

특화된 정보에 대한 Tag 기반 연속정보검색 시스템

이기은*, 박영호*

*숙명여자대학교 멀티미디어학과

e-mail : {kieun315, yhpark}@sookmyung.ac.kr

A Continuous Information Retrieval System Based-on Tag for Specialized Data

Ki-Eun Lee*, Young-Ho Park*

*Dept. of Multimedia Science, Sookmyung Women's University

요 약

정보화 사회에 접어들면서 정보를 판별하는 능력이 중요시 되고 있다. 그러나 정보가 점점 이질적이고 방대해 짐에 따라 사용자의 의도와 목적에 맞는 정보를 빠르고 정확하게 찾아내는 것이 어렵다. 정보검색 서비스를 제공하는 국내외 포털 사이트에서는 랭크 알고리즘을 이용하여 사용자에게 정보를 제공한다. 그러나 사용자의 요구를 충족시키기 위해 랭킹보다 더 중요한 것 정보를 압축시켜 사용자에게 사용자가 원하는 정보만 제공하는 것이다. 따라서 본 논문에서는 도메인을 제한하여 특화된 정보를 제공하며 사용자 위주의 더 특화된 정보를 제공하는 친구추가 기능을 제안한다. 동시에 주 검색 기능으로 사용자가 등록한 태그 링크를 따라 클릭하면서 연속적으로 정보 검색을 할 수 있는 연속정보검색을 제안한다. 그리고 제안한 시스템을 실제 웹 사이트를 구현을 통해 나타낸다. 제안한 시스템은 사용자에게 효율적으로 유용한 정보를 제공하는 기대효과가 있다.

1. 서론

우리는 특정 목적을 가지고 그 목적을 이루기 위해 검색을 한다. 그러나 정보가 점점 이질적이고 방대해 짐에 따라 사용자의 의도와 목적에 맞는 정보를 빠르고 정확하게 찾아내기란 쉽지 않다.

효율적인 정보검색을 위해 국내외 포털 사이트 [5][6]에서는 랭크 알고리즘을 구축하여 검색결과를 보여준다[2]. 그러나 방대한 정보 속에서 랭킹 기술보다 더 의미 있는 것은 사용자의 목적에 맞는 정보만 압축 하여 특화된 정보를 제공하는 것이다.

하나의 주제를 도메인으로 제한하여 정보를 제공하는 대표적인 사이트로 커뮤니티 형식의 카페[8]가 있다. 카페는 사용자의 참여, 개방과 공유가 활발한 Web2.0의 사례 중 하나이다. 그러나 카페는 사용자의 참여율이 높지만 게시판 중심의 사이트이기 때문에 검색을 통해 정보를 얻기에는 불편하다.

그러므로 본 논문에서는 특화된 정보만 다루어 사용자가 원하는 정보의 범위를 벗어나지 않되, 사용자가 중심이 된 연속정보검색을 통해 검색이 활성화 된 커뮤니티 구축 시스템을 제안한다.

본 논문에서 제안하는 방법은 도메인 제한 검색(Domain Restricted Search)이다. 포털 사이트의 정보과다 문제를 해결하기 위해 도메인을 제한하여 사용자에게 특화된 정보를 제공한다. 본 논문에서는 구축한 시스템을 테스트해보고자 테스트 정보로 도메인을 제한 하였다.

두 번째로 제안하는 방법은 사용자가 필요한 정

보를 연속적으로 검색 할 수 있도록 하는 연속정보검색이다. 하나의 Contents 속에 그와 연관된 키워드에 링크가 걸려있어 데이터를 읽다가 더 많은 정보가 필요할 경우 링크를 클릭하여 새로운 검색 결과를 얻을 수 있다. 연속정보검색은 User Created Contents와 사용자가 입력한 태그를 기반으로 검색이 제공된다.

세 번째로 제안하는 방법은 친구추가 기능기반 정보 제공이다. 나와 취향이 비슷한 사람이 갔던 레스토랑은 나도 좋아할 것이라는 것을 모티브로 한다. 관심 패턴이 비슷한 아이디를 친구로 등록하여 그 사람의 성향을 모방하는 데 의미가 있다. 서로 친구가 된 아이디끼리 Equivalent class로 묶어 서로의 글을 쉽게 접근할 수 있어 사용자들은 더 특화된 정보를 제공 받는다.

본 논문의 구성은 다음과 같다. 2장에서 시스템의 개요와 구성에 대해서 설명하고 3장에서 주요 기능에 대한 동작과 실제 구현한 것을 보이고 분석 한다. 마지막으로 4장에서 본 논문의 결론을 맺는다.

2. 시스템 개요 및 구성

2.1 시스템 개요

시스템은 크게 등록과 검색으로 나누어 진다. 등록은 클라이언트가 글과 태그를 등록하면 데이터베이스의 board 테이블과 keyword 테이블에 데이터가 저장되는 것이다.

검색은 세가지 방법이 있다. 첫 번째 방법은 일반적인 검색방법으로, 검색 창에 입력된 키워드는 text

parser 를 통해 단어 단위로 나뉜 후 Full Text Index 를 통하여 검색이 이루어 지고 작성자 검색에 대해서는 스트링 매치가 이루어 진다. 두 번째 방법은 도메인 제한 검색으로, BOARD 테이블의 SORT 속성을 통해 도메인을 제한하여 검색 결과를 보여준다. 마지막은 연속정보검색으로, 사용자가 글을 읽을 때 시스템은 내부적으로 Script 파일을 실행한다. Script 파일은 데이터베이스의 KEYWORD 테이블의 모든 태그를 출력한 후, 본 문의 단어와 비교하여 테이블의 태그와 같은 단어가 있으면 하이퍼링크를 걸어준다. 클라이언트가 링크를 클릭하면 링크에 대한 새로운 검색결과 리스트가 나타난다. 그림 1 은 클라이언트와 서버 간 검색이 이루어지는 방식을 나타낸다.

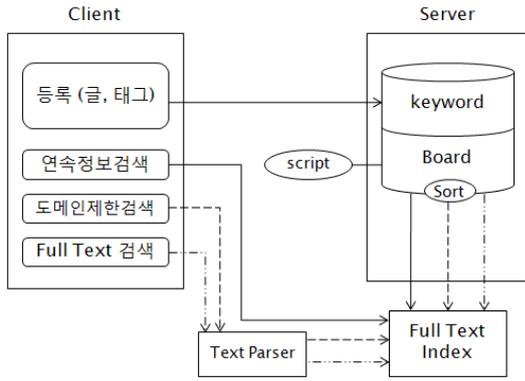


그림 1. 클라이언트-서버 간 검색 구조

2.2 시스템 주요 기능

본 절에서는 시스템의 검색기능과 친구추가 구성에 대해 설명한다.

2.2.1 검색

본 절에서는 본 시스템에서 제안하는 다양한 검색 기능을 설명한다. 본 시스템은 사용자의 편의에 맞는 다양한 검색 기능을 제안한다.

2.2.1.1 일반검색

본 논문에서 제안하는 시스템에서 사용하는 일반적인 검색방법은 두 가지가 있다.

첫째, 검색창에 키워드를 입력하여 검색하는 방법이다. 사용자가 입력한 키워드는 모든 게시판의 제목, 내용, 작성자에 대해서 검색된다.

제목과 내용에 대한 검색 방법은 BOARD 테이블의 TITLE 과 CONTENT 애트리뷰트에 대해서 인덱싱 후 MS_SQL 의 CONTAINS 연산자를 이용하여 Full Text 검색을 실시하고, 유니크한 데이터 추출을 위해 SELECT 절에서 DISTINCT 예약어를 사용한다.

작성자에 대한 검색 방법은 LIKE 연산자를 이용한 스트링 매치 방법을 적용한다.

둘째, 태그를 클릭하여 검색하는 방법이 있다. 메인 화면에서 보여지는 ‘인기 Tag’에서 태그를 클릭하거나 글을 읽을 때 글 내용 안에 있는 태그 링크를 클릭하여 검색하는 방법이다. 클릭한 태그 키워드에 대해 키워드 입력 검색과 같은 방법으로 제목, 내용, 작성자에 대해서 검색된다.

2.2.1.2 도메인 제한 검색

도메인 제한 검색은 검색 할 때 범위를 한정하여 사용자가 지정한 범위 내에서 검색 결과를 보여주는 것이다. 도메인을 제한해서 검색하면 사용자가 요구하는 정보 검색에 집중 할 수 있다. 그림 2 는 도메인을 선택하고 검색하는 화면이다.

검색 방법은 BOARD 테이블의 SORT 애트리뷰트를 이용한다. SORT 애트리뷰트는 글이 어떤 도메인에 속하는지를 나타낸다. 사용자가 도메인 선택창에서 데이트코스, 맛집, 베스트리뷰, 라운지 중 하나를 선택하여 검색하면 내부적으로 실행되는 쿼리는 WHERE 절에서 SORT 애트리뷰트를 도메인 선택한 것과 일치시키고 2.2.1.1 에서 설명한 키워드 입력 검색과 같은 방법으로 입력된 키워드를 파싱하여 제목, 내용, 작성자에 대해 검색한다.



그림 2. 도메인 선택창과 검색창

2.2.1.3 연속정보검색

본 시스템은 사용자가 등록한 글과 태그에 대해 사용자 위주의 연속정보검색을 제공한다. 사용자가 글을 읽을 때 시스템은 내부적으로 KEYWORD 테이블의 모든 KW 애트리뷰트 값을 추출하여 구분자 ‘|’로 연결하여 tag_list 변수에 저장한다. 그리고 tag_list 를 파라미터로 가지는 자동으로 링크를 걸어주는 JavaScript 함수 tag_link 를 호출한다. tag_link 함수는 구분자 단위로 tag_list 와 본문의 단어를 비교하여 같은 단어가 있으면 HTML 의 <A>태그를 사용하여 키워드에 링크를 걸어준다. 링크를 클릭하면 2.2.1.1 에서 설명한 태그 클릭 검색 방법과 같이 검색된다. 그림 3 은 연속정보검색의 흐름도이다.

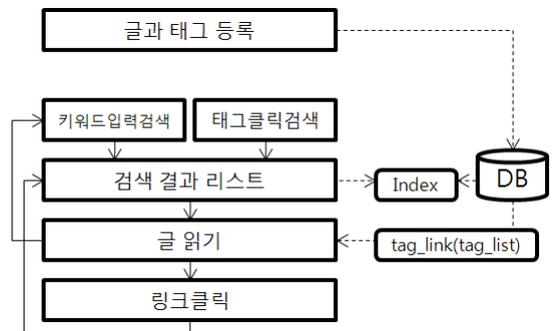


그림 3. 연속정보검색 흐름도

2.3.2 친구추가

본 시스템에서는 자신과 성향이 비슷한 아이디를 찾아 친구로 등록 할 수 있는 친구 추가 기능을 지원한다. 친구로 등록 하고 싶은 아이디를 클릭하고 친구 신청을 한 후, 친구 신청을 받은 아이디가 수락을

하면 두 아이디는 서로 친구관계가 된다. 친구관계에 있는 아이디끼리는 친구가 최근에 등록한 글을 로그인 했을 때 메인 화면에서 쉽게 볼 수 있다.

3. 시스템 구현

본 장에서는 실제 구현된 시스템의 세부메뉴인 회원관리, 게시판, 태그, 친구추가, 검색의 구현 내용을 그림과 함께 설명한다.

3.1 회원관리

본 시스템은 회원가입 후 로그인을 통해서 시스템 내의 기능을 이용할 수 있도록 설계되었다. 로그인을 하지 않은 경우는 글쓰기와 리플 쓰기를 할 수 없으며, 회원공개로 등록된 글을 읽을 수 없다. 또한 친구추가 기능을 이용할 수 없다. 로그인을 하지 않았을 때 일부 기능을 제한하는 이유는 원활한 사용자의 참여를 위해서 이다.

3.2 게시판

본 시스템에는 데이트코스, 맛집, 베스트리뷰, 라운지 네 개의 게시판이 존재한다. 네 개의 메뉴는 게시판 형식으로 구성되며 사용자는 데이트코스, 맛집, 라운지에 글을 등록할 수 있다. 베스트리뷰에는 데이트코스, 맛집, 라운지의 글 중에 조회수가 높은 글이 1 위부터 100 위까지 자동으로 등록된다. 메인 화면에서는 게시판에 등록된 글 중 조회수가 높은 인기글과 최근글을 볼 수 있다.

게시판은 기본적으로 글쓰기와 읽기, 태그 등록, 사진 등록, 비회원 이용 제한 설정, 리플 달기 기능이 있다. 태그는 글의 내용과 연관된 키워드를 다섯 개 까지 등록 할 수 있으며 영문 세 글자 이상, 한글 두 글자 이상만 태그로 등록 가능하다.

3.3 태그

메인 화면의 검색창 밑에 ‘인기 Tag’가 있다. 인기 Tag 는 작성자가 입력한 태그 중 가장 많이 등록된 상위 10 개의 태그를 1 위부터 10 위까지의 순서로 노출한 것이다. 사용자는 키보드를 사용하지 않아도 인기 Tag 의 태그를 클릭하여 자동으로 검색 결과를 얻을 수 있으며 글을 읽으면서 링크가 걸린 태그를 클릭하여 간편하게 연속정보검색을 할 수 있다. 그림 4 는 인기 Tag 화면이다.



그림 4 인기 Tag

3.4 친구추가

본 시스템은 회원인 사용자들에 대해 친구추가 기능을 제공한다. 친구 관계를 맺고 싶은 아이디를 클릭하면 친구 신청을 하는 창이 나타난다. 그림 5 는 친구 신청을 보여주는 창이다.



그림 5. 친구 신청

친구 신청을 한 후, 친구 신청을 받은 아이디가 수락을 하면 두 아이디는 서로 친구관계가 된다. 친구관계에 있는 아이디끼리는 친구가 최근에 등록한 글을 로그인 했을 때 메인 화면에서 쉽게 볼 수 있다. 그림 6 은 메인 화면에서 친구의 글을 볼 수 있는 ‘MY 친구글’ 화면이다.

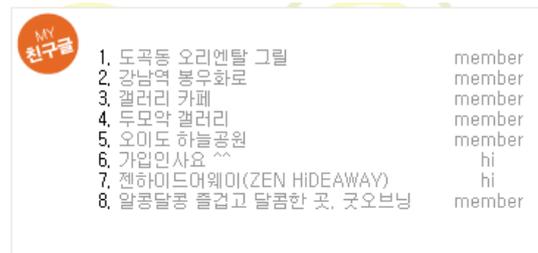


그림 6. 메인화면의 MY 친구글

3.5 검색

본 시스템은 검색기능에 중점을 두고 디자인 되었기 때문에 사용자는 사이트 내 어디서나 검색창을 볼 수 있으며 사이트에서 제공하는 다양한 검색 방법을 이용하여 원하는 정보를 얻을 수 있다.

본 시스템은 빠른 검색을 위하여 기본적으로 Full Text 인덱스를 이용한 텍스트 검색기능에 도메인제한 검색과 연속정보검색을 적용한다. 모든 검색은 글의 제목, 내용, 작성자에 대해서 검색하고 검색 결과는 조회수가 높은 글이 상위에 나오도록 구현하였다. 그림 7 은 하나의 키워드에 대한 일반검색과 도메인 제한 검색의 결과이다.



그림 7. 일반 검색과 도메인 제한 검색 결과

그림 8 은 연속정보검색을 할 수 있도록 페이지에 링크가 걸린 화면이다.

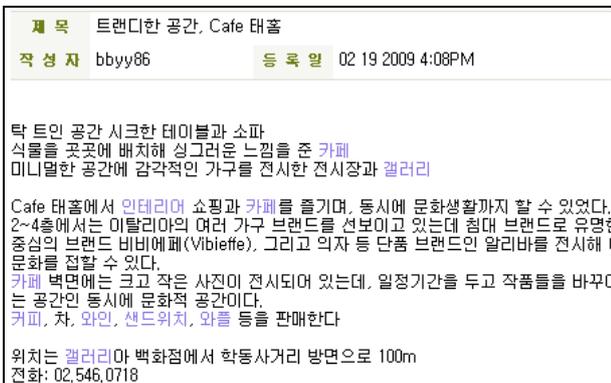


그림 8. 링크 걸린 페이지

4. 결론 및 기대효과

본 논문에서는 특화된 정보에 대해 User Tagging 기반의 연속정보검색과 도메인 제한 검색 시스템을 제안하였다. 제안된 방법은 사용자가 작성한 글과 태그를 기반으로 연속정보검색을 사용하고 사용자간에 친구관계를 맺어 친구의 글을 빠르게 접할 수 있는 친구추가 기능을 사용한다. 제안한 시스템을 실제 웹 사이트 구현을 통해 나타내었다.

참고문헌

- [1] 박동진, 김인환, 이재식, 이종운, “디렉토리형 검색 엔진의 문서분류 시스템에 관한 연구”, 한국경영 과학회, pp.485-488, 2001.
- [2] 이은정, 이민주, 이승희, 박영호, 김목련, 안후영, “신뢰성 있는 웹 전문지식 검색을 위한 하이브리드 랭크 매트릭스 제안”, 한국 디지털 콘텐츠 학회, Vol.9, No.4, pp.625-633, 2008.
- [3] Daniel C. Berrios, “Automated Indexing for Full Text Information Retrieval”, AMIA Symp., Vol.71, No.5, pp.71-65, 2000.
- [4] 손승범, 황운영, 이경하, 이규철, “순위부여를 지원하는 웹 서비스 검색 엔진의 개발”, 한국컴퓨터종합 학술대회논문집, Vol.33, No.1, pp.1-3, 2006.
- [5] 네이버, <http://www.naver.com/>
- [6] 다음, <http://www.daum.net/>
- [7] 미디어다음, <http://media.daum.net/>
- [8] 다음카페, <http://cafe.daum.net/>