

## 신뢰성 있는 웹 전문지식 검색을 위한 하이브리드 랭크 매트릭스 제안

이은정\*, 이민주\*\*, 이승희\*\*, 박영호\*\*\*, 김복련\*\*\*, 안후영\*\*\*

### 요약

최근 사용자들의 참여, 개방, 공유가 주요 이슈로 떠오르면서 전문적이고 정확한 정보를 웹에서 찾고자 하는 사용자의 요구가 증가하고 있다. 그러나 정보의 범람으로 사용자가 원하는 정보를 정확하게 검색하기 어려우며, 원하는 정보를 검색해도 그 정보에 대한 신뢰성을 판단하기가 어렵다. 따라서 본 논문에서는 신뢰성이 결여되기 쉬운 특정 정보에 대한 신뢰성과 검색의 효율성을 높이기 위해 새로운 랭크 매트릭스를 제안한다. 그리고 제안한 새로운 랭크 매트릭스를 적용한 민간의학 정보 웹 사이트를 구현한다. 제안하는 매트릭스를 사용자 레벨에 기반 하여, 레벨에 따른 평가 가중치를 차등화하여 글의 점수를 부여하는 방법이다. 이러한 방법은 참여자의 심리적 요소를 반영하여 글의 신뢰도를 높이는 방법으로 신뢰성이 결여되기 쉬운 정보의 신뢰도를 높일 수 있는 방안으로 사용될 수 있다.

## A Proposal on Hybrid-Rank Metrics for Retrieval of Reliable Expert Knowledge in Web

Eun-Jung Lee\*, Min-Joo Lee\*\*, Seung-Hee Lee\*\*, Young-Ho Park\*\*\*, Mok-Ryun Kim\*\*\*, Hoo-Young Ahn\*\*\*

### Abstract

Recently, the participation, opening and joint ownership of the users are important issue. The users want professional and accurate information from web. But users often suffer from retrieving accurate information. Even though the users find information they want, it is not guaranteed that the information is reliable since there are too much information placed on the web. Thus, we propose the novel rank metric to promote reliability and efficiency in information retrieval. In order to verify our approach, we implement a web site based on the proposed rank metric for nonofficial medical science information. The proposed rank metric based on user's level. This is to give score of text through differential rate depending on the user's level. The proposed rank metric enhances the reliability of text which is reflecting the user's mental factor. Thus, this method can be used for enhancing the reliability of text.

Keywords : Rank, Metrics, Hybrid, Rank Metrics, Retrieval, Web

### 1. 서론

최근 Web 2.0 시대가 도래하면서, 인터넷 사용자들에게 정보의 개방, 공유, 참여는 주요 이

슈가 되고 있다. 사람들은 참여를 통해 정보를 생성하고, 다른 사람들에게 개방함으로써 정보를 공유한다. 이를 통해 정보는 새롭고 방대한 지식으로 발전한다. 그러나 정보의 범람은 다음과 같은 문제를 야기하기 때문에 결국 사용자들은 정확하고 원하는 정보를 얻기에 많은 어려움이 있다.

첫째, 일반 사용자들이 제공한 정보는 신뢰도가 보장되지 않는다. 즉, 사용자가 찾은 정답이 사실성이 있는 정보인지, 정보의 원천이 확실한 것인지에 대해 신뢰성을 입증하기 어렵다.

※ 제일저자(First Author) : 이은정  
접수일:2008년 09월 05일, 완료일:2008년 11월 27일  
\* 숙명여자대학교 멀티미디어과학  
mm\_jung@naver.com  
\*\* 숙명여자대학교 생명과학과  
\*\*\* 숙명여자대학교 멀티미디어과학

둘째, 인터넷에는 방대한 정보들이 존재한다. 따라서 찾고자 하는 정보 외에 불필요한 정보가 제공될 수 있고, 필요한 정보는 얻지 못할 수 있다. 특히 전문 분야에 대한 정보검색의 경우는 더욱 그렇다[1].

셋째, 검색결과에 관한 문제점이다. 기존 검색 방법은 키워드 기반에 의한 질의만 지원하여 사용자의 정보 요구를 만족시키지 못한다[2]. 또한 사용자의 요구에 맞는 정보를 상위에 보여주는 것은 단연 중요한 문제이지만, 현재는 일반적인 선호도 위주로 결과가 산출되고 있다.

이러한 문제점들에 대해서 국내·외 포털 [3][10]이나 대형 웹 사이트들은 자체적인 랭크 알고리즘을 구축하여 검색 결과를 보여준다. 그러나 수익성 위주의 서비스로 인해 사용자의 요구보다는 상업적인 검색결과를 제공하고 있어 순위의 왜곡되어 찾고자 하는 정보를 정확히 획득하는 것이 어렵다.

전문 콘텐츠를 다루는 웹 서비스의 경우에는 정보의 평가방법이 합리적이지 못하다. 정보원의 신뢰성에 대한 판단은 최종적으로는 사용자가 하는 것이지만, 전적으로 사용자의 판단에만 맡겨 둘 수는 없는 문제로 평가 방법과 척도의 개발이 필요하다[5]. 그러므로, Web 2.0의 정신을 도입하되, 신뢰도가 높은 유용한 정보 환경을 제공할 필요성을 인식하고 언급한 문제점들에 착안하여 해결방법을 제시하고자 한다.

본 논문에서 제안하는 시스템은 다음과 같다.

첫째, 한정된 지식을 다루는 사이트를 기준으로 자체적으로 고안한 하이브리드 랭크 매트릭스를 사용하여 객관적이고 합리적인 평가가 이루어지도록 한다. 이 방법은 각 사용자가 쓴 글에 대해 세 가지 방향으로 평가한다. 그리고 글을 작성한 사용자가 점수를 얻어 사용자의 레벨을 높임으로써 사용자가 좋은 글을 통해 인정받고자 하는 욕구를 반영하고, 다른 이의 글을 평가할 때 사용자의 레벨에 따른 가중치를 적용하는 랭크 매트릭스를 제시한다.

둘째, 본 시스템에서는 민간의학이라는 입증되지 않은 지식을 다루는 콘텐츠를 바탕으로 객관성과 신뢰성을 가진 웹 서비스 시스템을 구현한다. 다양한 게시판을 구성하여 세부적인 지식의 교환이 가능하도록 하고, 이 중 명예의 전당에는 랭크 시스템을 통해 신뢰성이 있다고 판단된 적

합한 정보를 이동시키고, 신뢰도가 낮아지면 다시 명예의 전당에서 제외되도록 함으로써 현존하는 정보가 가장 신뢰도가 높을 수 있는 가치 체계를 확립한다.

셋째로 신뢰도 측정을 위해 신뢰성 입증 가능한 다양한 기법이 필요하고, 사람의 생각, 습관, 그리고 관습을 기반으로 하여 신뢰도가 낮은 지식을 보완하기 위해 멀티미디어 매체를 탑재할 수 있는 구조의 시스템을 구현하였다. 본 논문에서는 다음과 같은 공헌을 제시한다.

- 신뢰성이 결여된 도메인 지식의 신뢰도 보장을 위한 랭크 매트릭스의 제안
- 정보 검색의 다양한 방법을 제공하고, 랭킹 기반 검색 결과 정렬을 지원
- 멀티미디어 매체를 탑재할 수 있는 웹 서비스 시스템의 구현

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구와의 비교를 통하여 제안하는 시스템의 특징을 설명한다. 3장 본문에서는 시스템의 구성 및 주요 기능에 대해 설명한다. 4장에서는 시스템 구현에 대해 구체적으로 설명한다. 마지막으로 5장에서 본 논문의 결론을 맺는다.

## 2. 관련 연구

본 장에서는 전문적인 지식을 공유하기 위한 정보원(Information Source)들에 대한 비교분석을 통하여, 제안하는 방법과의 차이점을 보인다. 이를 위해 특정 주제에 대한 지식검색 기능을 제공하는 네이버지식IN[11], 노트북에 대한 정보를 다루는 NB인사이드[12], 웹에 산재된 민간의학 사이트, 카메라에 관한 정보를 다루는 DC인사이드[13]에 대해 비교한다.

네이버 지식IN[11]은 특정 지식에 대한 질문과 답변으로 구성되어 있다. 누구나 참여할 수 있다는 장점이 있지만, 답변에 대한 질문자의 주관적인 점수 부여 방식으로 인해 신뢰성에 대한 기준이 부족하다는 것이 단점이다[4].

NB인사이드[12]와 DC인사이드[13]는 디지털 카메라와 노트북에 대한 전문적인 정보를 제공한다. 사용자들이 스스로 디지털기기 정보, 리뷰

와 이미지를 제공하며 웹상에서 트렌드와 이슈를 만들어내고, 사용자들의 참여와 공유가 뛰어난 특성을 가진다. 그러나 글에 대한 평가체계가 없으며, 그로인해 글의 신뢰도는 평가되지 않는 단점이 있다.

또한, 웹에 산재한 민간의학이나 민간요법 사이트는 주로 운영자에 의해 주어진 정해진 정보로 구성되어 있으며, 상업적인 판매를 유도하기 위한 경우가 많다. 그러므로 사용자의 참여가 어렵거나 한정되어있고, 단순 추천 이외에는 평가체계가 없어 게시물에 대한 신뢰성이 떨어진다.

그 밖의 선행 연구들은 신뢰성과 상관없는 랭크 알고리즘을 제안하였다[5]. 또한, 검색에 있어서 풀 텍스트 인덱스를 사용하는 방법을 한정도메인이 아닌 웹상에서 적용하기 위한 연구가 선행되었다[6].

본 연구에서는 인간의 심리적 욕망에 기반한 [7] 네이버 지식IN과는 다른 방법으로, 글에 대한 신뢰도와 참여도를 높이는 방법을 제안한다. 그것은 사용자 피드백 방식에 기반한 사용자 레벨 변경과 사용자 레벨에 따른 평가 가중치 부여 기능이다. 이러한 방법은 사용자의 레벨 향상 욕구와 향상된 레벨에 부합된 영향력을 행사할 수 있기 때문에 사용자의 자연스런 참여가 이루어지며, 양질의 게시물이 우대받을 수 있다는 심리적 요소로 인해 신뢰도 높은 게시 환경을 만든다. 또한 사용자가 이미 작성한 글의 평가점수나 활동사항에 따라 사용자 레벨을 측정하고, 레벨에 따라 글에 대한 평가 점수가 차등 부여된다. 이를 통해서 전문가와 비전문가를 포함한 다양한 레벨의 사용자들의 참여와 평가를 반영할 수 있다는 장점이 있다.

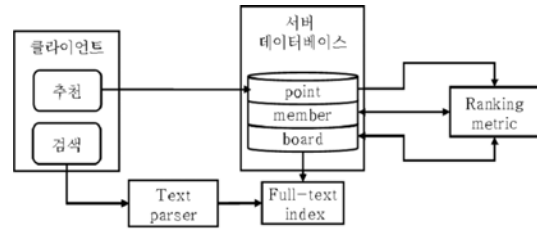
### 3. 본론

본 장에서는 시스템 구조 및 데이터베이스의 구성, 시스템의 주요 기능을 설명한다.

#### 3.1. 시스템 구조

제안하는 시스템의 구조는 (그림 1)과 같다. 시스템은 사용자가 작성한 정보를 저장하는 데이터베이스를 중심으로, 클라이언트의 추천, 검색요청을 처리하는 모듈과, 정보의 신뢰도 및 사

용자의 랭킹에 핵심적인 역할을 수행하는 Ranking Metric, 효과적인 검색을 위한 Text Parser, Full-text index로 구성된다.



(그림 1) 클라이언트-서버 간 랭킹과 검색 구조

시스템 데이터베이스의 구성은 3.2절에서 살펴보고, 랭킹 및 검색의 구체적인 방법은 3.3절의 주요 기능에서 살펴본다.

#### 3.2. 데이터베이스 구성

본 절에서는 서버에서 구성된 데이터베이스에 대해 설명한다. 서버는 크게 입력부와 질의부, 사용자 관리부로 구성되어있다. 입력부는 웹상에서 사용자의 요구를 받아 파싱하고, 파싱된 입력을 바탕으로 검색 대상을 선정하여 질의를 수행하는 과정이 질의부이다. 다음에서 입력부, 질의부, 사용자 관리부에 대해 순서대로 설명한다.

입력부는 일반 게시판에 대한 정보를 담고 있는 GBOARD 테이블, 전문가칼럼에 대한 정보를 담고 있는 EBOARD 테이블, 리플에 대한 정보를 담고 있는 REPLY 테이블이 있다.

로그인한 사용자가 일반 게시판에 글을 입력하면 GBOARD테이블의 TITLE 애트리뷰트에 글의 제목, AUTHOR 애트리뷰트에 작성자의 아이디, CONTENT 애트리뷰트에 글의 내용이 저장된다. GBOARD테이블의 BNAME과 CNUM 애트리뷰트에는 인덱스를 구성하였다.

KEYWORD 테이블의 KW1-KW5 애트리뷰트에 키워드들이 저장되고, MULTI\_DATA 애트리뷰트에 참고자료의 패스가 저장된다.

사용자 레벨이 6 이상인 사용자가 전문가 칼럼에 글을 입력하면 EBOARD테이블의 TITLE 애트리뷰트에 글의 제목, AUTHOR 애트리뷰트에 작성자의 아이디, CONTENT 애트리뷰트에 글의 내용, MULTI\_DATA 애트리뷰트에 참고자료의 패스가 저장된다. EBOARD테이블의 CNU

M 애트리뷰트에는 인덱스를 구성하였다.

작성된 글에 대해 리플을 입력하면, REPLY 테이블에 REPLYER 애트리뷰트에 리플을 작성한 사용자의 아이디가 자동으로 저장되고, REP 애트리뷰트에 사용자가 입력한 리플 내용이 저장된다.

질의부는 글에 대한 점수 정보를 담고 있는 POINT 테이블, 주요 키워드에 대한 정보를 담고 있는 KEYWORD, KW\_CNT 테이블이 존재한다. 로그인한 사용자가 글에 대해 점수를 입력하면 POINT 테이블의 PP 애트리뷰트에 실용성 점수, EP 애트리뷰트에 효과성 점수, RP 애트리뷰트에 신뢰도 점수가 추천인의 레벨 애트리뷰트 LV와 함께 저장된다. POINT 테이블에 존재하는 해당 글에 대한 점수를 불러들여 랭크 점수식을 적용하고, 이에 따라 도출된 점수는 GBOARD 에 저장된다. 해당 글 작성자에 대한 점수 역시 POINT 테이블에 존재하는 글 점수를 불러들여 랭크 점수식을 적용한 뒤 MEMBER 테이블에 저장된다.

사용자 관리부는 사용자 관리를 위한 MEMBER 테이블이 존재한다. ID 애트리뷰트에 사용자 아이디, PASSWD 애트리뷰트에 비밀번호, NAME 애트리뷰트에 사용자 이름, SSN1과 SSN2에 주민번호, MAIL 애트리뷰트에 메일주소가 저장된다.

GBOARD와 EBOARD의 TITLE과 AUTHOR, CONTENT 애트리뷰트에는 풀 텍스트 인덱스를 구성하였다 풀 텍스트 인덱스는 1분마다 전체 채우기를 시행한다.

### 3.3. 시스템 주요 기능

본 절에서는 시스템의 주요 기능인 정보의 랭킹 및 검색을 위한 정보의 점수 체계, 융합 검색, 레벨 중심 지식 체계 구성에 대해 설명한다. 그리고 사용자가 정의한 멀티미디어 매체를 탑재하는 멀티미디어 참여 광장에 대해 설명한다.

#### 3.3.1 점수 체계

본 시스템의 점수체계는 한 가지 정보에 대해서 가지 방면으로 평가함으로써 신뢰성 향상에 기여한다. 구현은 로그인 한 사용자가 해당 글에서 '글 추천' 버튼을 누르면 실용성, 효과성, 신뢰성 점수를 입력하게 된다. 입력된 점수를 기

으로 POINT 테이블의 애트리뷰트 PP, EP, RP 점수를 사용자 레벨 애트리뷰트 LV와 곱하여 사용자 레벨에 따른 가중치를 준다. 이것은 레벨이 n인 사용자가 10점을 주면 10점을 n번을 주는 것과 같은 효과이며, 레벨이 높은 사용자일수록 글의 점수에 큰 영향력을 가진다. 이 점수를 기반으로 실용성, 효과성, 신뢰성점수 각각의 평균을 구한다음 표준 편차를 구하여 최종 점수인 Ranking Score(rscore)를 계산한다. 시스템에 사용된 랭킹 점수 식은 (그림 2)와 같다.

$$\begin{aligned} \bar{X} &= \frac{\sum_{i=1}^n (x_1 + x_2 + x_3) \cdot rLV \cdot n}{3 \sum_{i=1}^n rLV \cdot n} \\ STDEV &= \sqrt{\frac{\sum_{i=1}^n ((x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + (x_3 - \bar{X})^2) \cdot rLV \cdot n}{3 \sum_{i=1}^n rLV \cdot n}} \\ \text{RankingScore} &= (\bar{X} \cdot 1.2) - (STDEV \cdot 0.5) + ((3 \sum_{i=1}^n rLV \cdot n) \cdot 0.001) + (mLV \cdot 0.1) \end{aligned}$$

$x_1, x_2, x_3$  : 추천자의 평가 점수  
 $rLV$  : 추천자 레벨  
 $\bar{X}$  : 게시물의 평균 점수  
 $mLV$  : 게시자의 레벨

(그림 2) 랭킹 점수 식

글에 점수가 부여됨에 따라 해당 글을 작성한 사용자의 레벨에도 영향을 준다. 추천된 글의 작성자 레벨은 작성자가 등록한 모든 글에 해당하는 점수를 POINT 테이블에서 불러들여 위의 랭킹 매트릭스를 적용한 뒤, 추천인수와 랭크점수를 바탕으로 갱신하게 된다. 점수는 누적식이 아니라 추천 받을 때마다 새로 계산하게 되므로 사용자의 레벨은 언제든지 올라가거나 내려갈 수 있다.

본 논문에서 제안하는 글 점수 측정과 사용자 레벨 측정 요소를 정리하면 표 1과 같다.

<표 1> 점수 산정 요소

<b>글 점수 추정 요소</b>	실용성, 효과성, 신뢰성, 추천 사용자의 레벨
<b>사용자 레벨 추정 요소</b>	작성 글의 점수, 추천인 수, 랭크 점수

점수체계는 검색과도 연관성을 가진다. 검색은 3.3.2장에서 설명한다.

3.3.2 융합 검색

본 시스템에서는 사용자의 편의에 맞는 다양한 검색 방법을 제안하고, 모든 검색에 대한 결과가 글의 점수가 높은 순으로 정렬되는 랭킹 기반 정보추출을 적용한다. 검색 방법은 다음과 같다.

첫째, 작성자가 글 작성시 입력했던 키워드에 대한 검색 방법이 있다. 메인화면에서 보여지는 '주요키워드'란에서 관심있는 키워드를 클릭하거나 글 읽기 화면에서 보여지는 해당 글의 '키워드'를 클릭하면 같은 키워드를 포함하고 있는 글을 검색한다. 구현은 MS-SQL의 LIKE 연산자를 이용한 exact match를 하며, 게시판 테이블의 글 번호와 키워드 테이블의 글 번호가 일치해야 하므로 GBOARD와 KEYWORD 테이블을 조인시킨다.

둘째, 작성자의 아이디에 대한 검색 방법이 있다. 작성자 애트리뷰트 AUTHOR에 텍스트 인덱싱 후 풀 텍스트 검색을 사용하면 완전히 일치하는 아이디만 검색되므로 사용자의 편의를 위해 LIKE 연산자를 이용한 검색을 하여 부분적으로 일치하더라도 검색 결과를 출력하도록 했다.

셋째, 제목, 내용, 제목+내용 검색 방법이 있다. 사용자가 검색 범위를 선택하면 해당하는 범위 내에서 검색하는데, 전체 게시판 또는 세분화된 게시판 내에서 각 범위에 따른 검색이 가능하다. 이 검색 방법은 제목 애트리뷰트 TITLE과 내용 애트리뷰트 CONTENT에 대해서 인덱싱 후 MS-SQL의 CONTAINS 연산자를 이용하여 풀 텍스트 검색을 실시하고, 유니크한 데이터 추출을 위해 SELECT 절에서 DISTINCT 예약어를 사용한다. 풀 텍스트 검색은 띄어쓰기를 지원하지 않으므로, 사용자로부터 입력 받은 strin

g타입의 검색어를 블랭크 중심으로 단어를 추출하여 쿼리의 where절에 넣는다.

넷째, 모든 게시판의 제목, 내용, 작성자에 대해서 검색하는 통합검색 방법이다. 사용자가 입력한 검색어가 제목, 내용 또는 작성자에 포함되어 있는 글을 모두 결과로 출력하므로 사용자에게 다양한 정보를 접할 수 있게 한다. 이 검색 방법은 제목, 내용에 대해서는 CONTAINS 연산자를 이용한 풀 텍스트 검색을 적용하고, 유니크한 데이터 추출을 위해 SELECT 절에서 DISTINCT 예약어를 사용한다. 작성자에 대해서는 LIKE 연산자를 이용한 스트링 매치 방법을 적용한다.

3.3.3 지식체계 그룹핑 및 구성

본 시스템에서는 도메인 내용과 사용자 레벨별로 구성된 카테고리를 제공함으로써 사용자들이 세분화, 전문화된 정보를 획득할 수 있도록 한다. 지식 체계의 구성은 <표 2>과 같다.

<표 2> 레벨 중심 지식 체계 구성 (Level-wise categorization)

Level 1	검증된 그룹	명예의 전당, 전문가 칼럼
Level 2	한방요법 그룹	민간요법, 식이요법, 운동요법
Level 3	건강상식 그룹	건강상식, 건강요리, 기타정보

Level 3 그룹은 일상생활에서 건강 유지를 위해 알아두면 좋은 기본적인 정보들이고, Level 2 그룹은 질병을 치료하기 위한 정보들이다. 회원 가입 후 로그인한 모든 사용자들이 글을 작성할 수 있다. Level 1 그룹은 명예의 전당과 사용자 레벨이 6 이상인 사용자들이 글을 쓸 수 있는 전문가 칼럼으로 구성된다. 명예의 전당은 Level 2그룹과 Level 3 그룹의 글 중에서 추천인 수가 100 이상이고, 글의 점수가 10.00 이상인 글이 자동으로 이동되고 글의 점수가 낮아지면 다시 원래 레벨의 게시판으로 자동 이동 한다.

3.3.4 멀티미디어 참여 확장

멀티미디어 참여확장은 사용자가 글을 작성할 때 그림, 동영상, 웹사이트, 텍스트, 애니메이션

등 다양한 멀티미디어 자료를 글에 첨부 가능하도록 하여 사용자의 편의와 신뢰성을 증가시켰다. 사용자가 글 입력 시 첨부파일 란에 멀티미디어 자료의 패스를 입력하면 GBOARD와 EBOARD 테이블에 있는 애트리뷰트 MULTL\_DATA에 멀티미디어 자료의 패스를 저장하고, 사용자가 글을 읽을 때 첨부파일에 보여 지는 패스를 따라가면 자료를 액세스 할 수 있도록 하였다.

민간의학과 같이 생각, 습관, 관습에 기반 한 정보들은 신뢰도가 하락한다. 따라서 정보의 신뢰성 입증에 가능한 다양한 멀티미디어 자료를 통해서 신뢰성이 결여된 정보에 대한 보완을 가능하게 할 필요가 있다.

#### 4. 시스템 구현

본 장에서는 실제 구현된 시스템의 세부 메뉴인 사용자 관리, 사용자 랭킹, 게시판, 추천 및 랭킹 매커니즘, 검색의 구현 내용을 그림과 함께 설명한다.

##### 4.1. 사용자 관리

본 시스템은 회원가입 후 로그인을 통해서 시스템 내의 기능을 이용할 수 있도록 설계되었다. 로그인을 하지 않은 경우는 글쓰기와 추천기능을 이용할 수 없다.

회원가입을 한 사용자에게는 다음과 같이 랭킹 점수가 부여된다. 랭킹 점수의 항목에는 사용자의 신뢰도, 사용자 레벨, 사용자의 인기도가 있다. 초기 값은 신뢰도 0, 인기도 0, 레벨 1이다. 사용자의 신뢰도는 자신이 작성한 글의 점수에 의해 계산되는데 글의 점수가 사용자의 신뢰도에 어떻게 적용되는지는 3.3.1절에서 다루었다. 사용자 레벨은 경험적 수치로 1부터 6까지 있으며, 초기 레벨 1에서 사용자가 작성한 글의 총 추천자수와 신뢰도를 바탕으로 하여 한 단계 높은 레벨로 갱신되고, 글의 점수가 떨어지면 사용자의 레벨도 낮아질 수 있다. 사용자의 인기도는 사용자가 작성한 글의 조회 수 합과 같고, 본인이 작성한 글을 읽는 경우에는 조회 수를 카운트하지 않는다.

##### 4.2. 사용자 랭킹

사용자 랭킹 리스트는 메인 화면의 왼쪽 아래에 위치한다. 사용자 랭킹 리스트는 신뢰도가 높은 사용자를 상위 5명까지 표시해준다. 레벨은 신뢰도를 기준으로 하여 등급을 나눈 것이며 인기도는 신뢰도와는 상관없는 글의 조회 수를 바탕으로 한 점수이기 때문에 신뢰도를 기준으로 사용자를 정렬한다. 로그인 한 상태에서는 사용자 자신의 랭킹 점수가 표시되며, 다른 사용자의 랭킹점수는 작성자를 클릭하여 나타나는 팝업창으로 확인 할 수 있다. (그림 3)은 사용자 점수 조회와 사용자 랭킹 리스트에 대한 예이다.

/127.0.0.1:8080/pro_d...		사용자랭킹 회원랭킹입니다			
127.0.0.1:8080/pro_d/my_point2		ID	신뢰도	인기도	레벨
Ish 님의 점수입니다.					
신뢰도 점수 : 10.41		test023	11.99	97	1
인기도 점수 : 131		Ish	10.41	131	5
레벨 : 5		wg0529	9.02	298	1
작성 글 수 : 6		kkk	9.01	284	4
		test115	3.74	97	1

(그림 3) 사용자 점수 조회와 사용자 랭킹 리스트

##### 4.3. 게시판

본 장에서는 일반 게시판과 특수 게시판의 기능에 대해서 설명한다.

일반게시판은 민간치료법, 식이요법, 운동요법, 건강 상식, 건강요리가 있다. 기본적으로 글쓰기와 읽기, 주요 키워드 등록, 도메인 제한 검색과 추천기능이 가능하며 사용자 등급에 따른 이용 제한이 없다. 주요 키워드 등록은 사용자가 글을 쓸 때 중요하다고 생각되는 키워드를 다섯 개까지 등록 할 수 있는 기능이다. 추천기능은 3.4절에서 다루도록 한다.

특수게시판은 신뢰도가 높다고 판단될 때 글이 등록되는 게시판이고, 전문가 칼럼과 명예의 전당이 있다.

전문가 칼럼은 레벨이 6 이상의 신뢰도에 도달한 사용자만이 글을 작성할 수 있다. 레벨이 높다는 것은 신뢰도가 높은, 글을 많이 작성하였고, 많은 사용자에게 평가받은 사용자라는 의미가 있다. 따라서 시스템 사용자는 전문가 칼럼에 있는 글을 신뢰도 높은 글이라고 판단 할 수 있다.

명예의 전당은 글의 랭킹점수가 기준치 이상일 경우 일반게시판의 글이 자동으로 이동되므로 사용자가 명예의 전당에 직접 글을 등록하는 것은 불가능하다. 명예의 전당에 있는 글도 추천이 가능하며, 명예의 전당에 등록된 글도 점수가 낮아지면 원래의 게시판으로 내려올 수 있다. 명예의 전당은 높은 랭킹점수를 가진 글을 등록한 게시판이므로 사용자는 높은 신뢰도를 가진 글을 따로 찾는 번거로움 없이 명예의 전당의 글만 읽어도 신뢰도 높은 민간의학 정보를 얻을 수 있다. (그림 4)은 명예의 전당 글 목록 화면이다.

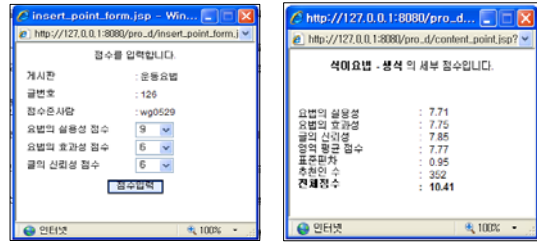
명예의전당		신뢰도 높은 민간요법게시물을 엄선한 곳입니다.		
게시판 제목	작성자	작성일	조회수	점수
운동요법 허리살을 빼고 싶을 때에는	wq052920080304	168	15.00	
식이요법 춘곤증에 좋은 음식	test11520080319	148	15.00	
민간치료 탈염증, 이형계 한번 해보세요	test02320080319	154	12.13	
건강요리 유자차 만들기	test02420080319	114	11.87	
식이요법 식이요법-생식	lsh	20080315	59	10.41

(그림 4) 명예의 전당 글 목록

#### 4.4. 추천 및 랭킹 매커니즘

본 시스템은 일반게시판과 명예의 전당 게시판의 글에 대해 추천하기 기능을 제공한다. 해당 글 작성자 외의 모든 사용자가 추천 가능하며, 하나의 글에 한번만 추천이 가능하므로 중복 피드백을 방지한다. 단, 동일 글에 대한 재 추천 기능을 허용한다.

추천 기능은 실용성, 효과성, 글의 신뢰성에 대해 경험적 범위(-10점-10점)내의 점수로 평가 가능하다. 한 가지 글에 대해 세 가지 항목을 융합하여 평가하여 세 가지 측면으로 부여된 점수를 바탕으로 글의 평균점수와 표준편차를 계산하여, 최종적으로 글에 대한 랭킹점수를 저장한다. 게시판 글 목록이나 글 읽기 창에서는 최종 랭킹점수만을 표시하며, 이 랭킹점수를 클릭하면 팝업창에 해당 글의 세부 점수가 표시된다. 표시되는 세부 점수는 세 가지 항목의 평균점수와 글의 전체 평균점수, 표준편차, 추천인수, 최종 랭킹점수가 있다. (그림 5)는 글 추천과 추천된 글의 세부점수를 보여주는 창이다.



(그림 5) 글 추천하기와 세부점수 표시

랭킹점수 계산 방법의 특징은 표준편차를 적용함으로써 같은 평균점수를 가지는 글이라도 높은 점수를 지속적으로 추천받은 글이 더 높게 평가된다. 또한 추천인수 적용으로 더 많은 사용자에게 추천받은 글이 더 높게 평가될 수 있으며, 레벨이 높은 사용자가 작성한 글은 신뢰도가 더 높다고 판단하여 이를 평가에 적용하였다. 레벨이 높은 사용자가 작성한 글은 레벨에 따른 가중치를 부여한다. 이는 레벨이 신뢰도를 의미하기 때문이다.

#### 4.5. 검색

본 시스템은 빠른 검색을 위하여 풀 텍스트 인덱스[8]를 이용한 텍스트 검색기능을 적용하고, 한 단어 이상의 검색어를 입력할 수 있는 다중 색인어 검색기능을 지원한다. 따라서 대규모 정보검색 분야에서 입증된 역 인덱스[9]를 기반으로 하므로 대용량성과 확장성이 우수하며, 기존 검색엔진과 유사한 한정된 도메인의 빠른 속도의 검색이 가능하도록 하였다.

통합검색에서는 (그림 6)과 같이 작성자, 제목, 내용에 대한 검색 결과를 얻을 수 있으며 검색 결과는 일반게시판 검색결과와 전문가칼럼 검색결과가 나뉘어서 나타나고, 랭킹점수가 높은 글이 상위에 나오도록 하였다.

번호	제목	작성자	작성일	조회수
1	공부방법 좋은 음식	mm_jung20080328	6	

번호	제목	작성자	작성일	조회수
12	감기에 좋은 여러가지 1스*	kkk	2008/03/19	16
11	감기에 좋은 여러가지2스*	kkk	2008/03/19	22
10	당뇨병 환자용 음식**	kkk	2008/03/15	15
9	따뜻한 배 성장차**	kkk	2008/03/18	31
8	겨울철 감기에방~~대추 성장차	lsh	2008/03/18	10
7	심장과 감기에 좋은 마늘꿀탕	mm_jung20080319	12	0.00
6	성장차 만들기	test023	2008/03/19	8
5	기침에 좋은 석이소법술름	test031	2008/03/19	15
4	심장과 감기에 좋은 마늘꿀탕	mm_jung20080319	10	0.00
3	감초차	test097	2008/03/25	2
2	선인장을 활용한 민간치료법	kkk	2008/03/15	16
1	감방를 위한 민간치료법	kkk	2008/03/17	13

(그림 6) 통합검색결과

또한 메인화면 왼쪽 하단에는 글의 작성자가 입력한 키워드 중 가장 많이 등록된 상위 20개의 키워드가 노출되며, 키워드를 선택하면 해당 키워드를 등록한 글들이 검색된다. (그림 7)은 주요 키워드 검색 결과를 나타낸다.

주요키워드 작성자 등록 키워드입니다.

키워드 검색결과 주요 키워드 검색 결과입니다.

번호	제목	작성자	작성일	조회수
8	심장과 감기에 좋은 마늘꿀탕	mm_jung20080319	11	
7	감기에 좋은 여러가지2스*	kkk	2008/03/19	23
6	감기에 좋은 여러가지 1스*	kkk	2008/03/19	16
5	심장과 감기에 좋은 마늘꿀탕	mm_jung20080319	12	
4	겨울철 감기에방~~대추 성장차	lsh	2008/03/18	10
3	따뜻한 배 성장차**	kkk	2008/03/18	31
2	감기에 는 성장차	kkk	2008/03/18	11
1	감기에 는 배탕 성장물	wj0529	2008/02/21	69

(그림 7) 주요 키워드 검색 결과

### 5. 결론

본 논문에서는 신뢰성이 결여된 정보에 대해서 정보의 신뢰성과 검색의 편리성 및 효율성을 부여할 수 있는 랭킹 매트릭스와 융합 검색 방법을 제안하였다. 제안된 방법은 사용자의 신뢰도와 추천을 기반으로 각 게시물에 대한 점수를 차등부여 하였고, 스트링 매치와 풀 텍스트 검색 방법을 융합한 새로운 검색 방법을 사용한다.

제안한 랭킹 매트릭스 및 검색 방법을 실제

웹 서비스 시스템 구현을 통해 분석하였으며, 데이터양이 많아지는 경우에도 효율적임을 보였다.

### 참 고 문 헌

- [1] 박동진, 김인환, 이재식, 이종운, “디렉토리형 검색엔진의 문서분류 시스템에 관한 연구”, 한국경영과학회, pp. 485-488, 2001.
- [2] 손승범, 황윤영, 이경하, 이규철, “순위부여를 지원하는 웹 서비스 검색 엔진의 개발”, 한국컴퓨터종합학술대회논문집, Vol.33, No.1, pp. 1-3, 2006.
- [3] Luiz Andre Barroso, Jeffery Dean, and Urs Hozle, “Web Search for a Planet : The Google Cluster Architecture”, IEEE Micro, Vol.23, No. 2, pp. 22-28, 2003.
- [4] 김영기, “웹 사이트의 신뢰성 평가에 영향을 미치는 요인과 각 요인의 중요도에 대한 연구”, 한국문헌정보학회지, Vol.41, pp.95, 2007.
- [5] Michael Collins, “Ranking Algorithms for Named-Entity Extraction:Boosting and the Voted Perceptron”, Association for Computational Linguistics(ACL), Vol.40, pp.489-496, 2002.
- [6] Sergey Melnik, Sriram Raghavan, Beverly Yang, and Hector Garcia-molina, “Building a Distributed Full-Text Index for the Web”, Transaction on Information System(TOIS), Vol.19, No.3, pp. 217-241, 2001.
- [7] 김희연, “웹에서 유통되는 정보· 지식의 신뢰연구”, 정보통신정책, Vol.19, No.8, 2007.
- [8] 장종인, “지식 생산 및 전달 양식의 변화: NAVER 지식검색 서비스에서 찾아본 건강지식 사례분석”, 정보통신정책연구원, Vol.18, No.16, 2006.
- [9] Daniel C. Berrios, “Automated Indexing for Full Text Information Retrieval”, AMIA Symp., Vol. 71, No. 5, pp. 71-75, 2000.
- [10] 임성채, 안준선, “역과일에 기반한 웹 검색 엔진의 랭킹 시스템 구현”, 한국정보과학회, Vol.34, No.2, 2007.
- [11] 네이버지식IN, <http://kin.naver.com/>
- [12] NB인사이드, <http://www.nbinside.com/>
- [13] DC인사이드, <http://www.dcinside.com/>





**이 은 정**

2005년-현재 : 숙명여자대학교  
멀티미디어과학(학사)

관심분야 : Data mining, Information Retrieval



**김 목 련**

2007년 : 숙명여자대학교  
멀티미디어과학(학사)  
2007년-현재: 숙명여자대학교  
멀티미디어과학(석사)

관심분야 : Information Retrieval(IR), indexing



**이 민 주**

2005년-현재 : 숙명여자대학교  
멀티미디어과학,  
생명과학(학사)

관심분야 : Bio-Informatics, microarray



**안 후 영**

2007년: 숙명여자대학교  
멀티미디어과학(학사)  
2007년-현재: 숙명여자대학교  
멀티미디어과학(석사)

관심분야 : 데이터마이닝, 디지털저작권(DRM)



**이 승 희**

2005년-현재 : 숙명여자대학교  
멀티미디어과학,  
생명과학(학사)

관심분야 : Bio-Informatics, Information Retrieval



**박 영 호**

1986년-1992년 : 동국대학교  
컴퓨터공학과(학사,석사)  
1999년-2005년 : 한국과학기술원  
전산학과(공학박사)

1993년-1999년 : 한국전자통신연구원(ETRI)  
교환전송연구단 선임연구원  
2005년-2006년 : 한국과학기술원  
첨단정보기술연구센터 연구원  
2006년-현재 : 숙명여자대학교 이과대학  
멀티미디어과학과 조교수  
관심분야 : 데이터베이스관리시스템, 정보검색, XML,  
Telecommunication System