

효과적인 콘텐츠 서비스를 위한 RSS피드 검색 엔진의 설계 및 구현*

이 해 성** · 권 준 희***

Design and Implementation of RSS feed search engine for Effective Contents Service

Lee, Hae sung · Kwon, Joon Hee

〈Abstract〉

In the Web 2.0, besides more gaining information on the web, the number of web sites that take advantage of RSS increases explosively. Commonly each users search RSS channels through the web search engine before registering RSS channel's url to the RSS reader. Users judge whether the site is RSS channel or not and register an RSS channel's url through their interests. Because accomplished by users themselves, those processes conflict to user's convenience and quick consumption of information. Techniques of current search engines can't provide users with reliable RSS feed information as search results. In this paper, we analysis appropriateness of current search engines' techniques that offer users RSS feed search service and discuss their limitations. Also, we make up RSS feed database through classification of RSS tag being possible to search RSS feed information effectively and apply update rate of each RSS channel's feed to ranking algorithm providing more reliable search results.

Key Words: RSS, Web2.0, Search Engine

I. 서론

웹은 폭발적으로 증가하는 정보의 양과 더불어 빠른 속도로 확장되어 가고 있다. 웹2.0시대의 도래로 UCC(User Created Contents) 형태의 정보가 나날이 생성되고 신속한 정보의 획득과 정보공유의 중요성이 증대되었다. 웹상에

산재되어 있는 수많은 정보원천(information resource)에서 개인의 관심사와 결부되는 정보를 획득하고 효과적으로 소비하는 것은 정보의 가치를 높이는 것이며 개방(공유)/참여(협력)를 통한 인터넷 서비스의 혁신을 이루는 것이다. 이는 웹2.0이 추구하는 핵심 요소이며 앞으로 웹이 발전해 나가야할 방향이다[1].

웹이 확장 되어가고 정보 가치에 대한 중요성이 증대됨에 따라 사회적으로 웹 검색 엔진에 대한 의존성은 커지고 있다. 세계 최대 검색엔진인 구글(Google)은 링크

* 본 논문은 경기도 지역협력연구센터인 경기대학교 콘텐츠융합소프트웨어연구센터의 지원으로 수행한 연구결과임.

** 경기대학교 일반대학원 컴퓨터과학과

*** 경기대학교 컴퓨터과학과(교신저자)

구조(Link Structure)와 링크 텍스트(Link Text)를 이용하여 관련성 있는 정보를 검색해내고 질 높은 정보 필터링(Filtering)능력을 보유하고 있다.

눈부신 기술 발전에도 불구하고 웹의 확장은 정보의 생성, 유통, 사용에 있어서 좀 더 새롭고 획기적인 기술적 접근 방식을 요구한다. 1994년 10,000개의 웹페이지 인덱스를 보유한 최초의 검색 엔진인 월드 와이드 웹 웜(World-Wide Web Worm)의 등장 이후 사용자에게 더욱 만족할 만한 결과를 제시하기 위한 검색 엔진의 정보 획득과 추천 메커니즘은 수많은 변화를 거듭했다[2]. 또한 웹상의 산재되어 있는 수많은 정보 원천으로부터 동시다발적으로 생성되는 정보를 거의 실시간으로 사용자에게 배달하는 RSS서비스는 차세대 정보 배달기술로서 각광 받고 있다.

좀 더 빠른 정보 획득과 배포, 이용에 있어서 사용자 만족의 극대화는 웹2.0시대의 검색엔진과 RSS기술이 추구하는 것이다. 그러나 웹2.0시대의 선도적 역할을 하는 검색엔진은 RSS서비스 제공 사이트의 웹 페이지를 획득하고 이를 필터링하여 RSS리더기(RSS Reader, RSS Aggregator)를 이용하려는 사용자에게 등록할만한 RSS 채널을 검색 결과로 제공 하지 못할 뿐 아니라 사용자가 흥미를 가질만한 정보를 효과적으로 제시하지 못한다. 왜냐하면 RSS가 갖는 기술적 특징은 검색엔진의 검색 알고리즘 적용에 있어서 적합하지 않기 때문이다. 이를 위해서는 빠르게 증가하는 RSS서비스 사이트들의 채널 주소를 자동적으로 획득하고 데이터베이스화 하여 사용자에게 RSS에 적합한 질의를 통한 검색 결과를 제시하는 것이 바람직하다. 따라서 RSS 피드 검색에 특수화된 검색시스템을 필요로 한다.

본 논문에서는 기존 검색엔진의 기술적 특징과 연구 현황을 살펴보고, RSS서비스와 관련한 검색엔진의 제한점을 분석한다.

이를 토대로 RSS특성을 고려한 효과적인 RSS피드 검색시스템을 제안한다. 제안된 시스템에서는 RSS태그별 분류를 통하여 데이터베이스를 구성한다. 이는 효과적인

RSS피드 검색이 가능 하도록 한다. 또한 각 RSS채널의 아이템 갱신률을 이용한 우선순위 적용 방식을 고려하여 좀 더 신뢰할 수 있는 검색 결과를 상위에 제시하는 RSS 피드 검색 시스템을 설계 하고 구현한다.

II. 관련 연구

2.1 RSS(Really Simple Syndication)

블로그는 web과 log를 합친 weblog의 약자로 웹상에서 개인 또는 그룹의 의견이 기록되고 공개되며 공유되는 웹2.0이 추구하는 마와 일치하는 대표적 사례이다.

초기에는 HTML을 수작업으로 편집하는 수동적인 기록 환경이었지만, 이후 자동화된 편집과 저장, 출판 기능이 개발되어 추가되면서 종합적인 개인 기록 플랫폼이자 개인미디어 플랫폼으로 발전하였다. 여기에 트랙백(trackable), 핑백(pingback), 퍼마링크(permalink)등의 연결 기능과 함께 콘텐츠 신디케이션(syndication) 기술인 RSS서비스가 결합되면서 블로그를 통해 생산된 콘텐츠를 보다 손쉽게 유통시킬 수 있게 되었고, 이를 통해 콘텐츠 생산 및 재생산과 유통기술이 적용된 핵심 플랫폼으로 진화하게 되었다[3]. 나날이 폭발적으로 늘어나는 다양한 서비스들과 정보를 처리하는데 대한 어려움이 생기기 시작하면서 사용자가 정보를 직접 찾아다니며 브라우징 하는 환경이 아닌, XML로 만들어진 RSS/Atom등의 정형화된 데이터를 자동으로 구독하여 다량의 정보를 습득하는 환경으로 변화되고 있다.

RSS기술은 데이터의 공유를 위한 포맷(format)이다. RSS 피드라 불리는 XML문서 형태로 정보가 가공되어 RSS리더기에 의해 사용자가 구독할 수 있는 형태로 정보가 제공된다. RSS 리더기는 사용자가 선택한 채널에 접근하여 가지고 온 RSS 피드를 파싱하여 RSS 규격에 나와 있는 태그별로 콘텐츠를 추출한다[4]. 일반적으로 RSS 리더기는 사용자가 등록한 채널에서 획득된 정보가 새로

운 것인가를 판단하여 구독자에게 배달시킬 것인지 아닌지를 결정하게 되는데 규격에서 제시된 다음과 같은 태그 정보를 이용하여 새로운 정보의 여부를 판단하게 된다[5].

- 1) RSS 2.0규격이 명시한 태그 중 포스트의 고유값을 나타내는 <guid>를 이용
- 2) 포스트의 원문 출처를 나타내는 <link>태그 이용

RSS리더기는 사용자에 의해 선택된 정보의 원천으로부터 정보를 수집하는데 정보의 공유가 효과적으로 이루어 질 수 있도록 정보 제공자와 구독자 사이의 중재자역할을 한다.

RSS 기술적용과 관련된 장점들을 요약하자면 다음과 같다.

- 1) 선택적 구독 -사용자가 직접 원하는 주제와 일치하는 채널 선택.
- 2) 빠른 구독 - 거의 동시에 다양한 RSS사이트 접근.
- 3) 히스토리(history) 관리 - 다양한 채널의 과거 기록들의 보관 가능.
- 4) 자동화된 콘텐츠 연동이 용이.
- 5) 콘텐츠 재사용성 - 구조화된 XML데이터로 손쉬운 변환 및 처리가 가능.
- 6) 커뮤니케이션의 변화 - 1:1에서 1:N으로 변화.

RSS기술적용의 궁극적 목적은 전달하고자 하는 정보의 많고 적음에 집중하는 것이 아니라, 사용자로 하여금 선택된 채널에서 개인 관심사에 부합하는 정보를 생성과 거의 동시에 지속적으로 제공하는 것에 있다.

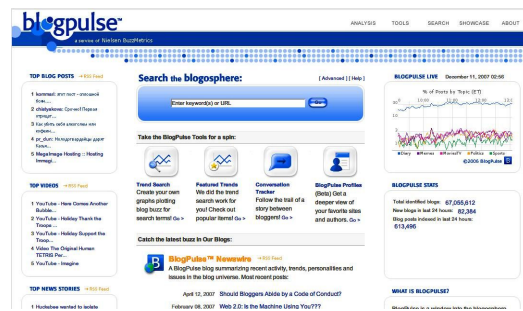
2.2 국내외 RSS피드 검색 엔진 기술 및 동향

현재 보편적으로 사용되고 있는 일반적인 RSS리더기

를 이용하기에 앞서 사용자는 검색엔진이나 블로그 포털 사이트의 검색도구를 통해 관심 주제를 제공하는 사이트의 링크정보를 획득한다. 결과로써 제시된 사이트들을 일일이 방문하면서 해당 사이트가 자신의 관심여부에 부합하는지 사용자 스스로 판단을 하고 RSS서비스를 제공하는지 여부에 따라 RSS리더기에 RSS채널 주소를 등록하게 된다. 이러한 일련의 과정은 사용자 입장에서 불편함을 야기 하며 RSS리더기를 통해 배포되는 정보에 대한 신뢰성이 떨어진다. 또한 아직 까지 사용자들에게는 RSS 문서구분과 링크위치 파악에 있어 어려움이 존재한다. 또한 사용자가 스스로 획득한 RSS채널 주소를 통해서 신뢰성 있고 유효한 정보를 계속적으로 제공하는지도 알 수 없다.

국외에서는 블로그와 관련한 여러 유용한 서비스를 제공하는 <그림 1>에서 보이는BlogPulse가 다른 어떠한 피드 검색 서비스보다 많은 피드기반콘텐츠의 인덱스를 보유하고 있다고 알려져 있다. BlogPulse는 텍스트매칭 검색뿐만 아니라 사용자에게 다양한 질의 방식을 제공한다.

BlogPulse 이외에도 다양한 검색 서비스를 제공하는 RSS피드 검색엔진들이 존재하고 발전을 거듭하고 있지만 이들 대부분은 사용자가 원하는 정보를 적절히 필터링(filtering)하거나 스팸성 RSS피드문제를 해결하지 못하므로 써 신뢰할 수 있고 만족할 만한 정보를 사용자에게 제시하지 못하고 있는 실정이다.



<그림1> 국외 RSS피드 검색엔진 - BlogPulse

국내에서도 초기에는 해당 포털 사이트 내에 블로그 끼리 정보교환을 위한 포맷으로 활용하는데 그쳤으나, 2004년 이후 뉴스 분야 사이트를 중심으로 RSS기술을 적용한 정보 배포 서비스를 시작하여 다방면으로 RSS기술을 이용하여 사용자에게 더 나은 서비스를 제공하고 있다[6]. 최근에는 블로그끼리 정보를 교환하고 연결하기 위한 포맷으로 RSS를 활용하는 것은 물론, RSS피드들을 서비스 콘텐츠로 규정하고 <그림 2>에서 보이는 국내의 RSS피드 검색엔진인 '올블로그'처럼 RSS 피드의 검색 및 신규 피드 정보를 제공해주는 일종의 RSS포털 성격의 사이트들이 늘어나고 있다. 그러나 전용 RSS피드 검색엔진의 수가 많지 않고 블로그 포털사이트의 경우 해당 포털의 RSS채널로만 검색이 제한되며 채널의 타이틀, 소개글 정도의 기준으로 사용자가 입력하는 질의를 처리하여 사용자가 만족할 만한 결과를 제시하지 못한다. 또한 글이 올라오지 않는 비어있는 블로그도 다수 검색되므로 효율적이지 못하다.



<그림 2> 국내 RSS피드 검색 엔진 - 올블로그

방대한 양의 정보를 효과적으로 배포하기 위한 RSS서비스는 사용자 관심을 고려한 정보 원천에서 낱말이 생성되는 최신의 정보를 꾸준히 제공할 필요성을 가진다. 기존의 블로그 포털 사이트에서 제공하는 RSS채널 검색 도구는 블로그의 제목과 소개 글을 기준으로 단순히 키워드 매칭을 통해 이루어지므로 해당 RSS채널의 정보 갱

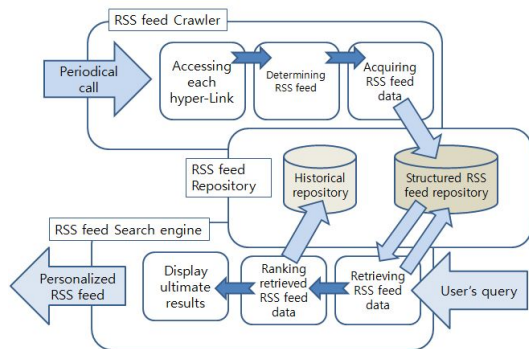
신 여부와는 무관한 결과가 제시된다. 이는 RSS기술의 사용을 무의미 하게 만드는 것이며 정보 제공자와 소비자 모두에게 있어서 정보의 가치를 떨어뜨리는 결과를 초래한다. RSS채널 정보를 검색 결과로 제시하기 위한 RSS피드 전용 검색엔진이 등장 하였지만 이들 대부분 현재 검색 엔진의 랭킹 방식을 그대로 RSS피드 검색 시스템에 적용하거나, RSS특성을 살리지 못한 검색 서비스를 제공하고 있다.

기존 검색엔진들의 유사도에 따른 순위부여 방식은 사이트 단위가 아닌 문서 단위의 유사도 계산에 적합하므로 갱신 주기를 고려한 RSS채널 순위부여 방식과는 거리가 멀다[6]. 예를 들어 사용자가 입력한 단어를 다수 개 포함한 글이 1개 올라온 RSS채널과 사용자가 질의한 단어를 한 개 이상 포함한 다수개의 글이 올라온 RSS채널의 경우를 볼 때, 단순한 유사도 계산으로 순위를 부여하면 전자의 경우가 우선순위를 높게 가질 수 있다. 즉 더 이상 새로운 RSS피드 정보가 갱신되지 않는 RSS채널이 검색 결과의 상위로 제시되는 문제가 발생하게 된다. 하이퍼링크 정보를 이용한 HITS나 PageRank와 같은 순위부여 방식 또한 블로그 형태의 개인 미디어 웹페이지의 순위를 부여할 때에는 적합하지 않다[7]. 블로그에 존재하는 타 블로그와의 하이퍼링크는 정보의 신뢰성과 유효성에 기반을 두기보다 단순히 싸이월드의 일촌과 같은 인간적인 친분관계를 반영하여 링크를 걸어두는 경우가 많다. 또한 링크정보는 유동적으로 변하며 링크의 개수를 제한한 블로그 포털 사이트도 있다.

앞으로 RSS서비스 적용 웹 페이지의 수가 계속적으로 늘어나고 차세대 정보 배포 방식으로 RSS기술의 다양한 응용과 도입 사례가 증가될 것이다. 따라서 RSS의 기술적 특성을 반영하여 RSS채널 정보 데이터베이스를 구성하고 해당 채널에 대한 정보 갱신을 고려해 채널 순위를 결정하는 좀 더 효과적인 검색 시스템의 도입이 요구 된다.

III. 자동 탐색 RSS피드 웹 수집기와 RSS피드 검색 시스템

본 논문에서는 사용자의 편의성을 도모하고 다양한 질의를 통해 정보 획득에 대한 사용자 만족도를 극대화하는 RSS 자동 탐색 수집기와 RSS 태그 분류를 통한 RSS피드 데이터베이스로 구성된 새로운 RSS피드 검색 시스템 모델을 제안하고 구현한다. 제안된 모델은 웹상에서의 RSS채널정보의 수집이 사용자와의 상호작용 없이 자동화되며 RSS 태그를 이용해 글의 작성자, 작성시간 등을 추출하여 저장함으로써, RSS피드 검색 시 RSS 특성을 활용한 다양한 질의가 가능하다. 또한 RSS 채널의 피드 갱신률을 랭킹 알고리즘의 주요 요소로써 고려하여 스팸성 RSS 피드를 효과적으로 필터링한다. 따라서 사용자에게 보다 신뢰할 수 있는 정보를 검색 결과로써 제시한다.

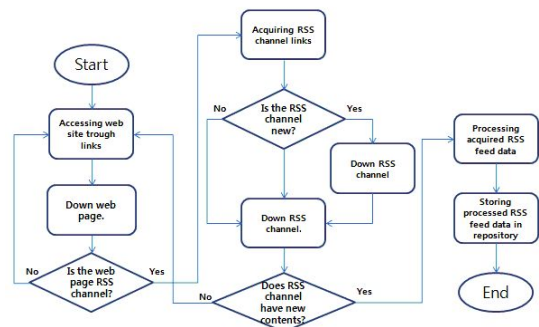


<그림 3> 시스템 수행 모듈 구성도

<그림 3>에 나와 있는 바와 같이 본 논문에서 제안하는 시스템은 RSS서비스 사이트정보를 획득하기 위한 자동 탐색 수집기모듈, RSS문서를 분석하고 데이터를 추출하여 구성된 데이터베이스와 이를 바탕으로 사용자의 질의에 따른 적합한 RSS피드를 검색 결과로 제시하는 검색 모듈로 구성된다.

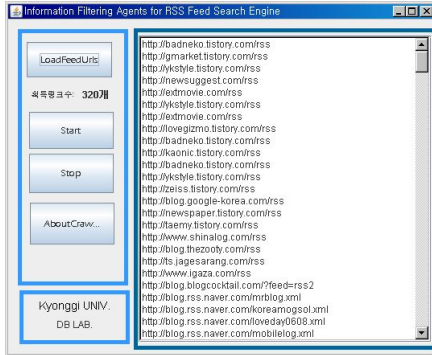
3.1 자동 탐색 RSS 피드 웹 수집기

웹 문서 내부에 포함된 하이퍼링크를 따라 이동하며 웹상의 정보를 수집하는 일종의 소프트웨어인 웹 수집기는 기본적으로 깊이우선 탐색(Depth First Searching), 너비우선탐색(Breath First Searching)방식으로 웹의 모든 문서 대상으로 탐색을 하게 된다[4]. 웹은 주로 HTML방식으로 구성되어 있기 때문에 일반적인 검색 엔진에서는 이에 걸맞는 검색이 가능하도록 웹 수집기를 설계한다. 하지만 RSS서비스제공 사이트의 경우 콘텐츠의 내용을 손쉽게 읽을 수 있는 RSS파일 규약을 지원하고 있으므로 이에 걸맞은 RSS피드 자동 수집기를 구현하는 것이 바람직하다.



<그림 4> RSS피드 수집기 수행 흐름도

<그림 4>는 본 논문에서 제안한 RSS피드 수집기의 수행 흐름도를 보여주고 있다. RSS피드 수집기는 주기적인 사이트 접근 호출 모듈을 통해 RSS서비스 사이트의 URL주소를 획득한다. 블로그와 같은 개인 미디어 사이트인 경우 RSS채널 주소 링크는 보통 블로그에 메인 페이지에 존재하기 때문에 탐색 범위를 제한하고, 해당 사이트의 링크 개수로 탐색 큐의 크기로 하여 RSS피드의 링크인지 조사한다. 수집기는 획득한 링크정보에 해당하는 URL로 접근하여 RSS채널 여부를 판단하고, 이미 획득된 채널인지의 여부를 판단하여 신규 채널일 경우 해당 RSS채널 주소를 저장하게 된다.



〈그림 5〉 RSS 피드 자동 수집기

〈그림 5〉는 본 논문에서 구현한 3개의 RSS 피드 자동 수집기들이 웹상을 돌아다니며 수집한 RSS피드의 URL 들을 보여주고 있다.

3.2 RSS 피드 검색 시스템

사용자가 입력한 질의에 따라 신뢰할 수 있는 RSS 채널정보를 검색결과로 제시하기 위해서는 RSS의 기술적 특성을 고려한 우선순위 부여 방식이 필요하다.

본 논문에서는 제안한 RSS피드 검색엔진에서는 사용자에게 좀 더 신뢰할 수 있는 RSS채널이란 RSS피드의 갱신주기가 짧으면서 보다 많이, 꾸준히 올라오는 채널이라 정의한다. 따라서 사용자의 질의 단어를 포함하는 게시물의 수, 자료의 갱신률, 자료의 갱신주기 등을 고려하여 우선순위를 부여한다. RSS채널의 순위를 정하기 위한 첫 번째 식은 (1)과 같다.

$$t p = \sum_{i=1}^{i=m} (k/n) * \alpha \quad 0.5 \leq \alpha \leq 1 \quad (1)$$

tp: 질의단어의 위치와 게시물수에 따른 우선순위 값
k: 질의 단어 수
n: 일정기간동안 질의 단어를 포함한 RSS피드의 수

식(1)은 일정기간동안에 질의 단어를 포함한 RSS피드의 양을 판단하여 순위를 부여하게 된다. 첫 번째 계산에

의해 필터링(filtering)된 검색결과 중에서 다시 식(2)를 적용하게 된다. 식(2)은 RSS피드 갱신의 일관성여부를 파악하기 위해 갱신률을 계산하는 식이다.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (k_i - m)^2 \quad (2)$$

n: 일정주기
k_i: i번째 주기의 질의 단어 포함 RSS피드 개수
m: 일정주기동안 발행한 RSS피드 평균 개수

일정기간동안 두 RSS채널의 질의 단어를 포함한 RSS 피드의 수가 같아도, 각 RSS피드들의 갱신주기가 다르기 때문에 두 RSS채널은 서로 다른 갱신 주기 분산값을 가질 수 있다. 따라서 보다 짧은 갱신 주기 값을 갖는 RSS 채널이 더 높은 순위를 가지게 되는 것이다.

〈그림 6〉은 본 논문에서 제시한 RSS서비스 사이트 검색 추천 시스템에서 사용되는 RSS 채널 갱신률을 계산하는 유사코드를 보여준다. 일정기간동안에 i번째 주기에 사용자 질의 단어를 포함하는 RSS피드가 존재하면 해당 RSS피드의 pubDate태그 값을 저장하고 이를 통해 일정주기 동안 해당채널의 RSS피드 갱신율을 계산할 수 있게 된다.

```

m ← the number of published contents within sample period
total ← 0
templ[sample period]

for i ← 1 to sample period do
  if ( existing published contents at i )
    d[i] ← the publishing date of contents at i
  else
    d[i] ← 0

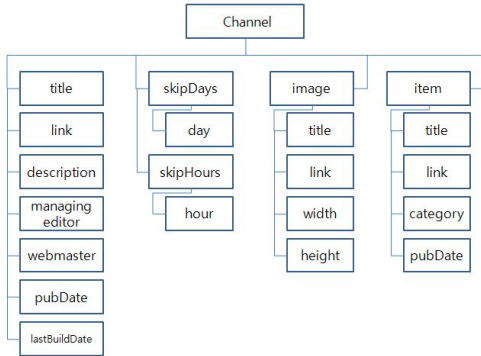
for j ← 1 to sample period do
  templ[j] ← d[j+ 1] - d[j]
  total ← total + templ[j]

return total / m - 1
    
```

〈그림 6〉 RSS채널 갱신률 계산 유사코드

이를 통해서 결과적으로 사용자의 관심 정보를 많이 포함하고 새로이 갱신되는 정보가 꾸준히 올라오는 RSS 서비스 사이트가 상위로 추천하게 됨으로써 기존 검색방식보다 신뢰할 만한 결과를 제시할 수 있다.

저장 공간의 효율성, 빠른 우선순위 계산과 질의 결과의 도출을 위해 웹 수집기가 획득 한 채널에 대한 정보는 RSS피드 단위로 저장하는 것이 아닌 <그림 7>에 나와 있는 아이템을 단위로 하여 <item>태그의 하위 태그 별로 구조화되어 저장된다.

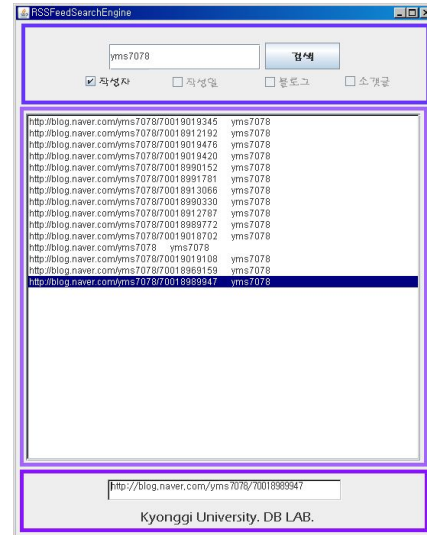


<그림 7> RSS 태그 트리

이는 블로그와 같은 RSS서비스 사이트가 담고 있는 콘텐츠에 대한 정보를 아이템 태그의 하위 태그들이 담고 있기 때문이다. RSS문서의 태그 분류를 통한 데이터 베이스 구성은 사용자로 하여금 RSS채널 검색 시 다양한 질의를 가능토록 한다. 예를 들어 작성자의 이름에 해당하는 문자열로 질의를 하면 같은 작성자의 다른 게시물도 함께 검색 결과로 제시될 수 있다.

<그림 8>은 RSS특성을 활용함으로써 일반적인 텍스트매칭 방법뿐만 아니라 좀 더 다양한 질의 방법을 제공하는 것을 보여주고 있다.

RSS서비스 사이트 검색 추천시스템 모델에서 제안된 우선순위 부여 방식은 사용자가 입력하는 질의어뿐만 아니라 해당 채널의 정보 갱신을 또한 고려하였기 때문에 RSS기술의 기본 목적인 사용자 맞춤 정보의 효과적인 자동 전달과 부합한다. 이는 정보의 효율적인 획득과 배포를 통한 공유, 재사용의 관점에 있어서 스팸성 피드의 필터링 문제와 같은 기존 RSS피드 검색도구가 가지고 있던 제한 사항을 극복한 신뢰할 수 있는 해결방안을 제시한



<그림 8> 작성자에 따른 RSS피드 검색

다. 결과적으로 기존의 블로그 검색도구나 RSS피드전용 검색엔진에 비해 좀 더 신뢰할 수 있고 사용자가 원하는 정보를 검색 결과로써 제시할 수 있다.

IV. 결론

웹2.0시대의 도래와 웹의 확장으로 인하여 나날이 방대해 지고 있는 정보의 바다에서 효과적인 정보 소비를 위한 새로운 기술들의 필요성이 대두 되었다. 이러한 기술들 중 차세대 인터넷 정보 배달 방식으로 여겨지는 RSS기술은 다양하게 선택된 정보원천으로부터 최신의 정보를 수집하여 동시에 특정 구독자에게 신속하게 제공한다. 그러나 주요 검색엔진과 블로그 포털 사이트들에서 제공하는 검색도구를 통해 획득된 RSS서비스 사이트의 URL정보는 RSS에서 고려되는 태그별 콘텐츠와 정보 갱신을 무시하였기 때문에 신뢰할 수 없는 RSS서비스 사이트가 제시될 가능성이 많다. 국내외적으로 RSS피드를 검색하기 위한 RSS전용 검색엔진이 등장은 하였지만 기술적인 면에서 나날이 증가하는 RSS서비스 적용 범위

를 포괄하지 못하고 있다. 때문에 RSS의 기술적 특징을 고려한 빠른 웹 수집기와 획득된 정보에 대한 우선순위 부여 방식이 적용되어야 한다. 웹상의 산재되어있는 수많은 RSS서비스 사이트에서 발생하는 정보를 신속하게 획득하고 이를 사용자가 효율적으로 소비하는 이상적인 정보 순환구조를 구축하기 위해서 정보제공자와 정보 소비자 사이의 중재자 역할을 하는 RSS 기술의 특성을 고려한 RSS피드 검색엔진의 개발이 요구된다.

본 논문에서는 빠르게 증가하는 RSS서비스사이트를 신속하게 획득하여 RSS문서의 특성을 고려한 데이터베이스를 구성하는 RSS자동 수집기와 사용자로부터 입력받은 질의 단어이외에도 해당 채널의 정보 갱신률을 고려한 우선순위를 부여하여 좀 더 신뢰할 만한 RSS피드를 상위에 제시하는 시스템을 제안 하였다. 또한 저장된 데이터의 질의를 통해 정보에 대한 효율적인 재사용과 효과적인 소비를 극대화하는 시스템을 설계하고 구현하였다.

참고문헌

- [1] Tim Berners-Lee, "Putting the Web back into Semantic Web", ISWC2005 Keynote, 11 .2005
http://www.w3.org/2005/Talks/1110-iswc-tbl
- [2] Sergey Brin and Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", the Seventh International World Wide Web Conference Vol. 30, Issues 1-7, April 1998, pp. 107-117
- [3] 이제윤, "질문 유형에 따른 인터넷 검색엔진의 성능 비교", 한국정보관리학회 제10회 학술대회 논문집, pp. 185-192, 2003
- [4] Ben Hammersley, "Content Syndication with RSS", O'Reilly & Associates, Vol.195, pp.117-124, Pub 2003.
- [5] 전중홍, 이승윤 "웹 2.0 기술 현황 및 전망", 전자통

신동향분석 제 21권 제 5호 pp. 141-153, 2006

- [6] Kowalski, G. and Maybury, M. "Information Storage and Retrieval Systems", Kluwer ACADEMIC PUBLICHER, Vol.313, pp. 10-24 Pub 2000
- [7] B. Yuwono "Search and ranking algorithms for locating resource on World Wide Web", Proc. of the Int. Conf. on Data Engineering, pp. 164-171, 1996

■ 저자소개 ■



이해성
Lee, Hae Sung

2008년 3월~현재
경기대학교 일반대학원 컴퓨터과학
과 석사과정
2008년 3월 경기대학교 정보과학부 전자계학과
학사 졸업
관심분야 : 유비쿼터스 컴퓨팅, 검색엔진
E-mail : hae2425@gmail.com



권준희
Kwon, Joon Hee

2003년 3월~현재
경기대학교 정보과학부 컴퓨터과학
전공 교수
1992년 3월 숙명여자대학교 전산학과 학사 졸업
1994년 3월 숙명여자대학교 대학원 전산학과
석사 졸업
2002년 숙명여자대학교 대학원 전산학과
박사 졸업
관심분야 : 유비쿼터스 컴퓨팅, 상황인식 컴퓨
팅, 모바일 컴퓨팅, 데이터베이스
E-mail : kwonjh@kyonggi.ac.kr

논문접수일 : 2008년 2월 22일
게재확정일 : 2008년 3월 5일