사용자 질의패턴 기반 에이전트에 의한 맞춤형 질의추천

임 요 한* 박 건 우 이 상 훈
국방대학교 전산정보학과
yohan0377@naver.com pjw4050@hanmail.com hoony@kndu.ac.kr

Customized Query Recommendation by Agent Based on User's Query Pattern

Yohan Lim* Gunwoo Park Sanghoon Lee
Dept. Computer Science and Information, Korea National Defense University

요 약

검색엔진을 사용해 질의를 입력 후 사용자가 원하는 정보를 얻을 때까지의 검색 결과정보의 탐색 범위에 대해 설문한 연구 보고서에 검색 결과정보의 첫 페이지만 보는 사용자가 설문인원의41%를 차지했고, 상위 3페이지만 사용하는 사용자는88%에 달한다고 하였다. 따라서 검색결과의 상위순위는 사용자의 정보 존재여부를 판단하는 중요한 척도가 된다. 또한 인터넷의 방대한 정보로 인해 정보 홍수에 빠진 사람들은 정보에 대한 까다로운 요구를 하고 있다. 이를 테면 개인화 또는 맞춤화된 정보를 제공 받기를 원하고 있다.정보검색 시 대다수의 사용자들은 질의의 길이를2단어 이하의 키워드를 사용하여 질의가 특정한 토픽을 지향하도록 하고 있다. 본 논문에서는 데이터 마이닝의 연관규칙을 적용 사용자 프로파일DB내 질의에 대한 사용자 질의패턴을 분석하여 '분석 Agent' 통한 연관 질의 리스트를 생성하고 '추천 Agent'는 사용자들의 취향변화 즉 시간에 따라 변하는 관심영역 또는 사용자 질의 변화에 대해서 날짜별 가중치를 부여하여 사용자와 상호교류를 통해 사용자에게 맞춤형 질의를 추천하는 방안을 제시하고자 한다.

1. 서 론

현재 웹 상에는 수많은 정보가 존재하고 있으며 정보의 종류 및 문서의 수도 방대하다. 오늘과 같은 정보화시대에서 사용자가 원하는 웹 문서의 검색은 매우 중요한 기술이다. 초기의 웹 검색 시스템은 정보 공유를 중요시 하여 사용자가 원하는 정보를 웹에서 대량으로 추출하여 주는 것만을 고려하였으나, 현재와 같이 방대한 정보가 내재되어 있는 웹에서는 사용자가 원하는 정보를 얼마나 정확히 추출할 수 있는가가 보다 중요하다고 할 수 있다.

2006년 검색엔진을 사용해 질의를 입력 후 사용자가 원하는 정보를 얻을 때까지의 검색 결과정보의 탐색 범위에 대해 설문한 연구 보고서에 검색 결과정보의 첫 페이지만 보는 사용자가 설문인원의41%를 차지했고, 상위 3페이지만 사용하는 사용자는88%에 달한다고 하였다[4]. 따라서 검색 결과의 상위순위는 사용자의 정보 존재여부를 판단하는 중요한 척도가 된다.

또한 기존의 단순 문서 읽기에서 사이트를 방문하여 음악을 듣고, 물건을 구입하는 등 다양한 활용법이 나타나고 있다. 그러나 인터넷의 방대한 정보로 인해 정보 홍수에 빠진 사람들은 정보에 대한 까다로운 요구를 하고 있다. 이를 테면 개인화를 원하고 있다.

정보검색 과정에서 사용자가 질의를 입력하면 검색엔진자료에서 질의와 일치하는 자료를 검색하고 검색알고리즘에 의거하여 검색결과를 출력하는 형태로 이루어진다. 이 때 대다수의 사용자들은 질의의 길이를2단어 이하의 키워드를 사용하여 질의가 특정한 토픽을 지향하도록 하고 있다[5].

예를 들어 검색엔진에 "자동차"라는 질의를 입력하게 되면 검색엔진에서는 사용자의 질의의도가 자동차의 무엇을 의미하는지 명확히 알 수가 없으며 검색결과도 자동차 관련 다양한 정보를 검색해 줄 것이다. 하지만 자동차 정비소, 자동차 극장, 자동차 인테리어 등 질의를 입력하게 되면 사용자의 질의 의도에 맞는 검색 결과가 상위에 보여 질 것이다. 또한 여러 검색엔진에서 연관검색어 서비스를 제공하고 있지만 사용자 관심(개인화)에게 꼭 맞는 연관검색어 서비스는 한계가 있다.

본 논문에서는 개인화 검색서비스 구현의 문제점[3]을 다소나마 해결하고자 사용자 질의패턴을 분석하여 사용

자들의 취향변화 즉 시간에 따라 변하는 관심영역 또는 사용자 질의 변화에 대해서 질의 패턴을 분석하고 추천 에이전트는 생성된 연관질의 리스또Rule)을 이용하여 사용자에게 적당한 질의를 추천하는 사용자 맞춤형 질의 추천 방안을 제시하고자 한다.

본 논문의 구성은 2장에서 데이터 마이닝 기법 중 연관규칙과 순차패턴과 개인화 검색 서비스의 연구 동향 및 문제점을 살펴보고 3장에서는 사용자 질의패턴의 구조화하여 추천하는 방법을 소개한다. 4장에서는 결론 및 향후 연구과제에 대해 기술한다.

2. 관련연구

이 장에서는 사용자 질의 패턴을 추출하기 위하여 필요한 데이터 마이닝의 기법 중에 연관규칙(Association Rule)과 순차패턴((Sequence Pattern) 연구하고 개인화 검색 서비스의 연구동향과 문제점을 기술한다.

2.1 데이터 마이닝 기법

2.1.1 연관규칙(Association Rule)

데이터베이스에서 알려져 있지 않은 숨겨진 패턴을 탐사하는 연구 중에 연관규칙에 대한 많은 연구가 이루어졌다. 연관규칙은 말 그대로 한 항목 그룹과 다른 항목 그룹 사이에 존재하는 강한 연관성을 찾아내어 그룹화하는 클러스터링의 일종이다. 또한 동시에 구매될 가능성이 큰 상품들을 찾아냄으로써 시장바구니 분석(Market Basket Analysis)에서 다루는 문제들에 적용할 수 있다. 연관규칙발견 알고리즘으로는Apriori, OCD, SETM, DHP알고리즘등이 연구 되었다. 연관규칙기법에 적용되는 데이터는 판매시점에서 기록된 거래와 품목에 관한 정보를 담고 있다. 연관규칙탐사 과정은 크게 두 단계로 진행이 된다. 첫 번째는 높은 지지도(Support)를 갖는 연관규칙을 도출하는 작업이다. 여기서 지지도와 신뢰도의 개념은 아주 중요한 개념으로 빈발항목 집합을 찾아내는데 있어 큰 역할을 한다.

지지도란 전체 트랜잭션에서 특정 패턴(A->B)이 차지하는 비율이고 신뢰도란 A를 구매하는 고객 중에 B를 구매하는 고객이 차지하는 비율을 말한다. 예를 들어 전체 거래건수 1000 건, A는 500건, B는 300건, A와 B는 250건에 대해서 특정한 패턴(A->B)에 관해 신뢰도와 지지도를 구해보면

신뢰도(Support) : 250/1000=25%

지지도(Confidence) : 250/500=50%

이를 해석하면 1000건 중에 A와 B를 동시에 구매한 고객이 전체의 25%이고, A를 구매 했을 때 B를 구매하는 고객은 25%중에 50%, 즉 전체의 12.5%를 차지한다는 것이다.[2]

2.1.2 순차패턴(Sequence Pattern)

순차패턴은 동시에 구매될 가능성이 큰 상품 군을 찾아내는 연관규칙에 시간의 개념이 포함되어 순차적인 구매 가능성이 큰 상품 군을 찾아내는 방법이다. 순차패턴에서는 연관규칙 "A->B"는 상품 A 가 구매되면 일정 시간이 경과한 다음 상품 A가 구매되면 일정 시간이 경과한 다음 상품 B가 구매된다 라고 해석된다. 즉 순차패턴은 구매 순서가 고려되어 상품간의 연간성이 측정되고, 이에 따라 유용한 연관규칙을 찾는 기법이다[2].

2.2 개인화 검색 서비스의 유형

개인화 검색의 유형은 개인을 정의하는 방법에 따라서 구분하는 방식과, 어떤 내용을 개인화하는가에 따라서 구분하는 방식으로 나누어 살펴볼 수 있다.

첫 번째 유형 구분 방식은 '개인'을 정의하는 방법에 따라서 서비스 유형을 구분한다. 첫째, 개인의 프로파일을 생성하는 방법으로서 사용자가 적극적으로 프로파일의 사항을 입력하면 이를 이용하여 기본 프로파일을 작성하는 방식이다. 둘째, 서비스 이용형태를 기반으로 하는 방식으로서 클릭의 흐름을 분석하여 사용자의 프로파일을 정의하는데 보통 웹 이용 마이닝 시스템을 통하여 작성한다. 셋째, 사용자의 사회화 프로파일 기반방식으로서 이용자가 누구와 관계를 맺고 있는가를 추적하여 협업 필터링을 통하여 개인을 정의한다.

두 번째는 어떤 내용을 개인화하여 검색 결과를 제공하는지에 따른 구분 방식이다. 링크 정보를 이용하거나, 질의를 확장하거나, 결과를 재순위화 하거나, 메타검색, 혹은 도메인별 검색 등이 있다. jeh and Widom은 사용자가 즐겨 찾는 페이지에서 링크되거나 해당 페이지가 링크한 페이지에 더 많은 가중치를 두어 검색랭킹에 반영하는 형태를 연구하였다. Liu, Yu, and Meng는 사용자의 프로파일에 기반해 입력된 질의를 사용자 관심 주제로 매핑시켜 검색 대상을 한정 시켰다. Websfter 프로젝트에서는 사용자의 검색 의도를 입력 받은 후 이를 해당 이용자의 검색 분류체계 안에서 가장 적절한 질의어로 변환시키는 방법을 연구하였다[3].

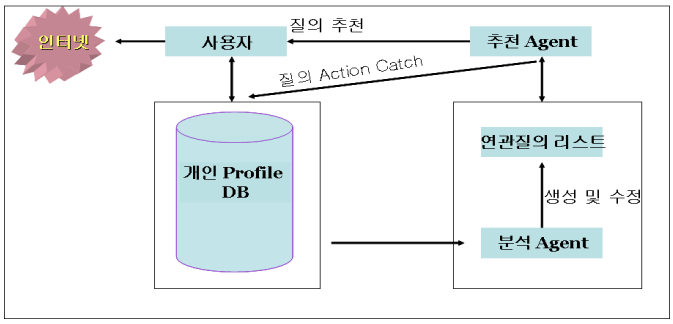
2.3 개인화 검색 서비스 구현의 문제점

개인화 검색 서비스를 제공하기에는 많은 현실적인 어려움이 따른다. 첫째, 사용자의 관심 주제 추출과 관련된 문제는 사용자가 직접 관심 주제를 입력하는 것이 개인화 서비스의 안정적인 출발선 일 수 있는데 사용자가 직접 관심 주제를 입력하는 작업이 매우 번거로워 참여율이 저조하며 또한 사용자의 관심주제가 시간에 따라 변하는데, 그때마다 사용자가 관심 주제를 수정할 것으로 기대하기는 어렵다. 둘째, 로그 데이터는 사용자의 서비스 사용에 대한 완전한 맥락 정보를 갖지 않은 단편적인 정보를 저장한다. 로그 데이터를 이용해 사용자의 관심

주제를 추출할 경우, 광범위한 사용자의 관심 주제를 추출할 경우, 광범위한 사용자의 관심 주제를 포괄적이고 입체적으로 반영하기 힘들다. 다시 말해 사용자의 의도나 해석과 같은 정량적으로 측정하기 힘든 부분에 대한 파악이 실질적으로 불가능하다는 것이다. 셋째는 처리해야 하는 방대한 데이터 양이다. 분석 기간이나 범위등에서 지나치게 포괄적인 분석을 시도하거나 분석 모델링이 복잡할 경우에는 처리속도 등의 문제로 적기에 사용자의 관심을 추출할 수 없다. 따라서 적절한 수준에서 분석 범위와 임계치를 설정하는 것이 필수적이다[3].

본 연구에서는 위와 같은 제한점을 다소나마 극복하기 위해서는 사용자가 스스로 관심사를 등록하는 것과 같은 적극적인 참여를 최소화 시키면서 시간에 따라 변화하는 사용자의 관심 주제를 효율적으로 반영하고 할 수 있도록 개인의 관심 주제를 사용자 질의패턴에서 추출하여 추천 Agent를 통한 맞춤형 질의를 추천 할 수 있도록 [2]에서 제안한'상품 추천서비스'와 추천 Agent 의 연관질의 리스트 추천하기 위해 [1]의'다중 프로파일을 이용한 문서검색비율관리를 참고하였다.

3. 제안하는 시스템 구조



<그림 1> 시스템 구조

사용자가 질의하게 되면 질의 내역이 사용자 프로파일 DB에 남게 된다. 분석 Agent가 이 정보를 이용하여 연관질의 리스트(Rule)을 추출하고, 구조화된 트리를 생성한다. 추천 Agent는 생성된 연관질의 리스트를 이용하여 추천할 질의를 결정해서, 이를 사용자에게 추천하게 된다

3. 1 사용자 질의패턴의 추출 및 구조화

추출된 패턴의 구조화는 5단계를 통해 이루어 진다.

1) Profile DB에서 모든 사용자의 질의내역을 추출
2) 사용자의 질의내역에 대한 형태소 분석
3) 원시패턴을 연관규칙 이용 단일 패턴으로 나눔
4) 생성된 단일패턴의 빈도수 테이블
5) 분포도에 의거하여 구조화된 트리 생성

Rule 1 : 자동차-> 정비소-> 극장-> 인테리어
Rule 2 : 극장-> 인테리어 -> 시간
Rule 3 : 자동차 -> 정비소 -> 시간 -> 인테리어
Rule 4 : 자동차 -> 극장 -> 인테리어
Rule 5 : 극장 -> 시간

<표 1> 사용자 원시 질의 패턴

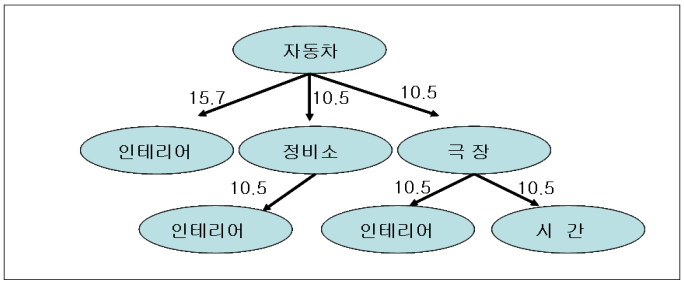
사용자의 원시 질의패턴을 연관규칙을 이용하여 단일 패턴으로 나눈다. 단일 패턴으로 나누기 위하여 고려해야 할 조합은 예로 든 <표 1>의 Rule 1 같은 경우는 {자동차,정비소}, {자동차,극장}, {자동차,인테리어],{정비소, 극장}, {정비소,인테리어},{극장,인테리어} 이렇게 6개($_4C_2$)이다. 이와 같은 방법으로 사용자의 원시 질의패턴의 모든 조합을 구한다. 이때 {자동차, 정비소}와 {정비소, 자동차}는 다르게 취급한다. 장바구니 분석에는 위와 같은 패턴이 같은 항목으로 취급되지만, 제안된 패턴의 구조화에서는 패턴에 방향성이 고려되기 때문이다. 모든 패턴이 조합되었다면 다음과 같은 <표 2>를 생성할 수 있다.

[전체 사용자 질의 19건]

자동차 -> 정비소	자동차 -> 극 장	자동차 -> 인테리어	자동차-> 시간
2 (10.5%)	2 (10.5%)	3 (15.7%)	1 (5.2%)
정비소 -> 자동차	정비소 -> 극 장	정비소 -> 인테리어	정비소-> 시간
0 (0%)	1 (5.2%)	2 (10.5%)	1 (5.2%)
극장 -> 자동차	극장 -> 정비소	극장 -> 인테리어	극장-> 시간
0 (0%)	0 (0%)	2 (10.5%)	2 (10.5%)
시간 -> 자동차	시간 -> 정비소	시간 -> 극 장	시간-> 인테리어
0 (0%)	0 (0%)	0 (0%)	1 (5.2%)
인테리어 -> 자동차	인테리어-> 정비소	인테리어-> 극 장	인테리어-> 시간
0 (0%)	0 (0%)	0 (0%)	1 (5.2%)

<표 2> 빈도수 테이블

이렇게 단일 질의로 추출된 패턴들은 각 질의와 얼마만큼의 연관을 가지고 있는지 나타내게 된다. 이 빈도수 테이블(support 5%이상)에 의거하여 트리를 구성한다. 트리를 구성함에 있어서 가장 빈도수가 높은 질의는 다른 질의와의 연관도가 가장 높다고 할 수 있다. 이 질의들을 중심으로 트리를 구성해 나간다. 하위 노드의 확장은 빈도수가 가장 높은 질의를 첫 번째 하위 노드 두 번째로 높은 질의는 그 다음 노드로 이러한 방식으로 확장 노드를 생성해 간다.



<그림2> 구조화된 트리 (Support 5% 이상인 항목)

<그림 2>에 따라 빈도수가 높은 질의를 우선으로 하여 하위 노드의 왼쪽에 두면서 점차 트리를 구성해간다. 이 때, support 5%미만인 연관 질의들은 빈도수의 저조에 의해 자동으로 트리이 구조화에 반영되지 않는다. 이와 같이 일정한 threshold를 두어서 많이 선택되지 않는 질의들을 걸러내는 작업이 필요하다. 이 값은 일정한 수치가 정해져 있는 것이 아니라, 통계치에 의해서 사용자가 선택 할 수도 있다.

3.2 추천 질의어 선택

개인 프로파일에서 질의 패턴 분석을 통해서 얻어진 연관 질의 리스트에 대하여 빈도수 측정을 통한 특정 질의와 관련이 깊은 질의를 찾아내었다. 이것을 빈도수 트리로 구조화 하여 연관 질의에 대한 구조화를 수행 하였다.

추천 Agent는 사용자의 질의 정보를 감시하다가 특정 질의 내용이 감지하면 그 즉시 구조화된 트리의 탐색에 의해 연관 질의 리스트를 추천하게 된다. 기본적으로 탐색은 pre-order방식으로 Breath-first Search를 수행하게 되며, 탐색을 시작하게 될 Root Node는 사용자 처음 질의한 키워드가 된다. 탐색 된 하위 노드들 중에 Support 값이 가장 큰 노드를 추천하게 된다. 추천한 질의를 사용자가 이용하게 되면 이는 다시 사용자 프로사일DB에 저장되며, 다시 트리의 탐색을 통해 다음 질의를 추천하게 된다. 만약 Root Node를 정할 때 트리에 여러 개의 동일 노드가 있다면 하위 노드의 Support값이 가장 큰 노드를 Root노드로 결정한다.

3.3 구조화된 트리의 수정

사용자 프로파일 DB에 질의한 데이터가 점점 쌓여 갈수록 사용자의 질의 패턴을 트리에 반영해 주어야 한다. 항상 최신 패턴을 반영해주기 위하여 사용자의 질의 데이터가 특정한 threshold를 넘을 때마다 트리를 수정한다. 예를들어 새로운 패턴에서 '자동차->정비소'의 support가 16.4%로 변경되었다면 '자동차->정비소'에는 $|S - S_{new}|$ 를 추가하고 같은 레벨의 나머지 노드에는 $|S_{dd} - S_{new}|$ / 동일 레벨 노드의 개수 -1 만큼 각각 감소시켜 준다. 위와같은 방식으로 트리를 점차 사용자의 질의 패턴에 맞추어 수정해 나간다면 사용자의 질의패턴을 효과적이고 빠르게 반영할 수 있을 것이다.

3.4 도태된 항목의 삭제

시간이 지나면 지날수록 점점 많이 질의 된 연관질의가 있는가 하면 점점 쓰이지 않아서 점차 사용자 프로파일 DB에서 찾아볼 수 없는 질의들이 생기게 된다. 이러한 질의들은 트리를 일정한 간격마다 한번씩 지지도가 특정한 Threshold를 넘지 않는 질의들을 삭제하면 점차 사라져 가는 질의를 트리의 구조화에 반영 할 수 있다.

3.5 추천 Agent 의 연관 질의 리스트 추천

추천할 수 있는 많은 연관 질의 리스트가 있겠지만 그 중 사용자의 관심이 높고 성격이 다른3개의 연관리스트의 대표적 연관 질의(자동차-인테리어, 자동차-정비소, 자동차-극장)를 선정하고 각각의 연관 질의에 대한 관심도는 아래<표 3>와 같다고 가정한다.

구 분	1일차	2일차	3일차	4일차	5일차	6일차	7일차
자동차-인테리어	14	24	33	34	38	43	54
자동차-정비소	33	33	33	33	33	33	33
자동차-극장	53	43	34	33	29	24	13
계	100	100	100	100	100	100	100

<표 3> user 연관 질의 관심도

사용자의 관심 영역과 질의를 위한 주요 연관질의는 어느 시간 정도는 일정할 수도 있고 주어진 상황속에서 많은 변화를 가져 올 수 있다. 업무를 하다가 필요한 물품이 생길 수도 있으며 가족의 생일로 인해 선물을 구입할 수 있다. 따라서 사용자의 연관 질의의 중요도는 달라질 수 있을 것이다.

<표 3>를 분석해 보면 '자동차-인테리어' 연관질의는 날짜가 지날수록 중요도가 증가하고 있고, '자동차-정비소'는 날짜의 지남과 상관없이 중요도가 유지되고 있으며, '자동차-극장' 연관 질의는 시간이 지남에 따라 관심도가 떨어지고 있음을 알 수 있다.

그렇다면 '8일차'때 추천 에이전트가 사용자의 관심도에 맞는 비율로 필요한 질의를 추천하여 자료를 수집할 수 있을 것이다.

본 논문에서는 각 날짜의 결과에 따라서 가중치에 대한 변화를 부여하여 최근 연관 질의의 선호도가 큰 질의어일수록 연관질의의 중요도를 높게 부여하는 방법을 적용해 보겠다.

$$[\ \sum_{i=0}^{n}(\frac{k}{100})w\ ] \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (1)$$

(1) 식은 날짜의 변화에 따른 연관질의에 대한 가중치를 부여한 식이다.

I : 사용자가 에이전트를 사용한 날짜

k : 각 연관 질의의 중요도

w : 가중치

날짜별 가중치 부여를 설명하기 전 각 날짜의 사용자 선호 연관질의에 대한 가중치가 없는 상태를 보겠다.

<표 3>에서 사용자가 첫날부터 7일차까지 결과를 분석하여 보면 전체 연관질의 중 '자동차-인테리어'의 사용자 관심율은 14%,24%,...,54% 임을 볼 수 있으며 가중치 없이 (1)식을 통해 '자동차-인테리어' $\frac{240}{100}$, '자동차-정비소' $\frac{231}{100}$,'자동차-극장' $\frac{229}{100}$ 이며 전체 대비 각각의 사용자의 선호도는 약 34%, 33%, 33%로서 가중치를 부여하지 않은 경우 시간의 지남에 관계없이 사용자 관심 연관질의는 모두 관심도가 비슷함을 볼 수 있다.

이러한 문제점을 해결하기 위해 일정 가중치를 두어서 연관 질의에 대한 예전의 중요도와 어제의 중요도를 적절하게 조정해 줄 필요가 있다. 그리고 가중치는 최근의 것 일수록 그 가중치의 비중을 높여 줄 필요가 있다.

$$w = \sum_{i=0}^{n}\frac{1}{n} \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (2)$$

w = 날짜별 가중치

이러한 가중치 적용하여 사용자이 선호도 연관 질의를 적용하려 백분율로 표현하면 '자동차-인테리어'는 94%, 자동차-정비소 85%, 자동차-극장 73% 이다.

이를 상대적으로 비교해 보기 위해 전체 합에 대한 연관질의의 각각의 중요도를 나타내면 자동차-인테리어 약 37%, 자동차-정비소는 33%, 자동차-극장 28%로 사용자 관심 반영 및 날짜의 변화에 따라 추천 에이전트는 적합한 비율로 연관 질의를 추천하여 자료를 수집할 수 있다.

이러한 사용자 선호 연관 질의에 날짜별 가중치를 적용함으로써 어제 하루 관심이 높았다는 이유로 사용자 연관 질의를 추천 할 우려를 예방할 수 있다.

4. 결론 및 향후과제

여러 검색엔진에서 연관검색어 서비스를 제공하고 있지만 사용자에게 꼭 맞는 연관검색어 서비스는 한계가 있다. 본 논문에서는 개인화 검색서비스 구현의 문제점을 다소나마 해결하고자 사용자 질의패턴을 분석하여 사용자들의 취향변화 즉 시간에 따라 변하는 관심영역 또는 사용자 질의 변화에 대해서 질의 패턴을 분석하고 추천 에이전트는 생성된 연관질의 리스트(Rule)을 이용하여 사용자에게 적당한 질의를 추천하는 사용자 맞춤형 질의 추천 방안을 제시하였다.

향후 연구계획으로는 실제 시스템 구현 및 평가하여 제한 사항을 도출해 보고 좀 더 효율적인 개인Profile 구축방안과 추천 Agent와 웹 클로링 Agent와의 협력을 통한 검색 효율 향상에 관해 연구가 필요하다.

참고문헌

[1] 김지하, 곽주현, 김효래, 이창훈 "비 감독학습 클러스터링을 이용한 웹 에이전트 효율성 향상에 대한 연구", 한국정보처리학회 추계학술발표 논문집 제7권 제2호, 2000

[2] 신민수,황준원,김성학,이창훈 "구매자의 구매패턴을 이용한 상품추천서비스에 대한 연구", 한국정보처리학회 추계학술발표논문집 제7권 제2호, 2000.

[3] 이소영, 정영미, "웹 포털 이용자 로그 데이터에 기반한 개인화 검색 서비스 모형의 설계 및 평가", 정보 관리학회지, 제23권 제4호, 2006.

[4] IProspect, Search Engine User Behavior Syudy, 2006.
http://iprospect.com/WhitePaper_2006_SearchEngineUserBehavir.pdf

[5] J. R. Wen, J. Y. Nie and H. J. Zhang. "Clustering user queries of a Search Engine". In Proceedings of the Internation World Wide Web conference, 2001.