

# 간접적으로 추출된 데이터를 활용한 사례기반 접속지역 추론기법

정용진

*Institute for Advanced Engineering*  
633-2 Goan-ri, Backam-myun, Yongin-si, Gyeonggi-do, Korea  
Tel: +82-31-330-7407, Fax: +82-31-330-7120, E-mail: [realizer@iae.re.kr](mailto:realizer@iae.re.kr)

Abstract

*The present much information of internet has to exist for innumerable user so that couldn't satisfy there's a variety of demand. so they have a demerit that search unnecessary information. However Web service is different with other mass media because It is possible that enable Mass Customization for Personalization strategy. In The paper suggest reasoning system that detect user connection location by using indirect abstraction techniques a kind of Case-based reasoning techniques.*

## Connection location Case-based reasoning technique Using indirect data

고등기술연구원 엔지니어링정보기술센터  
경기도 용인시 백암면 고안리 633-2 449-863  
Tel: +82-31-330-7407, Fax: +82-31-330-7120, E-mail: [realizer@iae.re.kr](mailto:realizer@iae.re.kr)

요약

현재 인터넷의 수많은 정보들은 다수의 임의 사용자들을 대상으로 하고 있으므로 각 개인의 다양한 요구 사항을 고려할 수 없고, 이로 인하여 수많은 불필요한 정보를 검색해야 하는 단점을 안고 있다. 하지만 웹 서비스는 기존의 미디어와는 달리 사용자 별로 차별화된 맞춤형 서비스를 제공하는 대량 개별화가 가능하다. 본 논문에서는 이러한 개인화 전략을 지원하는 서비스를 구현하기 위하여 사례기반 추론 방식의 한가지 분류인 간접적인 추출기법을 사용하여 사용자의 접속 지역을 추론해내는 시스템을 제안한다.

### Keywords:

*personalization strategy, case-based reasoning, Mass Customization, Context-Awareness Computing.*

## 1. 서론

인터넷이 보편화 되면서 수많은 정보들이 등장하고 빠른 시간 내에 갱신되거나 사라짐에 따라 사용자 들은 자신이 찾고자 하는 정확한 정보를 검색 및 수집하기 위해 많은 비용과 시간을 들이고 있다. 이러한 수많은 정보들은 다수의 임의

사용자들을 대상으로 하고 있으므로 각 개인의 다양한 요구 사항을 고려할 수 없고, 이로 인하여 수많은 불필요한 정보를 검색해야 하는 단점을 안고 있다. 따라서 이러한 사용자의 개별적인 요구를 만족시키는 개인화된 서비스(personalized service)의 필요성이 대두되었다. 이러한 개인화 전략을 위한 핵심 기술 중의 하나는 고객들의 특성이나 취향 등을 분석하여 개인별로 차별화된

정보를 자동적으로 필터링하는 패턴 분석 시스템이 요구된다. 본 논문에서는 패턴 분석 방법인 사례기반 추론(case-based reasoning) 기법을 이용하여 사용자의 지역정보를 추출해내는 방법을 제안하였다.

## 2. 관련 연구

### 2.1 개인화 전략 (personalization strategy)

인터넷의 대중화되고 사용자들은 언제 어디서나 필요한 정보를 쉽게 얻을 수 있게 되었다. 하지만 인터넷에서 제공하는 정보의 폭발적인 증가로 인하여 정보의 과부하(information overload) 현상이 발생하게 되었다. 이에 따라 사용자들은 필요한 정보를 찾거나 원하는 상품을 구매하기 위해서 오히려 더 많은 시간을 소비해야 하는 결과를 초래하였다. 이러한 문제점을 개선하기 위하여 인터넷 비즈니스 분야에서 등장하고 있는 기법이 개인화 전략 (personalization strategy)이다.

이러한 개인화 전략은 고객에 대한 행동 패턴을 분석하여 개인별로 차별화된 정보나 서비스를 제공하기 위한 전략이다. 예를 들면 콘텐츠를 제공하는 사이트에서 고객이 선호하는 콘텐츠를 제공함으로써 고객의 만족도를 높이고 고객과의 친밀한 관계를 지속시킬 수 있다. 이러한 개인화 전략을 이용하여 인터넷 사이트에서 고객과의 상호작용을 개인적으로 차별화 할 수 있는 대량 개별화(Mass Customization)가 가능하다.

IT분야 자문 및 정보서비스 그룹인 GarnerGroup에서는 개인화를 관련성 분석, 맥락형 추론, 클릭 스트림, 프로파일링, 선호도 매칭으로 구분하였다. 이러한 개인화는 크게 온라인과 오프라인으로 구분할 수 있다. 온라인은 명시적인 정보와 명시적인 정보로 구분된다. 여기서 명시적인 정보는 사용자의 인구통계학적 정보와 사용자들의 직접적인 평가정보를 의미하며, 묵시적인 정보는 시스템에 의해 자동적으로 수집되는 평가정보이다.

관련성 분석은 고객의 과거 행동에 근거 하여 과거 상호 작용을 분석하는 방식이며, 맥락형 추론은 사용자에게 보여진 콘텐츠의 내용을 분석하여 관련된 콘텐츠를 제공하는 방식이다. 클릭 스트림 분석은 방문자가 열람한 데이터를 분석하고 수집하여 관련된 콘텐츠를 제공해 주는 방식이다. 프로파일링은 사용자가 명시적으로 제공한 인구통계학적인 요인과 선호도에 대한 데이터를 미리 정의된 콘텐츠로부터 생성된 프로파일과 결합하여

정보를 제공하는 방식이다. 선호도 매칭은 선호에 대한 명시적인 자료를 수집하여 다른 사람들의 선호와 비교하여 제공해주는 방식이다.

개인화 전략은 결과적으로 사용자가 자신의 선호, 관심분야와 같은 정보를 웹 사이트에 제공하면 웹 사이트는 사용자가 제공한 자료를 기초로 사용자에게 가장 알맞은 정보를 제공한다. 개인화를 통해서 웹 사이트 운영자는 사용자에게 관련 자료를 얻고 사용자의 지속적인 이용이나 접속을 얻어낼 수 있게 되며 사용자는 자신에게 가장 알맞은 정보를 편리한 방법으로 얻을 수 있게 된다.

개인화 전략의 개념은 아직 초기단계로 다음과 같은 기술이 개발되어 있다.

1. 복합적인 프로파일 학습을 이용한 개인용 웹 정보 수집 시스템
  - 가. 사용자 프로파일 학습 알고리즘 개발
  - 나. 웹 페이지 필터링 시스템 개발
  - 다. 개인용 정보수집 에이전트 프로토타입 개발
2. 지식기반 자연언어 처리를 이용한 문서의 자동분류와 지능형 색인에 관련 연구
  - 가. 문서 자동 분류를 위한 색인기 개발
  - 나. 통계적 분류 알고리즘 개발
  - 다. 패턴에 의한 분류 알고리즘 개발

현재 개인화 전략은 초기 개발 단계이지만 앞으로는 동적으로 개인의 취향에 따라 정보를 찾고 클리핑 해 주며, 사용자가 원하는 정보를 추론해주는 등 개인화된 서비스를 제공하는 시스템에 대한 수요가 점차 늘어날 것으로 전망된다.

### 2.2 사례기반추론(case-based reasoning) 기법

인공지능 분야에서 주로 사용되는 사례기반 추론 기법은 사람이 새로운 문제를 해결하기 위해 과거에 경험한 사례 중에서 가장 유사한 사례를 검색하여 새로운 문제에 대한 해결과 교정을 통한 학습을 반복하는 기법이다. 이 기법은 새로운 문제가 발생하면 이미 경험한 사례 데이터베이스에서 가장 유사한 사례를 검색하여 이전 사례를 재사용하는 적응 과정을 통해 새로운 문제를 해결하는 방식이다 또한 교정과정에서 새로운 문제의 해가 문제 해결에 적합한지를 검정하여 새로운 사례로 학습한다.

사례기반 추론 기법을 위한 관심 분야 추출은 크게 직접적인 방법과 간접적인 방법이 있다. 직접적인 방법의 경우, 사용자에게 관심분야에 대해 직접적으로 질문을 하여 입력을 받는 것으로, 응답이 이루어진 시기에 정확한 정보를 확보할 수 있으나, 사용자가 작성한 시간의 데이터에 따라 추론을 하게 되므로, 사용자가 자신의 정보를 갱신

하지 않을 경우 처음에 응답이 이루어진 시기의 정보에 따라 항상 같은 내용을 보여주게 된다. 간접적인 관심분야 추출에 관한 연구는 사용자의 과거 행동 통계로부터 학습하는 방법으로 크게 사용자의 습관 정보로부터 사용자 정보를 학습하는 방법은 특성의 유사성 비교에 의한 것과 과거의 상황과 유사성을 비교하여 해법을 적용하는 것이 있다.

본 논문에서는 사례기반 추론기법의 한가지 분류인 간접적인 추출 기법을 사용하여 사용자의 과거 접속 위치 통계로부터 사용자의 인터넷 접속 지역에 대한 데이터 추출하여 추론하는 사례 기반 추론(case-based reasoning)기법을 제안한다.

제안된 방법은 인터넷 회원제 사이트에서 접속한 사용자의 접속 IP를 기초로 대략적인 위치를 수집하고 접속빈도에 의하여 패턴을 추론하여 사용자의 지역 정보를 동적으로 관리하는 기능을 구현한다.

### 3. 구현 및 결과

본 장에서는 개인화 된 웹 서비스 지원체계를 위한 간접적인 추출 기법의 구현을 설명한다.

Figure 1과 같이 본 논문은 특정 사용자가 접속하였을 경우 그 사용자의 접속 IP를 기초로 하여 whois서비스를 통해 대략적인 위치 정보를 검색하고 그 응답을 메타 데이터로 넘겨 받아 주소 정보를 나타내는 데이터를 추출하여, 그 데이터를 바탕으로 추론하는 사례기반 추론을 실행한다.

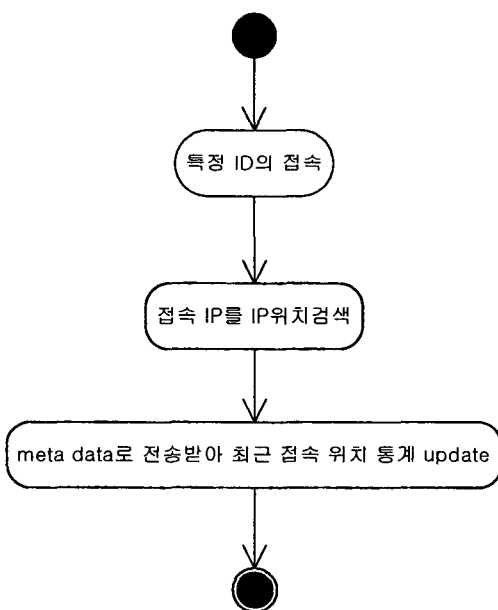


figure 1. 사용자의 인터넷 사용주소를 얻어내는 프로세스

### 3.1 정보 추출 과정

인터넷 통신에서 기본이 되는 프로토콜은 TCP/IP이다. TCP 프로토콜은 3way-handshake 형태의 접속과정을 통하여 서비스를 요청하게 된다. 즉, 인터넷에서 특정 호스트에 접속하여 서비스를 제공받기 위해서는 접속시에 서비스 제공자에게 자신의 IP를 전달 하고, 확인 받는 과정이 필수적으로 요구되므로 인터넷 서비스 제공자는 사용자의 IP 정보를 가지게 된다.

이렇게 얻어낸 사용자의 접속 IP는 Whois라는 서비스를 통하여 개략적인 등록자, 주소, 연락처 등의 정보를 얻어낼 수 있다. 그 데이터를 바탕으로 뽑아낸 Xml형태의 메타 데이터를 아래의 table 1과 같이 추출 할 수 있다.

<location>한국과학기술정보연구원</location> <address>유성구 유성우체국 사서함 122호</address>
<location>(주)케이.알.라인</location> <address>강남구 도곡동 902-55 신한빌딩 202호</address>
<location>한국통신</location> <address>성남시 분당구 정자동 206 한국통신 e-Biz본부 기획팀</address>
<location>호서대학교</location> <address>아산시 배방면 세출리 산29-1</address>

표 1. 추출된 주소 데이터

위와 같이 얻어진 주소 데이터를 패턴 저장 DB에 업데이트하여, 일정 횟수의 접속 데이터의 내용을 유지하고, 그 데이터를 바탕으로 사용자가 주로 인터넷을 사용하는 장소를 추론해 내는 프로세스를 거쳐 사용자의 지역정보를 대략적으로 얻어낼 수 있다.

위의 프로세스를 그림으로 정리하면 아래의 figure 2와 같다.

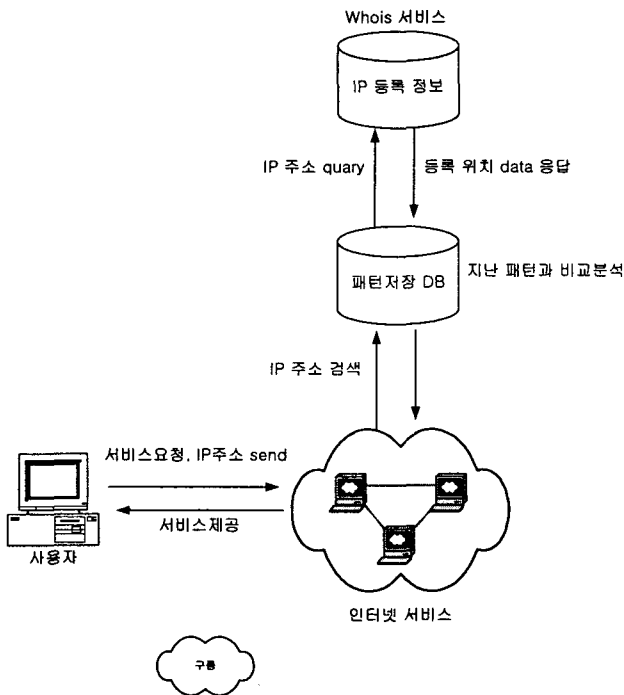


Figure 2. 사용자의 주소를 얻어내는 프로세스

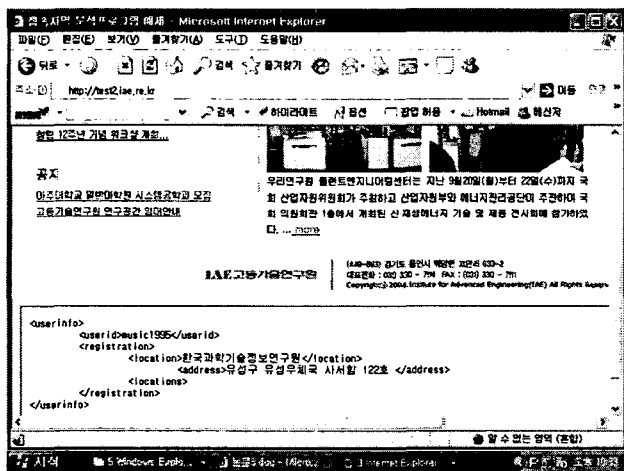


figure 3. xml로 얻어낸 사용자의 현재 접속위치

위에 figure 3과 같이 본 논문에서는 실제 사용자가 주로 인터넷을 사용하는 지역을 추출하는 부분까지 구현 하였다. 추출된 정보를 이용하여 사용자의 지역의 날씨 생활 정보를 지원하는 등 여러가지의 관련된 컨텐츠 서비스가 가능할것으로 기대된다.

#### 4. 결론 및 향후 연구과제

오늘날 인터넷의 정보 과다로 인하여 사용자는 자신에게 적절한 정보의 검색 시간이 점점 길어지고 있으며, 검색 과정에서 필요한 정보들을 놓쳐 버릴 가능성이 증대되어 사용자 개인의 필요에 맞는 자료를 제공하는 서비스가 갈수록 요구되는 실정이다.

본 논문에서 제시한 간접적으로 추출된 데이터를 활용한 사례기반 접속지역 추론기법에 대하여 실제 개인화된 서비스에 적용한 실험 결과는 제시하지 못하고 있다. 이러한 개인화 서비스에 대한 평가는 다양한 계층의 사용자들을 대상으로 상당한 기간에 걸쳐 이루어져야 하므로 연구기간 중에 수행할 수 없었음을 밝히며 향후 연구 개발의 과제로 실제적인 효과의 검증 과정을 거칠 것이다.

#### [참고 문헌]

- [1] 한선미, 유진운. (2001). 지능형 에이전트를 이용한 유무선 뉴스 검색 시스템. 정보처리학회논문지B 제8-B권 제6호.
- [2] 김영지, 문현정, 옥수호, 우용태. (2002). 사례기반추론 기법을 이용한 개인화된 추천시스템 설계 및 구현. 정보처리학회논문지D 제9-D권 제6호.
- [3] 김철수. (2002). 웹 고객의 개인화를 지원하는 지식기반 통합시스템. 정보처리학회논문지 B 제9-B권 제1호.
- [4] 이경호, 윤창현, 박두순. (2003). 웹마이닝을 이용한 M-Commerce 추천시스템 설계 및 구현. 한국컴퓨터교육학회 논문지 제6권 제3호.
- [5] 김형도. (1999). 개인화된 웹 광고를 지원하기 위한 요구 통합조정 체계의 설계. 한국정보처리학회 논문지 제6권 제6호.
- [6] 김영민, 이상준. (2002). 제목 내의 키워드 역할정보를 이용한 연구 논문 검색 시스템. 제주대학교 첨단기술연구소 논문집 제13권 2호.
- [7] O. Etzioni,. (1996). The world wide web : quagmire or gold mine?. ACM comm.. Vol.39 No.11.
- [8] Jason Compton. (2004). Web Extra: Gartner's Five categories of Personalization. <http://www.destinationcrm.com/articles/default.asp?ArticleID=603>.
- [9] XML : eXtensible Markup Language, <http://www.w3.org/XML>.
- [10] XML Query. <http://www.w3.org/XML/query>.