 문제들을 도출하고 그에 따라 대응하면 되었다. 하지만 인터넷에서의 이용자 중심은 이용자가 주체가 된다. 도서관과 정보센터는 이용자를 위해서 무엇을 해주는 것이 아니라 이용자가 원하는 대로 움직여야 하는 즉, ‘이용자가 필요로 하는 것을 얻어가는 대상’인 것이다. 이런 관계 속에서 진정한 이용자의 만족이 이루어 질 수 있는 수단이 바로 개인화가 고도로 적용된 개인화 정보 서비스 시스템일 것이다.

개인화 정보 서비스 시스템은 이용자가 제공한 정보나 로그데이터 또는 이용자의 선호도와 웹상에서 이용자의 정보탐색 패턴 등을 발견하여 개인별로 적절한 정보

를 추천해 주는 자동화된 정보 필터링 시스템을 말한다. 예컨대, 이용자에 대한 인구 통계학적 정보, 가장 이용 빈도수가 높은 정보에 대한 이용자의 활용패턴, 이용자 선호도 등을 데이터마이닝(Data mining) 기법 등에 의해 분석을 하고, 그 결과를 토대로 이용자가 구매하고 싶은 상품을 쉽게 찾을 수 있도록 이용자만을 위한 개인화된 웹 페이지를 만들어 개인화 서비스를 이용자에게 공급하는 시스템인 것이다(황성희 2001).

따라서 도서관과 정보센터에서도 보다 적극적이고 능동적으로 이용자의 요구에 적합한 정보를 보다 신속하게 제공하기 위한 일련의 노력들이 필요하며, 이러한 노력의 일환으로서 이용자의 개인차에 따

른 정보요구 특성을 반영한 개인화 정보 서비스 시스템을 구축하는 것이 필요하다.

3.1.3 개인화 서비스의 적용 현황 및 효과

Cyber Dialogue사의 조사에 따르면, 이미 미국 인터넷 이용자의 29%가 웹사이트 중 개인화된 페이지를 이용하고 있으며, 88%가 개인화된 웹 페이지를 보다 선호하는 것으로 밝혀진 바 있다. 이제 우수한 웹사이트가 가지고 있어야 할 기본적인 조건으로 개인화를 꼽고 있을 정도로 이미 개인화는 웹사이트의 중요한 요소로 자리 잡아가고 있다. 전자상거래 업체들은 개인화 서비스 시스템 구축이 기업의 수익증대를 가져다준다는 실질적인 통계 자료를 발표한 바 있다. 미국의 Jupiter Communication에서 전자상거래 분야의 상위 25개 업체를 대상으로 한 조사에 의하면, 40%의 업체가 이미 개인화 서비스를 제공하기 시작하였고, 나머지 대부분도 1년 내에 이 서비스를 도입할 예정이라고 응답했다. 또한 Jupiter는 개인화를 도입한 24개의 전자 상거래 업체에서 고객이 47% 증가하는 동시에 매출액은 52% 증가했다는 내용을 발표하였다(www.trendsetters.com).

오늘날 Yahoo!, Lycos, Excite와 같은 거의 모든 검색 엔진들이 My Yahoo!나 My Excite와 같은 이용자 중심의 맞춤 정보 기능을 가지고 있다. 또한 Amazon을 필두로 한 전자상거래 업체들도 이용자의 개인 취향에 따라 자신의 페이지를

구성하고 선별적인 정보만을 받을 수 있도록 시스템화되어 있다. 그리고 몇몇 사이트는 이용자의 구매기록과 취향에 맞는 제품을 이용자들에게 자동으로 추천해 주는 시스템을 갖추고 있다.

부분적으로 도서관 및 정보센터 등에서도 유사한 서비스를 제공하고 있다. 국내의 경우 연세대학교에서 My Digital Library 형태로, 서울대학교에서는 My Library로 개발, 운영 중에 있고, 현재 많은 대학도서관 및 전문도서관에서도 이러한 서비스를 제공하거나 계획하고 있다. 국외의 경우에는 미국의 듀크대학교의 의학도서관에서는 Ovid를 통해 유사한 서비스를 제공하고 있으며, 코넬대학교의 MyLibrary 서비스, 워싱턴대학 도서관의 My Gateway 서비스 등이 있다. 이러한 유형의 서비스를 일반적으로 개인화(Personalization) 또는 맞춤화(Customization) 서비스의 일종이라고 말할 수 있다.

3.1.4 기술적 기법

개인별 개인화 정보 서비스를 제공하기 위해서는 먼저 개인에 대한 정보를 수집한 후 관심분야를 다면적으로 분석하는 기술과 가장 적절한 서비스를 찾는 기술 그리고 실제로 적합한 정보를 제공할 수 있도록 하는 기술 등이 필요하다.

개인에 대한 성향분석은 물론 가장 적절한 서비스를 도출하는 기술은 개인화 서비스의 성패를 좌우하는 핵심적인 기술이라고 할 수 있다. 이 기술은 고객에 대

한 각종 정보를 기반으로 고객이 원하는 서비스가 무엇인지를 찾아내는 메커니즘을 시스템화 한 것이다. 개인에 대한 분석을 기반으로 적절한 개인화 서비스를 결정하기 위해 사용 가능한 기술적 접근법은 다음과 같이 설명할 수 있다.

가. 데이터베이스 주도형

고객 데이터베이스를 기반으로 간단한 애플리케이션을 통해 개인화 하는 방법이다. 이러한 접근법에서는 고차원적이고 비즈니스적인 노하우를 요구하지는 않으므로 시스템을 구현하는데 그리 어렵지는 않다. 따라서 소형이나 중형 사이트에 적용될 수 있는 방법이라 하겠다. 이 접근법을 통해 가능한 개인화 서비스로는 방문자가 찾아 왔을 때 쿠키나 등록ID를 체크한 후 해당 방문자의 선호도를 추출할 수 있다. 이와 같은 결과를 추출하기 위해 방문자별로 자기가 원하는 방식으로 간단한 질의문 또는 키워드 정보를 작성할 수 있도록 지원하는 기법 역시 이러한 범주에 속한다고 할 수 있다.

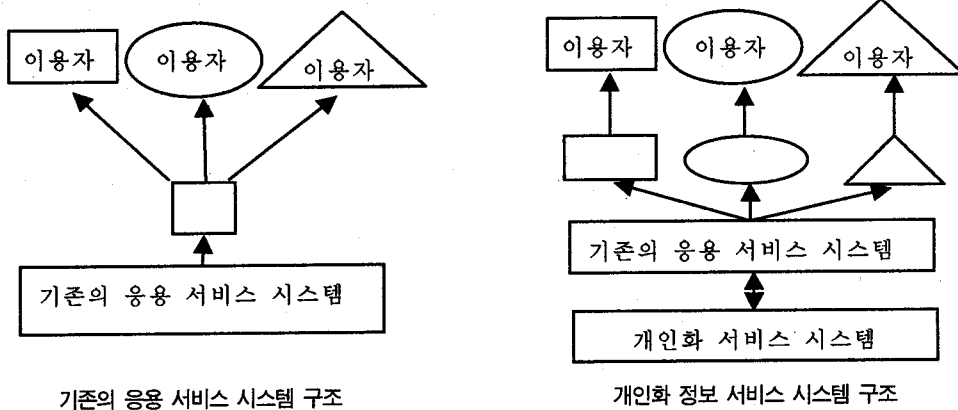
나. 지능형 에이전트

웹 사이트에서 방문자로 하여금 원하는 정보를 쉽게 찾을 수 있도록 하는 소프트웨어이다. 이는 일종의 인공지능기술을 활용하여 방문자의 선호도와 방문자의 사이트 활용 정보를 토대로 사이트 내에 존재하는 많은 정보를 방문자의 관심분야에 맞도록 제공한다. 지능형 에이전트 기술

을 활용할 경우 이용자가 웹 사이트에서 그 동안 조회한 정보와 이용자의 특성을 기반으로 현재 구매할 가능성이 높은 제품을 소개하거나, 이 고객이 특정한 제품을 검색할 경우 가장 마음에 들 것으로 보이는 제품을 안내하는 기능을 구현할 수 있다.

다. 공동 필터링

이용자의 프로필과 비슷한 유형의 이용자 집단에 대한 선호도와 관심사를 추정하여 서비스하는 방식이다(Baillus 1998). 이용자가 자신의 프로필을 제공하면 비슷한 성향의 이용자가 지닌 행동양식을 분석하여 얻어낸 기준을 적용해서 적당한 제품이나 서비스, 정보를 제공하는 형태로 동작하게 된다(Krypis 2000). 이 기법은 정보 이용자에 대한 초기 정보가 부족한 경우 가장 적절하게 이용될 수 있다. 즉, 정보 이용자가 보여주는 정보탐색행위나 정보탐색 후의 피드백에 대한 정보가 충분하지 않은 경우 해당 정보 이용자와 비슷한 반응을 보여주는 이용자들의 선호도를 기준으로 정보를 제공해 주는 방법이다. 따라서 이용자들의 유사성을 계산하고, 유사한 이용자 그룹을 형성해야 한다. 각각의 이용자는 유사성을 판별하기 위해서 정보에 대한 선호도를 기본 요소로 삼는다. 궁극적으로 이러한 기법은 이용자의 주관적인 선호도의 평가에 기반하기 때문에 예상하지 못한 새로운 정보를 제공할 수 있다는 장점을 가지고



〈그림 6〉 응용 서비스 시스템과 개인화 정보 서비스 시스템 구조의 차이점

있다. 공동필터링 기법은 도서관이나 정보센터에서 개인화된 정보 서비스를 제공하는데 가장 적절한 방법이라고 할 수 있다.

라. 전문가/를 기반형

이미 구축된 경험과 지식을 바탕으로 세밀한 추론에 근거하여 서비스를 수행하는 방법이다. 이 접근법은 이용자들의 행동을 통해서 지속적으로 새로운 룰을 생성하거나, 학습을 통하여 이미 존재하는 룰을 변형하기도 하면서 개인화 방식을 발전시켜 가게 된다.

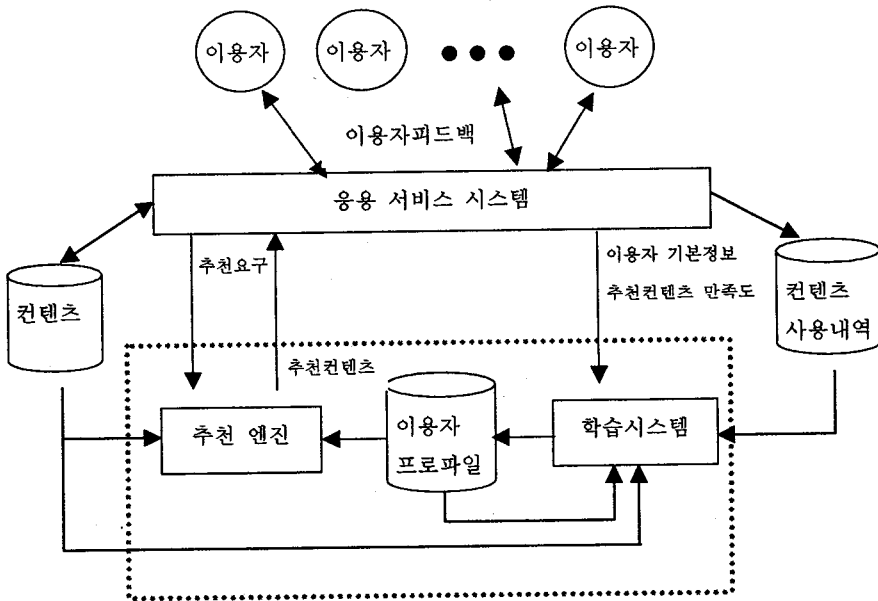
이상의 방법들이 개인화 정보 서비스 시스템을 구현하기 위한 기술적 접근 방법들이다. 물론 아직 시스템에 충분히 반영되지 못한 부분들도 상당히 많기 때문에 실제 부분적으로만 구현되고 있는 단계이다. 그러나 개인별 정보 이용자들이 보여주는 탐색행동은 매우 다양함으로 이용자들의 선호도에 따라 가장 적절한

정보를 제공할 수 있는 개인화된 정보 서비스에 대한 많은 논의와 연구가 필요하다. 이를 위해 본고에서는 보다 진보된 개념으로써 학습시스템에 기반을 둔 개인화 정보 서비스 시스템을 제안하고자 한다. 학습시스템은 정보 이용자에 대한 기본 프로파일 정보와 함께 정보 이용자가 정보탐색을 하는 행동양식 및 정보 이용자로부터의 피드백 등을 입력 값으로 지속적인 학습과정을 통해 정보 이용자에게 가장 적절한 정보를 제공하기 위한 방법이라고 할 수 있다.

4 학습시스템에 기반한 개인화 정보 서비스 시스템

4.1 시스템 특성

기존의 인터넷 응용 서비스 시스템은 이용자 개인별로 콘텐츠 이용의 특성을



〈그림 7〉 제안한 개인화 정보 서비스 시스템 구조

고려하지 않기 때문에 모든 이용자에게 동일한 콘텐츠(서비스)를 제공한다. 개인화 정보 서비스 시스템은 여러 가지 방법을 이용하여 이용자의 특성을 분석한 후 각각의 이용자에게 적합한 콘텐츠를 제공함으로써 응용 서비스 시스템의 이용자 만족도를 향상시킬 수 있다. 〈그림 6〉은 일반적인 응용 서비스 시스템과 개인화 정보 서비스 시스템을 적용한 시스템 구조의 차이점을 보여 준다.

4.2 시스템 구성

〈그림 7〉은 개인화 정보 서비스 시스템의 전체구조 및 개별 요소간의 구조관계를 보여 주고 있다. 응용 서비스 시스템과 연동하는 개인화 정보 서비스 시스템

은 크게 이용자 프로파일 및 추천 엔진 그리고 학습시스템 등 3가지 모듈로 구성된다. 그림에서 보는 바와 같이 각 모듈간의 상호 순환과정을 통해 이용자에게 적절한 정보를 제공하기 위한 탐색과정을 살펴보면 다음과 같다.

- ① 먼저, 정보 이용자는 시스템에 자신의 기본정보를 입력하게 된다. 기본정보로는 이용자의 개인 신상정보와 관심분야 및 키워드가 될 수 있다. 그리고 추가 이용자 프로파일을 생성하기 위해서 학습시스템에 입력되는 입력값은 이용자가 입력하는 기본정보를 포함하여 이용자의 정보탐색행위 및 개인화 정보 서비스 시스템에 의해서 추천된 정보에 대한 만족도와 같은 이용자 피

드백 등이 해당된다.

- ② 이러한 입력값은 학습시스템에 의해 지속적인 학습과정을 거쳐 이용자의 이용성향을 분석하여 적합성 있는 정보를 추천하기 위한 출력값으로 추출되어지며, 출력값은 이용자의 프로파일정보가 되어 해당 프로파일정보를 갱신한다.
- ③ 이렇게 생성되고 갱신되어지는 이용자 프로파일 정보는 이용자에게 가장 적절한 정보를 제공하기 위한 추천엔진의 입력값으로 사용되고, 이용자 프로파일에 가장 적합한 정보가 추천엔진을 통해 정보 이용자에게 추천된다.
- ④ 이용자의 추천 콘텐츠에 대한 만족도 및 이용성향은 다시 학습시스템의 입력값으로 입력되어지고, 이를 통해 이용자의 프로파일이 계속적으로 갱신되어 진다.

따라서 본 시스템에 의해 추천되는 정보는 이용자의 정보 이용행태를 분석하는 기초자료로 재활용하기 때문에 이용자의 관심분야에 보다 적합한 이용자 프로파일을 구성할 수 있으며, 이를 통하여 이용자에게 불필요한 정보는 제거하고 개인화된 적합정보를 제공할 수 있다는 장점이 있다. 즉, 이용자에 대한 정보추천은 지속적인 이용자의 피드백을 통한 이용자 프로파일의 갱신이 이루어지므로 보다 정확한 정보추천이 가능하다.

4.2.1 이용자 프로파일

본고에서 제안한 이용자 프로파일은 이용자의 콘텐츠에 대한 이용 특성을 반영한 DB로서, 학습시스템에 의해 생성되는 최적의 이용자 정보를 포함하고 있다. 이용자 프로파일은 이용자의 입력으로부터 얻어지는 이용자 기본정보 부분과 학습에 의해 계속 변경되는 부분으로 구성된다. 학습에 의해 변경되는 부분은 이용자에게 제공되는 서비스 방법에 따라 이용자 별 키워드 벡터와 그룹별 키워드 벡터 그리고 그룹별 키워드 리스트로 구분되며, 서비스에 따라서 적합한 프로파일이 사용된다. 학습시스템은 다양한 입력 소스로부터 여러 형태의 데이터를 받아서 이용자에게 적합한 프로파일을 생성해 낸다. 이러한 입력요소로는 이용자의 기본 정보, 추천 콘텐츠에 대한 피드백(Feedback), 콘텐츠 사용 내역(웹서버 및 콘텐츠서버 로그 등), 사이트의 구조 등이 될 수 있다. 특히, 이용자의 정보이용 행태를 파악할 수 있는 대표적인 방법은 이용자가 직접 등록한 자신의 프로파일 정보와 온라인 행동 데이터 분석기법 등이 있다. 웹상에서의 이용자 행동을 분석하는 온라인 행동 데이터는 기술적으로 다음 세 가지 방법으로 얻을 수 있다.

첫 번째, 로그파일(Log File) 방식은 온라인 행동 데이터를 얻는 일반적인 방법으로 안정적인 데이터의 확보가 가능하나 배치(Batch)로 작업이 이루어져야 하기 때문에 실시간 분석이 불가능하다는 단점

이 있다.

두 번째 방식은 서버추가(Server Add-in) 방식으로서 웹서버의 내부 모듈로서 존재하여 정보가 파일형태로 남겨지기 전에 수집된다. 이 방법은 정보를 가져오는 모듈이 서버에 통합(integrated)되었기 때문에 실시간 분석이 가능하여 신속한 의사 결정을 하는데 유리하다. 그러나 서버에 붙어서 작동하기 때문에 웹서버의 성능에 영향을 줄 수 있고, 특히, 성능이 떨어지는 추가모듈(Add-in Module)인 경우, 서버를 손상(Crash)시킬 수도 있음을 유의해야 한다. 이 방식을 채택하려면 추가모듈이 정지하였을 경우에 대비하여, 로그 파일(Log File)에 근거한 백업 시스템을 갖추는 것이 필요하다. 이 방식은 웹 행동 정보의 실시간 취합, 분석 및 리포팅이 가능하다는 점 외에도 자체적인 쿠키(cookie) 관리 기능을 이용하여 익명(anonymous) 방문객의 정보 취합과 방문객 관리가 가능하다는 장점이 있다.

세 번째, 패킷 스니퍼스(Packet Sniffers) 방식은 웹 서버의 네트워크 세그먼트(Network Segment)에 위치하여, 지나가는 TCP/IP 패킷을 직접 분석하는 방법이다. 이 방법은 서버가 인식할 수 없는 정보를 분석할 수 있다는 점과 사이트 모니터링이 가능하다는 장점을 가지고 있다. 이용자가 사이트에 접속하여 웹서핑 중 페이지가 완전히 로드되기 전에 접속을 끊은 경우, 웹서버에서는 요청된 페이지를 전송하였다는 자체만 인식할 뿐 이용자가

그 페이지를 완전히 열람했는지에 대한 사실은 정확하게 파악할 수 없다. 패킷 스니퍼스 방식은 패킷의 송수신 여부를 확인하므로 이와 같은 사실을 확인할 수 있으며, 네트워크를 지나가는 정확한 바이트 수를 알 수 있어 자원 사용에 대한 정확한 정보를 얻을 수 있다.

그러나 현실적으로 이 방식은 거의 사용되지 않는다. 그 이유는 암호화되어 있는 패킷을 해독할 수 없고, 해독하기 위해서는 암호화기가 필요하며, 항상 서버의 네트워크 세그먼트 단에 설치되어야 한다는 점에서 호스팅 환경에 부적절하고 고가의 장비가 추가로 필요하다는 단점이 있다. 무엇보다도 이용자의 분석 데이터는 대부분 웹 서버에서 획득 가능하다는 점에서 패킷 스니퍼스 방식보다 로그 파일이나 서버 애드인 방식을 더 선호하고 있다.

본고에서는 이용자의 정보이용성향에 대한 분석을 위해 웹서버 자체에 대한 부하가 가중되는 서버추가 방식이 아닌 배치작업에 의해 획득한 데이터를 웹로그에 의해서 이용자의 성향을 분석하는 웹로그 방법을 적용하고자 한다. 즉, 도서관이나 정보센터에서 이용자의 프로파일을 구성하고, 이를 통하여 정보를 제공하는데 있어서 실시간 분석이 크게 요구되지 않으므로 서버에 많은 부하를 제공하는 서버 추가 방식 보다는 배치작업을 통한 웹로그 분석 방법이 보다 효율적이라고 할 수 있다. 따라서 웹로그 분석 방법을 사용하여 이용자의 성향을 분석하고 이에 따른

개인화작업을 수행하고자 한다. 이러한 개인화 작업을 위해서는 각 요소별 가중치를 설정하여 가중치에 따른 유사도를 분석함으로써 적합성 여부를 판별하는 기준이 된다. 가중치에 대한 기준요소는 아래와 같다.

가. 이용자 기본 정보

이용자의 기본적인 인적사항(아이디, 이름, 성별, 나이 등)이다. 응용서비스 시스템에 의해서 데이터를 입력받아 학습시스템을 통해서 이용자 프로파일에 저장된다.

나. 이용자별 키워드 벡터

이용자의 관심 분야를 키워드로 하여 가중치를 가지는 벡터이다. 이용자의 입력에 의하여 초기화 된 후 학습시스템에 의하여 값이 변한다. 키워드 벡터의 초기 값을 구하는 방법으로는 이용자가 직접 입력할 수도 있고, 비슷한 성향을 가지는 이용자 그룹의 프로파일로부터 구할 수 있다.

다. 그룹별 키워드 벡터

비슷한 성향을 가지는 이용자 그룹을 대표하는 키워드 벡터이다. 새로운 이용자의 초기 프로파일을 생성하거나 이용자의 분류(Clustering)에 사용될 수 있다.

라. 그룹별 키워드(URL) 리스트

이용자 그룹이 선호하는 키워드(URL)의 리스트를 가중치 순으로 정렬한 리스트이다.

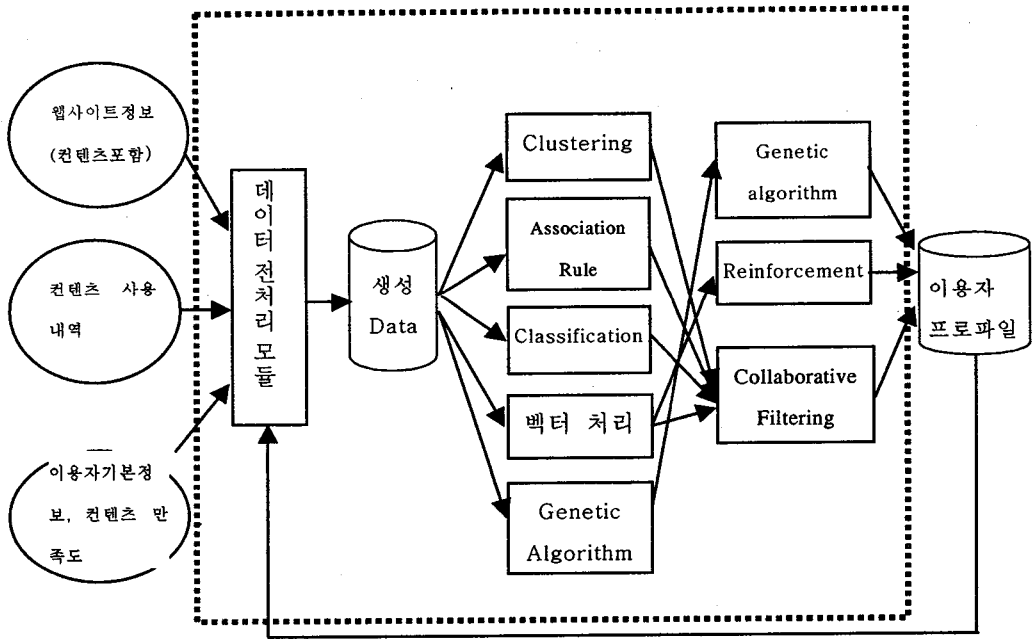
트이다.

4.2.2 학습시스템

학습시스템은 이용자의 초기 프로파일을 생성하고 정보이용 성향을 프로파일에 반영하는 시스템이다. 학습시스템은 한가지로 고정되어 있지 않고 서비스 특성에 따라서 그에 적합한 학습방법이 적용될 수 있도록 구성되어야 한다. 학습시스템이 이용자의 프로파일을 현행화하기 위해서는 이용자의 기본 정보, 정보사용내역 외에도 추천된 정보에 대한 이용자의 만족도 등이 학습시스템의 입력 값으로 사용될 수 있다. 여기서 이용자의 만족도는 정보이용자의 피드백으로 측정된다. 이처럼 다양한 입력 데이터가 학습시스템에서 사용될 수 있도록 하는 데이터 전처리 모듈이 학습시스템의 구성요소로 포함되어야 한다.

개인화 정보 서비스 시스템을 구축하는데 본 연구에서는 다음과 같은 학습 방법을 적용하고자 한다.

- 유전알고리즘(Genetic Algorithm) : 이용자로부터 입력받은 정보를 이용하여 초기 프로파일을 만드는데 사용
- 가중치보강(Reinforcement) : 키워드 벡터로 표시된 이용자 프로파일에 대해 이용자 피드백을 사용하여 가중치를 변경하는 방법으로, 이용자 프로파일과 가장 비슷한 콘텐츠를 추천하고자 할 때 사용한다. 지속적인 가중치 보강을 통해 이용자의 정보요구에 가



〈그림 8〉 제안한 학습시스템 구성도

장 적합한 정보를 추출할 수 있다.

- 공동필터링(Collaborative Filtering) : 이용자와 비슷한 특성을 갖는 유사한 그룹을 찾고, 그 그룹 내의 사용자들의 컨텐츠 이용 성향을 분석하여 이용자 그룹의 선호도에 의해 개인화된 정보를 추천한다.

그리고 상기 학습 방법을 구현하기 위해서는 다음의 기능이 수행되어야 한다.

- 그룹화(Clustering)
- 연관규칙(Association Rule)
- 분류(Classification)
- 유전 알고리즘(Genetic Algorithm)
- 벡터 처리

〈그림 8〉은 앞에서 언급한 여러 가지

학습방법과 학습에 필요한 요소 및 데이터 전처리 모듈로 구성된 학습시스템을 보여 주고 있다. 먼저, 데이터 전처리 모듈을 통해 이용자의 프로필을 생성하기 위해 입력되는 초기 입력값으로는 웹사이트 정보, 컨텐츠 사용내역, 이용자의 기본 정보 및 이용자로부터의 피드백 등이 여기에 해당되며, 데이터 전처리 모듈에서 생성되어진 1차 출력값은 다양한 가중치 처리, 그룹화 및 분류작업, 연관규칙, 벡터 처리 등을 통해서 2차 출력값이 생성되고, 2차 출력값은 보다 정확한 사용자 프로필을 생성하기 위해 유전알고리즘, 가중치 보강, 공동필터링 방법을 거쳐 최종 사용자 프로필을 생성하게 된다. 또한 생성되어진 사용자 프로필은 다시

전처리 모듈의 초기 입력값으로 입력되어 학습과정을 통해 갱신되어 진다. 이렇게 반복적인 학습과정에 의해서 생성되는 사용자 프로파일은 이용자에게 보다 개인화된 정보를 제공할 수 있는 기반이 된다.

4.2.3 추천 엔진

추천 엔진은 응용서비스 시스템으로부터 추천요구를 받아서 사용자 프로파일과 가장 적합한 콘텐츠를 찾아서 응용서비스 시스템에 전달한다. 추천 엔진은 온라인 상에서 추천요구를 받아서 처리해야 하기 때문에 응답 속도의 제한을 받는다. 따라서 응용 서비스 시스템의 추천 요구에 앞서 미리 추천 콘텐츠를 구축해 놓은 후 추천요구가 있을 경우 즉시 추천 콘텐츠를 전달하는 방법도 고려할 수 있다. 추천요구에는 어떠한 방식으로 추천을 할 것인가에 대한 정보 및 추천 콘텐츠의 수를 명시한다. 추천 엔진은 다음의 세 가지 추천 방식을 제공한다.

가. 개인별 추천

이용자 개인의 성향과 가장 잘 부합되는 콘텐츠를 추천한다. 이는 이용자별 키워드 벡터로 표현된 사용자 프로파일과 콘텐츠의 키워드 벡터를 비교하여 가장 근사도가 높은 콘텐츠를 추천하게 된다. 벡터 사이의 근사도를 측정하기 위해서는 Cosine Similarity 등의 방법을 사용한다.

나. 그룹 추천

그룹 추천은 이용자와 성향이 비슷한

그룹 내의 이용자들이 선호하는 콘텐츠를 추천한다. 이를 위하여 그룹별로 가중치 순으로 정렬된 키워드(URL) 리스트를 이용한다.

다. 혼합형 추천

이 방법은 우선 그룹 추천 방법을 이용하여 이용자와 성향이 비슷한 그룹의 키워드 리스트를 구한 후에 개인별 추천 방식을 사용한다.

이상과 같이 학습시스템의 각 구성요소의 기능이 상호협력을 통해 이용자별 성향에 대한 종합적인 분석을 수행해서 최종 이용자 프로파일을 구성하고, 이를 통하여 개인별로 이용자에게 적합한 정보를 쉽고 편리하게 제공할 수 있다. 따라서 본고에서는 데이터의 전처리 과정과 함께 학습시스템에 의한 이용자 개인의 특성을 반영하여 가장 적절하게 개인화된 정보를 추출할 수 있는 효율적인 시스템 요소 및 기법을 제시하였다.

5 결 론

인터넷의 급속한 발전은 정보 관리자 관점에서 이용자 중심으로 정보환경을 변화시키는 계기가 되었다. 도서관과 정보센터에서도 이러한 일련의 변화에 적극적이고 능동적으로 대응하기 위해 선택적 정보배포 서비스의 일종인 맞춤형 정보 서비스 (Customized Information Service)를 제공하게 되었다. 최근에는 활발히 연구

가 되고 있는 고객관계관리(CRM) 개념을 이 시스템에 접목하여 개인별로 이용자 정보요구의 성향에 맞추어 적합한 정보를 제공할 수 있는 개인화 정보 서비스가 등장하게 되었다. 일부 도서관과 정보센터에서는 기존의 정보 서비스 형태보다 발전된 형태인 개인화 정보 서비스를 적용하여 이용자의 만족도를 향상시키려는 노력을 시도하고 있다. 그러나 발전된 형태의 개인화 정보 서비스를 제공하기 위해서 무엇보다도 해결해야 할 문제는 개인화에 필요한 가장 필수적인 개인 정보를 수집할 수 있는 능력이다. 즉, 이용자의 신상정보, 이용자의 선호도, 정보이용 행태 등 다양한 개인정보를 수집, 분석하여 이를 바탕으로 이용자에게 가장 적합한 정보를 추출해야 한다.

따라서 본고에서는 개인화 정보 서비스 시스템에서 핵심적인 요소인 사용자 프로파일을 생성하는데 필수적인 정보수집을 위해 이용자의 입력으로부터 얻어지는 기본정보 뿐만 아니라, 이용자 개인의 정보요구 특성을 다양한 기법을 통해 심층적으로 분석해서 이용자에 관한 유용한 정보를 획득할 수 있도록 학습시스템에 기반을 둔 개인화 정보 서비스 시스템의 구조를 제안하였다. 학습시스템은 다양한 입력 소스로부터 이용자와 관련된 여러 형태의 데이터를 받아서 이용자에게 적합한 프로파일을 생성하는 기능을 수행한다. 제안한 학습시스템은 유전알고리즘(Genetic Algorithm), 가중치 보강, 공동필

터링의 방법을 적용하였고 이 학습방법을 수행하기 위해서 그룹화, 연관규칙, 분류, 백터처리 등의 요소로 구성되어 있다. 이러한 일련의 학습과정을 통해서 이용자의 정보이용 행태를 분석할 수 있는 기초자료로 활용하여 이용자 프로파일을 지속적으로 갱신하고, 이를 통해서 이용자의 관심분야에 적합한 개인화된 정보를 주기적으로 제공할 수 있다.

본 연구에서 제안한 학습시스템에 기반을 둔 개인화 정보 서비스 시스템이 도서관이나 정보센터에서 이용자의 다양한 정보요구 형태를 추적, 분석하여 광범위하게 반영함으로써 한층 더 진전된 서비스를 제공할 수 있는 토대가 마련되기를 기대한다. 향후 이용자의 성향을 분석하는데 따른 각 요소별 가중치 설정방법을 구체적으로 제시하고, 학습알고리즘 및 개인화 기술을 적용하여 실제 서비스가 될 수 있도록 시스템을 구현하고 이에 대한 평가 작업도 병행되어야 할 것이다.

참 고 문 헌

- 강현규, 박세영. 2001. 정보검색. 『한국정보처리학회지』, 5(5):37-47.
- 김은주. 2003. 『개인화 서비스를 위한 유사도를 이용한 신경망 협력적 여과』. 석사학위논문, 숭실대학교 대학원.
- 김학균. 1999. 『문서의 구조기반 저장 및 검색과 사용자 기호를 반영한 정

- 보 추천시스템』. 석사학위논문, 연세대학교 대학원.
- 김현희, 구내영. 2002. 맞춤형정보서비스를 위한 MyCyberLibrary 모형설계와 평가에 관한 연구. 『정보관리학회지』, 19(2) : 132-157.
- 사와노보리 히데야키. 2000. 『e-CRM 마케팅』. 서울: 국립증권경제연구소.
- 이성희. 2001. 『인터넷 서비스의 개인화가 고객만족에 미치는 영향에 관한 연구』. 석사학위논문, 연세대학교 대학원.
- 조현진. 2002. 『웹 개인화(Web Personalization)가 웹 사이트 재방문 의도에 미치는 영향에 관한 연구』. 석사학위논문, 한국외국어대 경영정보대학원.
- 한국전자통신연구원. 1998. 인터넷 에이전트와 로봇. 『ETRI 주간기술동향』, 867: 45-47.
- 황성희, 김영지, 이미희, 우용태. 2001. 인구조통계학적 특성에 따른 협동적 필터링 알고리즘의 추천 효율 분석. 『한국데이터베이스 학회 춘계 논문집』, 362~368.
- Billsus, Daniel. and Pazzani, Michael J. 1998. *Learning Collaborative Information Filters*. University of California Press.
- Karypis, George. 2000. *Evaluation of item-based Top-N Recommendation Algorithms*, University of Minneapolis Press.
- Lakos, Amos and Gray, Chris. 2000. "Personalized Library Portals as an Organizational Culture Change Agent." *Information Technology and Libraries*, 19(4): 169-174.
- Machovec, G.S. 2001. Library Portals : "Customizing and Focusing the User's Experience." *Online Libraries and Microcomputers*, 19(1): 1-3.
- Peppers, D. and Rogers, M. 1997. *The One-to-One Future: Building Relationships One Customer at a Time*. Double Day Publications.
- Pointcast Network.
< <http://www.pointcast.com> >
- Pretschner, Alexander and Gauch, Susan. 1999. *Personalization on the Web*, ITTC-FY2000-TR-13591-01.
- Tennant, Roy. 1999. "Personalizing the Digital Library." *Library Journal*, 124(12): 36-38.
- Trendsetters.
< <http://www.Trendsetters.com> >
- Winter, Ken. 1999. "MyLibrary Can Help Your Library." *American Libraries*, 30(7): 65-68.