

웹의 검색 성향을 고려한 웹 방송정보 재구성 기법

김원철⁰ 황인준

아주대학교 정보통신전문대학원 정보통신공학과
{wc323⁰, ehwang}@ajou.ac.kr

Web contents reorganizing scheme based on user web browsing

Woncheol Kim⁰ Eenjun Hwang
The Graduate School of Information and Communication, Ajou University

요약

최근 다양한 무선 단말기의 보급과 네트워크 기술의 발전으로 인하여 무선 단말기를 이용한 인터넷 접속이 증가하고 있다. 특히 텍스트 기반의 방송사 웹 콘텐츠들은 무선 단말기의 제한된 환경에서도 비교적 손쉽게 접근이 가능하기 때문에 이용률이 높다. 그러나 대부분의 방송사 웹페이지들은 한 페이지에 세분화된 섹션과 많은 내용을 담고 있기 때문에 제한된 화면과 입력장치를 가진 무선 단말기를 이용하여 사용자가 원하는 부분에 접근하기까지 반복적인 스크롤링을 해야 하는 불편한 점이 있다. 이러한 문제점을 해결하기 위해 본 논문에서는 방송사 웹 페이지내에서 사용자가 선호하는 방송사 웹의 섹션을 추출하고, 무선 환경에 적합하도록 방송사 웹페이지들의 각 섹션의 순서를 재구성하는 기법을 제안한다. 본 논문에서 제안된 기법을 통해 사용자는 무선 단말기의 각종 단점을 극복함과 동시에 방송사 웹에서 선호하는 섹션의 맞춤형 방송사 웹 서비스를 제공받을 수 있다.

1. 서론

최근 디지털 기술의 급속한 발전으로 디지털 방송의 도입은 방송환경에 커다란 변화를 가져오고 있다. 가용채널과 부가 방송 서비스의 폭발적인 증가로 인해, 사용자에게 다양한 방송 프로그램과 많은 서비스의 기회가 제공 되고있다.

이로 인해, 다양한 사용자의 요구에 부응할 수 있는 시청자 중심의 방송 서비스 시대가 열릴 것으로 예상된다. 이러한 변화에 발맞추어, 무선 단말기의 대중화와 무선 인터넷 기술의 발전으로 사용자들은 무선 단말기를 이용하여 언제 어디서나 자신이 원하는 방송 정보를 얻기를 원한다. 이와 같은 환경 변화로 인해, 신문이나 책자에 활자로 출력된 프로그램 가이드 대신 전자적인 정보를 통해 방송 프로그램을 조회할 수 있는 서비스인 전자 프로그램 가이드 (EPG: Electronic Program Guide)[1]가 출현 하였다.

현재 디지털 방송을 위해 제공되고 있는 대부분의 전자 프로그램 가이드 서비스들은 디지털 TV 화면이나 각 방송사 웹 사이트 및 무선 단말기 등을 통해서 서비스되고 있다. 그러나, 방송 채널의 수가 증가하여 사용자가 시청할 수 있는 프로그램이 다양해 지면서, 시청자는 많은 방송 프로그램들 가운데 자신의 취향에 적합한 프로그램을 찾아야만 하는 부가적인 노력이 필요하게 되었다. 또한, 실질적으로 사용자 입장에서 원하는 방송 프로그램 정보를 찾기 위해서 수백개의 채널에 달하는 방송 프로그램에 대한 정보를 일일이 검색하는 것은 매우 힘든 일이며, 무선 단말기를 이용하여 각 방송사의 웹 사이트 정보를 일일이 찾아보는 것도 힘들다. 사용자가 원하는 방송에 대한 키워드를 직접 입력하는 방식 또한 일시적인 사용자의 기억에만 의존함으로써 폭넓은 검색이 어렵다.

이러한 문제를 해결하기 위해 기존의 연구들은[2, 3, 4] 방송 메타데이터를 이용한 프로그램 정보의 단순 검색뿐만 아니라 사용자가 미리 작성한 사용자 프로파일을 이용하여 시청자 취향에 맞는 프로그램을 제공하고자 노력 하였다.

그러나 이러한 연구는 사용자가 프로파일 작성에 대하여 무지하거나, 프로파일 작성을 번거롭게 생각하여 제대로 이용되

지 못한다. 또한 프로파일을 작성하였다고 하더라도 작성시점 이후에 사용자의 취향변화에 대해 사용자가 적절히 반응하지 않음으로 불필요한 정보를 제공받게 된다.

본 논문에서는 사용자의 웹의 검색성향을 통하여 사용자가 관심을 가지고 있는 방송 정보들에 대해 추출한다. 웹 사이트 내에서 사용자의 의미있는 접근 성향을 추출하기 위해서는 웹 검색시 사용하는 링크의 링크앵커 텍스트를 이용한다. 사용자가 무선 단말기를 이용하여 방송정보및 방송 웹페이지를 요청할 때, 파악된 선호도를 바탕으로 웹페이지를 재구성해 줌으로써 무선 단말기의 각종 단점을 극복할 수 있으며 방송에서 선호하는 프로그램의 맞춤형 서비스를 제공받을 수 있다.

2. 관련 연구

최근 대부분의 방송이 아날로그방송에서 디지털 TV방송으로 전환됨에 따라 점차적으로 많은 채널이 신설, 추가, 변경, 삭제등 채널변경이 많아지면서 되면서 각 채널을 소개해 주어야할 필요성이 증가하였다. 사용자는 선택의 폭이 늘어난 만큼 자신의 구미에 맞는 프로그램을 선택하는 시간과 노력을 더 들여야만 하게 되었다. 게다가 이제는 인터넷 방송국까지 생겨나면서 수 많은 프로그램 정보를 한꺼번에 모두 얻기 힘들게 되었다. 예를들어, 사용자가 원하는 프로그램을 찾기 위해 수많은 채널의 정보를 일일이 브라우징 하는 것은 사용자에게 매우 번거로운 일이기 때문이다.

이러한 문제를 해결하기 위해 MPEG-7 UsageHistory DS(Description Schemes)나 MPEG-21을 이용하여 사용자의 정보(프로그램 시청정보, 시간대별 선호장르)를 저장하고, 사용자 정보를 동적으로 추출하여 일치하는 EPG 데이터를 우선적으로 표시함으로써 EPG를 효율적으로 제공하는 방법을 제시하는 연구들이 활발히 진행되고 있다. 그러나 이러한 연구는 사용자 프로파일 작성에 대한 작업이 필요하고, 사용자의 성향의 변화에도 불구하고, 사용자가 이를 수정해 주지 않으면 불필요한 정보들이 제공되는 문제점을 안고 있다.

3. 사용자 웹 검색성향 추출 기법

웹 사이트내에서 사용자의 의미있는 접근 성향을 추출하기 위해 웹 사이트에서 사용자의 검색 패턴을 알아야한다. 따라서 본 논문에서는 웹 환경에서 사용자의 검색 패턴을 찾기위해 순회패턴(Traversal pattern)[5]을 이용 한다. 본 논문에서는 순회패턴 알고리즘을 이용하여 사용자의 웹 사이트 검색순서를 분석한다. 웹 사이트의 페이지들이 서로 연결되어 있다고 가정할 때, 사용자는 자신이 원하는 정보를 얻기위해 웹 사이트의 링크를 따라 앞이나 뒤로 페이지를 검색하는 경우가 일반적이다. 그 결과로 어떤 페이지의 링크는 그 내용보다는 위치 때문에 방문될 가능성이 있다. 예를 들면 웹 환경에서 웹 사이트를 하나의 트리구조형식으로 생각해 볼 경우, 형제 노드위치에 있는 페이지를 방문하기 위해서 사용자는 새로운 URL을 사용하는 대신에 보통 역방향 아이콘을 누르고 다시 형제 노드위치의 페이지를 방문한다. 결과적으로 웹 사이트 내에서 사용자의 기록을 가지고 있는 로그파일에서 의미있는 사용자 접근 패턴을 추출하기 위해서는 이러한 역방향 검색의 영향을 고려하여 관심있는 실제 접근 패턴을 발견하도록 해야 한다. 이런 관점에서 역방향에 대한 참조는 주로 내용을 검색하기 보다는 검색의 편리를 위해 이루어진다고 가정한다. 그러므로, 역방향 참조는 같은 사용자 접근에 의해 이전에 방문한 객체를 재방문하는 것을 의미한다. 역방향 참조가 일어나면, 순방향 참조 경로는 종결된다. 이렇게 얻어지는 순방향 참조 경로를 최대 순방향 참조(Maximal forward reference)라고 한다.

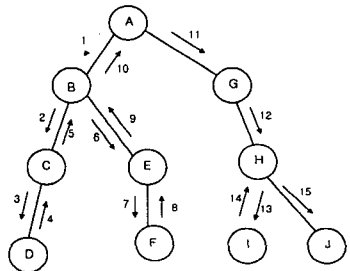


그림 1. 순회 패턴(Traversal patterns)의 예

본 논문에서는 사용자가 웹페이지 검색시 그림 1과 같은 (A, B, C, D, C, B, E, F, E, B, A, G, H, I, H, J)순회패턴을 가지고 있다고 가정한다. 알고리즘을 적용하여 순방향 참조 집합을 추출하면 (ABCD, ABEF, AGHI, AGHJ)와 같다. 즉 사용자가 웹을 검색함에 있어서 역방향 링크를 다시 검색할 경우 순방향 참조가 종료되고 하나의 최대 순방향 참조를 만들게 된다. 알고리즘을 통해 순방향 참조 집합을 구할 때 각 노드에 해당하는 링크앵커의 키워드들을 추출한다. 링크앵커는 링크내 <a>태그 사이에 존재하는 문자로, 예를 들면 korea news homepage같은 태그가 존재할 경우 "korea news homepage"을 링크 앵커 텍스트라 한다. 링크앵커 텍스트는 링크에 대한 가장 중요한 정보들을 포함하고 있다. 본 논문에서 링크앵커 텍스트의 키워드들을 링크앵커 키워드라고 정의하며, 링크에 대해 가장 잘 설명해주는 링크앵커 키워드들을 이용하여 방송사 웹의 섹션을 구분한다.

순방향 참조에서, 각 링크에서 결정된 방송사 웹 섹션들을 비교하고 분석하여 최대 순방향 경로의 방송사 웹 섹션을 최종 결정한다. 예를 들면 사용자가 (A, B, C, D)의 순서로 링크를 방문했을 경우 각 링크앵커 텍스트에서 결정된 방송사 웹의 섹션들이 (news, movie, news, news)라고 하면, 최대 순방

향 경로가 news에 관련된 것이라 판단할 수 있다. 본 논문은 추출된 각 순방향 참조 집합에 대하여 반복적으로 링크앵커 키워드들을 추출한 후에 방송사 웹 섹션을 판단한다. 그림 2과 같이 웹 사이트에서 사용자의 링크정보와 사용자의 정보를 이용하여 방송사 웹 사이트에서 사용자의 일반적인 접근 순서가 N→L→E→S→M→N→S와 같음을 알 수 있다.

Link information	Link anchor keywords	News section
newsArticle.jhtml?type=politicsNews&storyID=2844205	Congress, Solve, Medicare Problems	News (N)
newsArticle.jhtml?type=politicsNews&storyID=2844971	Israel, Palestinians, Peace, Steps	
.../tsclsi/markets/marketstory/10090182.html	lifestyle, babies, Adoption.	Lifestyle (L)
~/financeNewsArticle.jhtml?type=storyID=2845238	style, life, Food, Wine Festival	
.../newsArticle.jhtml?type=entertainmentNews&storyID=2844570	Thanks, Memory, Americans, Rock	Entertainment (E)
.../newsArticle.jhtml?type=sports&storyID=2843665	star, Premier, League, club	Sports (S)
...

그림 2. 웹에서 사용자의 검색 성향

각 순방향 참조과정에서 링크앵커 키워드를 추출한 후 각 노드의 방송사 웹 섹션을 결정한다. 사용자가 웹페이지를 검색함에 있어 순방향 참조과정 중 역방향 참조가 일어나면, 순방향 참조가 끝나고 최대 순방향 참조가 결정된다. 최대 순방향 참조가 결정되면, 각 순방향 참조에서 결정하였던 방송사 웹 섹션의 정보를 통하여 최대 순방향 참조에 대한 방송사 웹 섹션을 최종 결정한다.

사용자가 검색을 재시작 할 경우, 위와 같은 과정을 반복적으로 수행하여 접근 순서에 대한 성향을 분석하고 사용자의 최근 성향이 점차 어떻게 변하고 있는지를 분석한다. 사용자가 재구성된 정보를 요청할 경우 분석된 접근 성향 순서를 이용하여 문서를 재구성한다.

4. 실험 및 분석

대부분의 웹페이지들이 데스크톱에 최적화되어있어 무선 단말기를 이용하여 웹페이지를 검색할 경우 콘텐츠의 전체 구조를 이해하기 위해 많은 스크롤 작업이 필요하고 낮은 대역폭으로 인해 페이지의 내용을 가져오기까지 많은 시간이 걸린다. 따라서 본 논문의 제안된 시스템의 성능 평가하기 위해 사용자가 방송사 웹페이지를 검색하기 위한 스크롤 수와 단말기를 통해 전달받게되는 페이지사이즈에 대해 성능 평가를 한다. 표 1과 2는 사용자 선호사항을 기준으로 방송사 웹페이지를 재구성한 경우의 이점을 보여준다. 방송사 웹에 포함되는 일반적인 6개의 섹션 News, Lifestyle, Drama, Sports, Entertainment, Movie를 각각 P, T, B, S, E, M이라고 하였다. 실험에서는 모바일 화면에 제공되는 전형적인 방송사 웹페이지의 순서가 (P, B, T, S, E, M)라고 가정한다. 재구성된 방송사 웹섹션이 이점을 갖는지 그렇지 않은지를 판단하기위해 위/아래로의 스크롤링의 수를 계산한다. 주어진 순서의 방송사 웹섹션을 읽기 위해 사용자 스크롤의 총합을 이용하여 비교한다. 예를 들면, 무선 단말기 화면에 단지 3개의 섹션만 보인다고 가정하고, 사용자 A의 방송사 웹 사이트에서 검색패턴이 일반적으로 (N, D, E, L, M, N) 같을 경우, 재구성하지 않은 모바일 단말기를 이용할 경우 (N, D, L, S, E, M)로 나열된 화면을 통해 방송사 웹을 검색해야한다.

모바일 단말기의 화면을 통해 단지 3개의 섹션을 보여주기 때문에 처음 모바일 단말기를 통해 (N, D, L)의 섹션을 볼 수 있다. 처음 사용자가 관심을 가지는 방송사 웹섹션이 N이므로 사용자는 스크롤을 하지 않아도 된다.

표 1. 재구성하지 않은 경우 스크롤의 수

User	Display sequence	User sequence	The number of scrolling	Total number
A	N.D.L.S.E.M	N.D.E.L.M.N.E	0,0,2,0,1,3,2	8
B		M.D.E.N.D.M.D	3,2,1,2,0,3,2	13
C		M.E.D.N.S.L.S	3,0,2,1,1,0,0	7
D		M.L.S.M.E.L.M	3,1,0,1,0,1,1	7
E		E.M.S.L.N.D.N	2,1,0,1,2,0,0	6
F		N.D.S.E.L.M.L	0,0,0,1,1,0,1	3

사용자가 일반적으로 이후에 검색을 하는 이후의 방송사 웹섹션은 D이므로 역시 스크롤을 하지 않아도 된다. D를 검색한 후 사용자가 검색하는 방송사 웹섹션은 E이므로 사용자는 위로 2번의 스크롤을 통해 (L, S, E)의 화면이 보이게 하고 E의 방송사 웹 섹션을 검색한다. E의 방송사 웹섹션을 검색후, 사용자가 일반적으로 검색하는 방송사 웹 섹션은 L이므로 사용자는 스크롤을 하지 않아도 필요한 방송사 웹섹션을 검색할 수 있다. 방송사 웹섹션을 검색후 M을 보기 위해 사용자는 위로 1번의 스크롤 작업을 해야 하고 모바일 스크린상에 (S, E, M)화면을 보이게 하여 방송사 웹 섹션을 검색한다. 마지막으 로 N와 E의 섹션을 보기위해 5번의 스크롤 작업을 해야한다.

그러나 테이블 2와 같이, 방송사 웹 섹션을 재구성한 경우 사용자 A는 단지 4번의 스크롤 작업을 필요로 한다; 예를 들 면, 무선 단말기 화면에 단지 3개의 섹션만 보인다고 가정하 고, 사용자 A의 방송사 웹 사이트에서 검색패턴이 일반적으로 (N, D, E, L, M, N)같은 경우, 재구성을 한 모바일 단말기를 이 용할 경우 (N, D, E, L, M, S)로 나열된 화면을 통해 방송사 웹 을 검색해야한다. 모바일 단말기의 화면을 통해 단지 3개의 섹션을 보여주기 때문에 처음 모바일 단말기를 통해 (N, D, E)의 섹션을 볼 수 있다. 처음 사용자가 관심을 가지는 방송사 웹섹션이 N이므로 사용자는 스크롤을 하지 않아도 된다.

사용자가 일반적으로 검색을 하는 이후의 방송사 웹섹션은 D와 E이므로 역시 스크롤을 하지 않아도 된다. E를 검색한후 사용자가 검색하는 방송사 웹섹션은 L이므로 사용자는 위로 1 번의 스크롤을 통해 (D, E, L)의 화면이 보이게 하고 L의 방송 사 웹 섹션을 검색한다. 방송사 웹섹션을 검색후 M을 보기위 해 사용자는 위로 1번의 스크롤 작업을 해야 하고 모바일 스크린상에 (E, L, M)화면을 보이게 하여 방송사 웹 섹션을 검색 한다.

표 2. 재구성한 경우 스크롤의 수

User	Display sequence	User sequence	The number of scrolling	Total number
A	N.D.E.L.M.S	N.D.E.L.M.N.E	0,0,0,1,1,2,0	4
B	M.D.E.N.S.L	M.D.E.N.D.M.D	0,0,0,1,0,1,0	2
C	M.E.D.N.S.L	M.E.D.N.S.L.S	0,0,0,1,1,1,0	3
D	M.L.S.E.N.D	M.L.S.M.E.L.M	0,0,0,0,1,0,0	1
E	E.M.S.L.N.D	E.M.S.L.N.D.N	0,0,0,1,1,1,0	3
F	N.D.S.E.L.M	N.D.S.E.L.M.L	0,0,0,0,1,1,0	2

마지막으로 N와 E의 섹션을 보기위해 2번의 스크롤 작업을 해야한다. 즉, L를 검색하기 위해 1번, M을 검색하기 위해 1 번 그리고 N를 검색하기 위해 2번이 필요하다.

지금까지 방송사 웹섹션을 재구성한 경우가 재구성하지 않은 경우보다 더 적은 양의 스크롤을 필요로 한다는 것을 알 수 있다. 결과적으로 사용자 선호하는 섹션으로 재구성하는

방법이 작은 화면을 가진 무선 단말기에 편리하다는 것을 알 수 있다.

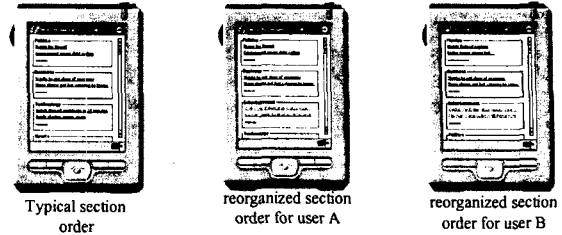


그림 3. 방송사 웹페이지의 재구성

그림 3은 사용자가 정치, 경제, 오락등의 순으로 방송사 웹 섹션을 선호할 때, 방송사 웹페이지를 요청했을 경우의 화면 과 경제, 오락, 기술등의 순서로 방송사 웹섹션을 선호하는 사 용자가 방송사 웹페이지를 요청했을 경우 재구성되지 않은 경 우와 사용자의 선호 사항을 통해 재구성된 경우의 화면을 보 여준다.

5. 결론 및 향후 연구 과제

무선 단말기 사용자의 인터넷 접속이 증가하면서 점차 웹페 이지는 유선 사용자뿐 아니라 무선 사용자의 요구를 만족시켜 야 할 필요성이 증가하였다. 이러한 필요성으로 인해 현존하 는 웹페이지들이 각각의 기기에 맞게 표현되어야 하지만, 무 선 단말기는 데스크톱에 비해 제한된 인터페이스와 화면 크기 를 가지므로 사용자에게 사용의 불편함을 가져온다. 따라서 본 논문에서 무선 단말기 사용자의 요구를 만족시키고 편리한 인터페이스를 제공하기 위하여 선호사항 분석 기반의 방송사 웹 페이지 전달기법을 제안하였다. 특히 무선단말기에 맞는 페이지의 별도 준비없이 현존하는 웹페이지를 통하여 맞춤형 서비스를 제공하게 되었다. 사용자 선호사항을 추출함으로써 방송사 웹의 각 섹션을 사용자의 검색 성향을 이용하여 재구 성함으로써 무선 단말기의 작은 화면에서 스크롤링을 줄였고, 사용된 페이지셋은 본래 페이지와 비교하여 작은 파일 사이즈 로 변형시켰다. 따라서 무선이란 환경의 낮은 대역폭의 제한 을 완화시켰고 작은 화면에서 가독성을 증가시켰다.

향후 계획으로는 사용자의 선호사항을 분석한 후에 무선 환 경에서 방송사 웹 사이트만이 아닌 일반적인 사이트에도 웹 컨텐츠를 적용시키는 것이다.

6. 참고 문헌

- [1] <http://www.worlddab.org/>
- [2] Buyukkotken, H. Garcia-Molina, and A. Paepcke, T. Winograd, "Power Browser: Efficient Web Browsing for PDAs," Proceedings of CHI' 2000, ACM Press, Amsterdam, 2000.
- [3] S. Ahn, Y. Cho, J. Choi, J. Kim and C. Ahn " Personalized and Interactive Broadcasting System," Workshop on Personalization in Future TV Sonthofen, July 13th and 14th, 2001
- [4] Baudisch, P. and Brueckner, L. " TV Scout: Guiding users from printed TV program guides to personalized TV recommendation," In Proceedings of TV'02: the 2nd Workshop on Personalization in Future TV (Collocated with AH2002), May 28, 2002, Malaga, Spain.
- [5] Ming-Syan Chen, Jong Soo Park, Philip S. Yu " Efficient Data Mining for Path Traversal Patterns" , IEEE Transactions on Knowledge and Data Engineering, Vol 10, No. 2, pp. 209-221.