Journal of the Korea Academia-Industrial cooperation Society
Vol. 15, No. 11 pp. 6856-6862, 2014

http://dx.doi.org/10.5762/KAIS.2014.15.11.6856
ISSN 1975-4701 / eISSN 2288-4688

사용자 웹 로그를 이용한 적응형 웹 검색

윤태복[1*], 이지형[2]
[1]서일대학교 컴퓨터소프트웨어과, [2]성균관대학교 컴퓨터공학과

Adaptive Web Search based on User Web Log

Taebok Yoon[1*], Jee-hyong Lee[2]
[1]Dept. of Computer Software, Seoil University
[2]Dept. of Computer Engineering, Sungkyunkwan University

요 약 웹 사용 마이닝은 웹 사용자의 로그 정보를 기반으로 의미 있는 패턴을 추출하는 방법이다. 하지만 기존의 웹 사용 마이닝을 이용한 패턴 추출에는 사용자들의 다양한 성향을 고려하지 않은 개별적인 모델을 생성하는데 주를 이루고 있다. 웹에서 사용된 사용자들의 검색 키워드는 그들의 검색 의도나 배경지식에 따라 다양한 의미를 가질 수 있고, 그런 개개인의 검색의도에 맞는 검색 서비스가 제공할 수 있는 기술이 요구된다. 본 논문은 사용자 검색 키워드에 대한 웹 페이지 사용 행위 정보 및 방문한 웹 페이지 리스트를 수집하고 분석하여 웹 사용자의 패턴을 추출한다. 웹 사용자 패턴은 사용자들의 검색 키워드에 대해 가질 수 있는 다양한 검색 의도에 따른 방문 웹 페이지 연결망을 생성한다. 또한, 웹 사용자 패턴은 웹 페이지 추천을 위하여 유용하게 사용할 수 있으며, 실험을 통하여 제안하는 방법의 유효함을 확인하였다.

Abstract Web usage mining is a method to extract meaningful patterns based on the web users' log data. Most existing patterns of web usage mining, however, do not consider the users' diverse inclination but create general models. Web users' keywords can have a variety of meanings regarding their tendency and background knowledge. This study evaluated the extraction web-user's pattern after collecting and analyzing the web usage information on the users' keywords of interest. Web-user's pattern can supply a web page network with various inclination information based on the users' keywords of interest. In addition, the Web-user's pattern can be used to recommend the most appropriate web pages and the suggested method of this experiment was confirmed to be useful.

Key Words : Web Mining, User Modeling, Pattern extraction

1. 서론

IT기술과 발달과 함께 웹 정보는 기하급수적으로 증가하는 모습을 보이고 있으며, 대량의 데이터로 부터 사용자는 자신이 원하는 정보를 얻기 위하여 많은 시간과 노력을 들이고 있다. 하지만, 소비하는 시간과 노력에 비해 만족할 만한 결과를 얻기는 쉽지 않으며, 이런 문제를 해결하기 위하여 패턴 분석, 마이닝 등 다양한 연구가 시도되고 있다[1]. 웹 환경에서 사용자가 원하는 정보를 보다 지능적으로 서비스하기 위해서 크게 웹 컨텐츠 및 구조를 이해하기 위한 연구와 사용자 웹 사용 정보를 분석하는 방법으로 나뉠 수 있다. 특히 사용자의 웹 사용 정보를 분석하는 연구는 웹 페이지 추천을 위한 기반 기술로 매우 유용하게 사용된다. 예를 들어 사용자들이 방문한 웹페이지를 평가하고, 그 평가 결과를 신뢰도에 반영하여 웹 검색 추천에 사용하는 방법, 또는 사용자가 관심있게 사용한 키워드나 마우스 및 키보드 등을 통한 행위 정보를 분석하여 웹 페이지를 선별하고 추천하는 방법 등이 모두 사용자에게 의미 있는 정보를 제공해 주기 위한 연구이다[2]. 하지만 기존의 웹 사용에 따른 평가 및

본 논문은 2013년 서일대학교 교내연구과제로 수행되었음.
*Corresponding Author : Taebok Yoon(Seoil Univ.)
Tel: +82-2-490-7441 email: tbyoon@seoil.ac.kr
Received August 26, 2014 Revised September 11, 2014 Accepted November 6, 2014

분석 방법은 다수 사용자의 성향을 고려한 서비스를 제공하기에는 어려운 문제가 있다. 예를 들어 “축구”라는 키워드가 있다고 가정하자. 그럼 당신은 무엇이 가장 먼저 떠오르는가? 아마도 어떤 사람은 월드컵이나 챔피언스 리그와 같은 축구 경기를 생각하는 사람도 있고, 또 어떤 사람은 팰레나 마라도나, 박지성과 같이 축구 선수를 떠 올리는 사람도 있을 것이며, 또 다른 사람은 축구공, 축구화, 유니폼 등의 축구 용품을 생각하는 사람도 있을 것이다. 이 처럼 하나의 키워드는 사용자가 가지고 있는 배경지식이나 성향 그리고 현재 처한 상황에 따라 다양한 의미를 가지게 된다. 하지만, 이렇게 사용자의 성향이 다양함에도 불구하고, 사용자 관심 키워드에 대하여 공통된 결과를 서비스 한다면 지능적이고 적응된 서비스라고 말하기 곤란할 것이다. 본 논문은 사용자의 관심 키워드 중심의 웹 검색 및 웹 사용 로그 정보를 수집하고 분석하여 웹 사용자 패턴을 추출한다. 이는 사용자 관심 키워드에 대하여 사용자가 방문 했던 의미 있는 웹 페이지들을 이용하여, 사용자들의 다양한 성향 정보를 포함할 수 있는 네트워크 형태로 표현한다.

2. 관련 연구

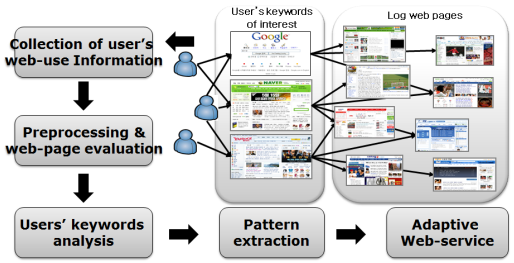
웹 사용자의 패턴 대하여 의미 있는 정보 제공을 위한 웹 페이지 추천과 관련된 연구는 다음과 같다. Joh et al.[3]과 Hay et al.[4]은 웹에서 사용자의 활동을 시퀀스로 나타내고 사용자간 유사성을 비교 분석하는 연구를 하였고, Sufyan과 Ahmad[5]은 사용자의 웹페이지 사용 정보를 분석하기 위하여 사용자의 행위 정보를 이용한 웹 페이지 평가 방법을 연구 하였다. 또한, White와 Drucker[6-9]은 단순히 하나의 웹 페이지가 아닌 여러 웹 페이지의 연관된 탐험 행위를 조사 분석하는 연구를 실시하였다. 기존 연구들의 형태는 웹 페이지 사용에 대한 로그 정보를 수집하고 분석하여 패턴을 찾고 웹 사용 정보를 모델링한다. 이 모델은 자동화, 지능적, 개인화 및 적응형 등의 서비스를 위한 기반 기술로 활용되지만, 다수 사용자의 성향이 고려되지 못한 모델 생성으로 사용 범위의 제한적인 모습을 가지고 있다. 다양한 사용자의 성향을 반영한 분석과 모델 생성에 대한 연구의 필요성이 요구된다.

국내의 상용화된 서비스로 큐로보, 다음, 네이버, 넷스루 등이 있으면 특징은 다음과 같다. 큐로보는 시멘틱 검색을 기반으로 단어의 의미를 구별할 수 있는 시멘틱 의미 검색 기술을 제공한다. 그러나 매우 큰 계산량에 따른 시스템 부하문제와 다국어 서비스가 어렵다. 다음의 카페 검색은 전문 집단 지식이 축척된 카페 정보를 이용한 검색 서비스를 제공하나 폐쇄적 카페 정보로 인해 열람 등의 어려움이 있다. 네이버의 지식검색은 사용자들의 의견들을 모아 랭킹 후, 질의에 대한 답변으로 제공하는 기술이나 답변을 평가할 수 있는 체계적인 방법이 없어 답변 지식의 유효성을 판단하기 어렵다. 넷스루 와이즈로그 프리미엄은 웹 사이트 방문 고객의 패턴을 이해하여 성공적인 서비스 전략을 세울 수 있도록 지원하는 패키지를 제공한다. 그러나 구성이 복잡하고 관리가 어려운 단점이 있다. 해외 연구 및 서비스 사례로는 구글, MS 등이 중점적으로 실시하고 있으며 그 특징은 다음과 같다. 구글의 PageRank은 구글에서 사용하는 웹 페이지 랭킹 방법으로 많이 참조될수록 중요한 페이지로 평가하는 방식이나 새로 생성된 페이지에 대해 정당한 평가가 어려움이 있다. Knol의 경우 구글에서 제공하는 커뮤니케이션 기반 지식공유 서비스이며, 다양한 주제에 대한 수많은 지식을 담는 공간으로서 지식의 저자가 저작권 보유 및 자격내용을 제공하고 리뷰와 댓글을 통해 쌍방향 커뮤니케이션을 이루어 지식을 공유하는 방식이다. 마이크로소프트 Learning User Behavior Model은 학습자 행동 모델을 정의하여 분석하고 이를 이용하여 실제 웹에서 필요한 정보를 찾는데 사용하나 모든 쿼리에 대해 동등한 성능을 나타내지는 않는다. 유럽의 경우 WAB cluster에서 UWEM을 통하여 유럽의 3개 프로젝트와 WAB cluster에 속하는 23개 단체의 협업결과를 사용하여 웹의 표준화 및 웹 사이트 접근성 평가에 관한 방법론을 제시하였다. 그러나 유럽 지역 웹에 대해서만 분석하여 사용되기 때문에 타 지역에서는 사용이 제한된다.

3. 사용자 웹로그 기반 적응형 웹 검색

웹 사용자 패턴은 사용자들이 이용한 키워드를 기반으로 웹페이지 정보를 수집하고 의미 있게 본 페이지를 분류하여 사용한다. 사용자는 구글이나 야후 등과 같은 검색 엔진을 이용하여 자신이 원하는 키워드를 입력하고 결과 페이지를 열람하게 된다. 이때, 사용자의 관심 키워

드, 열람한 웹 페이지, 웹 페이지에서의 사용자 행위 등의 정보를 수집한다. 수집된 데이터는 전처리 과정을 거쳐 유효한 웹 페이지를 분류한다. 키워드에 대하여 의미 있는 웹 페이지를 분류하고, 그 연결 형태를 연결망으로 표현한다. 생성된 연결망은 다시 성향간에 유사도를 측정하여 함축적으로 표현한다. 사용자 웹 사용정보를 수집하고, 웹사용자 패턴을 추출하는 전체 과정은 Fig. 1과 같이 나타낼 수 있다.



[Fig. 1] Work flow for the adaptive web service

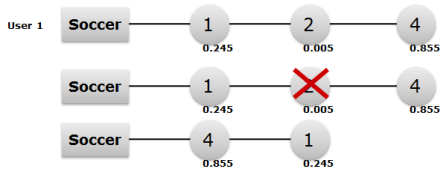
이 과정에서 다음과 같은 3가지 고려사항이 있다. 첫째, 사용자가 열람한 웹페이지 중에서 의미 있는 웹페이지 선별 방법, 둘째, 사용자의 패턴 표현 방법, 셋째, 다수 사용자의 웹 페이지 사용 정보에 대한 함축적 표현 방법이다.

웹 환경에서 사용자들은 자신이 원하는 정보를 얻기 위하여 다양한 검색 엔진(Google, Yahoo, Naver 등)을 이용하여 웹 페이지에 접근한다. 사용자가 어떤 키워드를 이용하여 검색을 하고 어떤 웹 페이지를 의미 있게 보았다면, 그 정보는 웹 검색 추천을 위한 유용한 정보로 활용 될 수 있다. 사용자 관심 키워드, 사용자 ID, 그리고 사용한 웹페이지에서의 사용자의 행위 정보는 웹페이지가 얼마나 사용자에게 의미 있게 사용하였는지를 측정할 수 있는 요소들이다. 웹페이지를 사용한 사용자의 수집할 수 있는 행위 정보로는 사용자를 구분하기 위한 사용자 ID와 관심 키워드를 이용하여 열람한 웹페이지 URL, 페이지 사용 시작 시간, 웹페이지 사용 종료 시간, 다운로드 유무, Copy & Paste 명령 (Ctrl +C, Ctrl +V) 사용 유무, 웹 페이지의 컨텐츠 크기 등 다양하다. 예를 들어 A라는 웹 페이지에서는 3분 동안 머물면서, 다양한 키보드 및 마우스 이벤트가 발생 했고, B라는 웹 페이지에서는 10초간 머물었고 아무런 키보드 및 마우스 이벤트가 발생하지 않았다고 가정하자. 이때 우리는 B라는 페이지 보다는 A라는 페이지가 사용자에게 의미 있는 웹 페이지였다고 생각 할 수 있다. 또한, 사용자의 관심 키워드에 따른 수집된 웹 페이지 사용 로그 정보를 이용한 분석은, 전처리 작업이 필요하다. 방문한 웹 페이지에서 머문 시간이 얼마 되지 않는다고 하면 사용자가 원하는 내용이 아니라고 판단할 수 있는데, 이런 경우 의미 없는 웹 페이지라고 판단하여 분석에서 제외 시켜야 한다. 또한 웹 로그 수집 과정에서 시스템 오류로 인한 잘 못된 데이터도 마찬가지이다.

웹 페이지가 사용자에게 얼마나 유용했는가에 대한 수치적 표현을 위하여 웹 페이지 점수(Web Page Scoring)[3] 방법을 이용한다. 여기에서 고려해야할 사항으로, 점수 계산에 사용되는 각 요소간의 관계가 얼마만큼 상호간에 영향을 미치는가 하는 것이다. 각 요소는 가중치 값을 이용하여 중요도를 결정한다. 예를 들어 웹 페이지 평가에 사용되는 요소가 웹 페이지 열람 시간, 마우스 클릭, 즐겨찾기 유무 3가지가 있다고 가정하자. 이때 세 가지 요소를 이용하여 웹 페이지가 얼마 유용했는가에 대한 가중치를 얻어야 한다. 동등한 의미를 부여하여 가중치를 계산할 수 도 있겠지만, 경우에 따라서는 시간이 마우스 클릭이나 즐겨찾기의 의미보다 높은 경우도 있고, 또 어떤 경우에는 웹 페이지 사용시간이나 마우스 이벤트 보다는 즐겨찾기 행위가 더 중요하다고 여겨 질 때도 있을 것이다. 적용되는 환경에 따라 각 요소의 가중치를 의미 있게 부여하여야 한다. 요소별 가중치 부여를 고려하여 웹 페이지가 사용자에게 얼마나 유용했는가를 측정하기 위하여 아래와 같은 수식을 이용하였다.

$$PageWeight_j = 1 - \left(\frac{1}{\sum_{i=0}^{n} (C_i \bullet Attribute_i)} \right)$$

$PageWeight_j$는 사용자가 어떤 키워드를 기반으로 참고한 여러 페이지들 중 j번째 웹 페이지의 가중치를 나타내며, n은 웹 페이지 평가를 위해 사용되는 요소의 개수를 의미한다. $Attribute_i$는 i번째 요소를 나타내며, C_i는 i번째 요소의 가중치상수이다. 여기서 Attribute은 웹 페이지 사용시간, 마우스 클릭, 마우스 휠 클릭, 마우스 드래그, 키보드 클릭, 복사 횟수 등 웹 페이지에서 수집된 사용자의 행위 정보를 의미 한다.

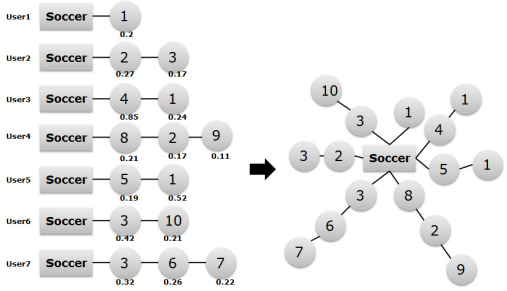


[Fig. 2] Deleting meaningless Web pages by using page weight

예를 들어 사용자 1이 "축구"란 키워드를 이용하여 웹 페이지 1,2 그리고 4을 보았다고 가정하자. 이 때 앞서 설명한 PageWeight를 이용하여 웹 페이지의 가중치를 계산하였다. 이때 페이지 2의 가중치는 0.005의 값을 가지며, 다른 페이지 보다 아주 현저하게 낮은 수치를 보이고 있기 때문에 제거 되고, 가중치가 높은 페이지 1과 4가 남겨진다.

Fig. 2와 동일한 방법을 이용하여 전처리된 사용자 7명에 대한 웹 페이지 집합은 Fig. 3의 좌측 그림과 같이 수집되었다고 가정할 때, 이는 다시 Fig. 3의 우측과 같이 통합된 연결망 형태로 표현할 수 있다.

생성된 연결망은 전처리 과정을 거쳐 의미 없는 웹페이지를 제거하였으나, 사용자의 수가 증가 할수록 연결망의 표현은 복잡하고 거대한 모습을 보이게 된다. 유사한 웹페이지를 참고한 사용자들 간의 유사성을 이용하여 적절한 통합 과정이 필요하다.

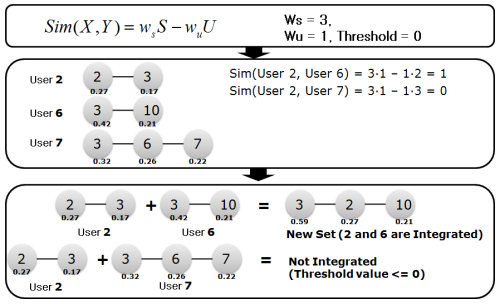


[Fig. 3] (Left) List of Web pages from many users' keywords of interest. (Right) Expression of a network from the web-page lists

만약 n명의 사용자 정보가 수집될 경우 연결망은 n개의 가지(branch)를 가지게 되어 연결망의 관리 및 탐색 연산에 드는 비용이 증가하게 된다. 관심 키워드를 기준으로 단순히 사용자가 참고한 웹 페이지의 집합을 나열하는 것을 넘어서 유사한 웹 페이지를 참고한 사용자들 간의 병합이 가능하다면 생성된 연결망을 이해하는데 더 도움이 될 것이다. Fig. 3의 경우 7명의 웹 페이지 열람 정보는 7개의 성향으로 표현된다. 연결망의 의미 있는 표현을 위해 성향간의 유사도 및 포함관계를 비교하여 통합과정을 거친다. 성향의 병합은 일치형, 포함형, 상호부분 일치형으로 크게 3가지 경우로 나뉜다. 먼저 일치형의 경우 두 성향이 동일한 경우를 나타낸다. Fig. 3에서 사용자 1과 사용자 5의 경우 웹 페이지 리스트가 동일하다. 이런 경우 한쪽을 제거 한 후 $PageWeight_j$부터 계산된 가중치를 합산한다. 포함형의 경우 사용자 1과 사용자 3에 해당한다. 사용자 1의 정보가 사용자 3의 정보를 포함하는 경우, 사용자 3의 정보를 제거하고 가중치만 합산한다. 마지막으로 상호부분 일치형일 경우 아래와 같은 수식을 이용하여 유사 정도를 판단하고 통합 유무를 결정한다.

$$Sim(X,Y) = w_s S - w_u U$$

Sim(X,Y)에서 S는 두 Set이 공통으로 포함하는 웹 페이지 개수이고, U는 두 Set이 공통으로 포함하지 않는 웹 페이지 개수이다. 또한, W_s는 두 Set이 공통으로 갖는 웹 페이지에 대한 가중치이고, W_u은 두 Set이 공통으로 갖지 않는 웹 페이지에 대한 가중치를 의미한다. 두 집합의 유사도가 임계값을 넘으면 병합하게 되고, 웹 페이지 가중치는 서로 합하여 하나의 가중치로 만든다.



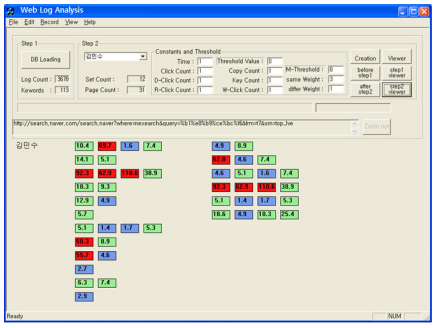
[Fig. 4] Combination through comparing similarities among users' web-page sets

웹 페이지 유사도 분석을 통한 병합 방법은 Fig. 4와 같이 중복되는 웹 페이지의 개수와 중복되지 않는 웹페이지의 개수에 가중치를 곱하여 두 집합의 유사함을 측정하였다. 그림에서와 같이 동일한 경우에 가중치를 3, 틀릴 때 가중치 1이라고 하면, 사용자 2과 6의 경우 중복

된 페이지가 1개, 중복되지 않은 페이지가 2개이므로, (3 · 1) - (1 · 2) = 1의 유사도를 얻는다 또한, 사용자 2과 7의 경우 중복되는 페이지가 1개이고, 중복되지 않는 페이지가 3이므로, (3 · 1) - (1 · 3) = 0의 유사도를 얻는다. 측정된 두 집합의 유사도는 사용자가 참고한 페이지 리스트를 통합하는 기준으로 사용된다. 만약 임계값(Threshold)을 0이라고 정의 했을 때, 유사도의 값이 0초과 이면 두 리스트를 통합하고, 0이하이면 통합하지 않는다.

4. 실험

제안하는 방법의 유효함을 확인하기 웹 환경에서 사용자의 웹 검색 키워드 및 사용 정보를 수집하고 분석하였다. 사용자들의 로그 데이터 수집을 위해 IE(Internet Explorer)에 Add-on 할 수 있는 프로그램을 제작했다. 또한, 수집된 사용자 웹 로그를 분석하여 그 결과를 확인하기 위하여 Fig. 5과 같은 툴을 제작하였다.



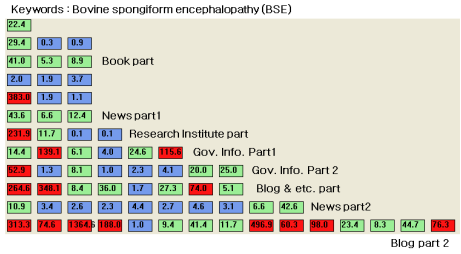
[Fig. 5] Web log Analysis tools.

사용자의 웹 사용 정보를 수집하기 위해서 현재 사용자의 IP address, IE 버전, 다중 탭 구분을 위한 ID, 열람한 URL, 이전 URL, 웹 페이지 Title, 파일 크기, 문자 길이, 시작시간, 종료시간, 마우스 클릭 횟수, 마우스 더블클릭 횟수, 오른쪽 마우스 클릭 횟수, 마우스 휠 횟수, 드래그 인 드랍 횟수, 키보드 클릭 횟수, 복사 & 붙이기 횟수, 문서 종류, 웹 페이지 프레임별 구분 URL, 클릭된 URL으로 구분되어 수집한다.

사용자 웹 로그는 S대학교 교내에 자율 PC실 150대에 사용자 웹 사용 정보를 수집할 수 있는 프로그램을 설치하였다. 이 프로그램은 IE (Internet Explorer)를 이용하여 네이버, 야후, 다음, 구글 등의 검색 사이트에서 키워드 입력 후의 웹 페이지 사용 정보에 대하여 수집한다. 전체 수집된 로그 파일 용량은 약 950MByte 이다. 우리는 수집된 웹 사용 로그 정보에서 27,965개의 검색 키워드를 찾았고 87,738개의 키워드와 연관된 웹 페이지 사용 정보를 얻을 수 있었다. 27,965개의 키워드 중에서 약 90%이상이 한 번씩만 검색되었던 키워드이다. 의미 있는 분석을 위해서는 한 개의 키워드에 대하여 다양하게 검색 했던 데이터를 이용하는 것이 유용하게 사용될 수 있다. 그래서 하나의 키워드에 최소 5회 이상의 검색 정보가 존재하는 경우에만 분석에 이용하였다. 전체 수집정보에서 5회 이상의 검색 정보가 있는 키워드는 395개 이었다.

대부분의 검색 키워드가 특정 사이트를 찾기 위한 것으로 검색 후 한 번의 클릭으로 웹사이트 이동이 가능한 것이다. 앞서 설명한 바와 같이 어떤 키워드 대해서 모든 사용자가 원하는 절대적인 한 개의 사이트가 존재한다면, 웹 사용자 패턴 생성에 의미가 없는 데이터이다. 그래서 본 실험에서는 국회도서관이나 싸이월드와 같은 특정 기관이나 웹 사이트를 검색하기 위한 키워드를 제외하고 광우병, 전기 분해 등과 같은 사용자에 따라 원하는 결과가 다양한 키워드를 직접 선별하였다. 분류된 키워드는 113개이며, 이 키워드를 통하여 검색된 횟수는 1,614회, 전체 열람한 웹 페이지의 개수는 3,678개 이다.

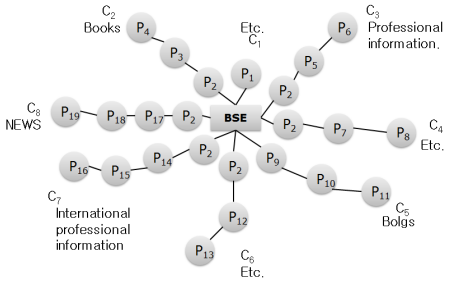
113개의 키워드에는 영화, 지도, 자기소개서, 여성과학자, 대운하, 아스피린, 인버터, 광우병, 과외, 광섬유 등 다양하게 나타났으며, 이중에서 '광우병'의 경우 Fig. 6과 같은 결과를 보였다.



[Fig. 6] Result of web log Analysis tools.

Fig. 7에서와 같이, 광우병 키워드는 도서, 뉴스기사, 전문 정보, 해외전문정보, 블로그, 기타 분류로 각각의 성향으로 나눠진다. '기타'를 제외한 각 성향은 의미를 가지

고 뭉쳐진 모습을 보이고 있다. C1, C4, C6의 경우 검색 정보, 블로그, 카페 등 혼합된 웹 페이지 정보를 가지고 있다. C2의 경우 처음에는 검색 페이지 이지만 그 이후 광우병과 관련된 도서 정보 웹 페이지를 참고하였다. C3는 보다 전문적인 정보를 찾는 모습을 볼 수 있었다. C5는 키워드와 관련된 블로그를 방문하였다. C7는 광우병의 최초 발병지인 영국의 정부 기관(환경식품농촌부) 사이트에서 정보를 열람하였다. C8은 조선일보 및 YTN에서 키워드와 관련된 뉴스 기사를 열람한 것을 알 수 있었다. 이러한 웹 사용자 패턴은 Fig. 8와 같이 브라우저에 검색을 지원하기 위한 서비스 기반 기술로 활용한다.



[Fig. 7] Pattern for adaptive web service

[Fig. 8] Browser for adaptive web service

5. 결론 및 향후 연구

본 논문은 사용자의 웹 검색 키워드에 대한 다양한 성향 정보를 포함할 수 있는 웹사용자 패턴추출 방법을 제안하였다. 사용자의 키워드 기반의 웹 사용정보를 기반으로 웹 페이지 연결망을 생성하고, 사용자 성향 간에 유사도를 측정하고 통합하여 보다 의미 있는 연결망을 생성하였다. 추출된 웹 사용자 패턴은 웹 페이지 추천서비스에 가능하고, 키워드 중심의 연결망간의 비교 및 분석을 통하여 의미 유사성을 판단하는 기반 기술로 활용가능하다. 또한 실험에서는 임의 사용자에 대하여 자유로운 키워드 검색을 통한 웹 로그 정보를 수집 분석하는 실험을 하였다. 실험에서 추출된 웹 사용자 패턴은 사용자 검색 행위에 대한 정보를 잘 나타내고 있었으며, 추천 서비스를 위해 유용하게 적용 가능할 것이다. 향후 연구로는 상용화된 웹 브라우저를 이용한 다양한 웹서비스 방법과 보다 많은 사용자를 대상으로 하는 실험이 요구된다.

References

[1] Robert Cooley, Bamshad Mobasher, Jaideep Srivastava, "Web mining: information and pattern discovery on the World Wide Web", *Ninth IEEE International Conference on Tools with Artificial Intelligence*, pp.558~567, 1997.
DOI: http://dx.doi.org/10.1109/TAI.1997.632303

[2] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan, "Web usage mining: discovery and applications of usage patterns from Web data", *ACM SIGKDD Explorations Newsletter*, Vol. 1, No. 2, pp.12~23, 2000.
DOI: http://dx.doi.org/10.1145/846183.846188

[3] Chang H. Joh, Theo A. Arentze, Harry J. P. Timmermans, "A position-sensitive sequence alignment method illustrated for space-time activity-diary data," *Environment and Planning A 2001*, vol. 33, pages 313~338, 2001.
DOI: http://dx.doi.org/10.1068/a3323

[4] Birgit Hay, Geert Wets, Koen Vanhoof, "Clustering navigation patterns on a website using a Sequence Alignment Method," *Proc. Intelligent Techniques for Web Personalization: 17th Int. Joint Conf. Artificial Intelligence*, 2000.

[5] M. M. Sufyan Beg, Nesar Ahmad, "Web search enhancement by mining user actions," *Information Sciences,* vol. 177, pp.5203~5218, 2007.
DOI: http://dx.doi.org/10.1016/j.ins.2006.06.011

[6] Ryen W. White, Steven M. Drucker, "Investigating Behavioral Variability in Web Search," *The International World Wide Web Conference* 2007.

[7] C.-K. Park, H.-S. Park, Y.-S. Hong, "Traffic Safety System based on WEB", The Journal of The Institute of Internet, Broadcasting and Communication (IIBC), Vol. 14,

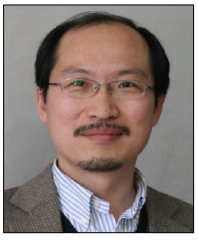
No. 3, pp.81-88, Jun. 2014.
[8] K. Kim, D. Nam, "Web Service for Traffic Information Using Focus+Context Visualization Technique", The Journal of The Institute of Internet, Broadcasting and Communication (IIBC), Vol. 14, No. 2, pp.101-106, Apr. 2014.
[9] M.-J. Lim, M.-G. Kim, K.-Y. Lee, "WPAN Based Semantic-Web Health Monitoring", The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 13, No. 6, Dec. 2013.

윤 태 복(Taebok Yoon) **[종신회원]**

- 2001년 8월 : 공주대학교 전자계산학과 (이학사)
- 2005년 2월 : 성균관대학교 컴퓨터공학과 (공학석사)
- 2010년 8월 : 성균관대학교 컴퓨터공학과 (공학박사)
- 2011년 3월 ~ 현재 : 서일대학교 컴퓨터소프트웨어과 조교수

<관심분야>
데이터 마이닝, 사용자 모델링, 게임인공지능

이 지 형(Jee-hyong Lee) **[정회원]**



- 1993년 2월 : 한국과학기술원 전산학과 (학사)
- 1995년 2월 : 한국과학기술원 전산학과 (석사)
- 1999년 2월 : 한국과학기술원 전산학과 (박사)
- 2002년 3월 ~ 현재 : 성균관대학교 컴퓨터공학과 교수

<관심분야>
퍼지이론, 지능시스템, 기계학습