모바일 웹 검색을 위한 루티드 최단 트리 알고리즘

이우기 이정훈○ 송종수
인하대학교 산업공학과
wookeylee@gmail.com, bigjames.lee@gmail.com○, jaegal83@gmail.com

Rooted Shortest Tree Algorithms for Efficient Retrieval in Mobile Web

Wookey Lee, Junghoon Lee○, Jongsu Song
Department of Industrial Engineering, Inha University

1. 서 론

지난 10년 동안, 인터넷의 성장과 기하급수적인 웹 정보의 증가는 정보검색의 다양한 발전을 이루었다. 특히 최근에는 아이폰, 안드로이드폰과 같은 스마트폰의 등장으로 모바일 기기를 이용한 웹 검색의 수요는 더욱 더 증가하게 되었다[1]. 그러나 이러한 모바일 장치들은 일반 PC에 비해 상대적으로 자원적으로 제한이 있기 때문에 (예를 들면 낮은 주파수 대역폭, 적은 저장공간, 제한된 배터리 수명, 낮은 CPU속도 등) 모바일 장치를 이용한 웹 검색 기술 발전에 커다란 장애물로 작용한다. 따라서 모바일 웹에서 좀 더 효율적인 검색 방법론이 제시될 필요가 있다. 웹 검색의 방법론 중 하나인 웹 그래프 검색은 각 웹 페이지들을 그래프의 하나의 노드로 표현하고, 각 페이지간의 링크를 각각의 노드를 연결하는 아크로 표현한다. 이렇게 웹 페이지에 대한 하나의 웹 그래프가 생성되면 생성된 그래프를 이용해 PageRank, Kleinberg, Simrank, tf-idf 등 각 검색 특성에 맞추어 메저를 적용하면 각 링크에 가중치를 적용 할 수 있고, 그 값을 이용하여 사용자가 원하는 정보를 제공한다. 일반적인 웹 페이지에서는 충분한 네트워크 환경과 하드웨어자원이 있으므로 일반적인 웹 검색을 통해 충분한 검색이 가능하다. 하지만 모바일 환경에서는 전체 웹 페이지에 대한 구조를 이용해서 사용자에게 제공하기에는 리소스 및 네트워크 사항이 매우 제한적이다. 여기서 구조를 이용한다는 말은 포탈검색과 같은 검색이 아닌 홈페이지 단위를 검색할 때를 의미한다. 즉, 특정 홈페이지나 사이트 단위에서 메인 페이지에 도달했을 때, 하위의 여러 단계를 거쳐서 접근할 하부 페이지들을 미리 구조화함으로써 사용자에게 검색하는 홈페이지 내비게이션 역할을 할 수 있도록 구조화 하는 것이다. 이때, 홈페이지란 메인 페이지에서 링크로 닿을 수 있는, 즉 같은 도메인에 속한 웹페이지 그룹을 뜻한다. 여기서 구조화란 트리 형태로 페이지를 만드는 것을 의미하는데 이러한 구조는 모바일 검색을 할 경우 순환이 제거된 가장 짧은 거리로 원하는 페이지에 도달 할 수 있는 트리로 구성하는 것을 의미한다[2,3,4]. 트리 구조 생성 시 검색 지연시간이 생길 수 있으므로 트리 구조로 생성된 홈페이지를 상위 페이지 즉 트리의 루트로부터 가까운 페이지들을 비동기 로딩을 함으로써 사용자는 느끼지 불편함을 느끼지 못하면서 빠르게 검색을 진행할 수 있다. 본 논문에서는 *RST*라는 트리를 통해서 이것이 모바일 검색에서 유의함을 입증하고자 한다.

2. 본론

지금까지 모바일 검색에 대한 여러 가지 많은 이슈가 존재했지만 트리를 이용해 구조화하여 검색을 하는 방법은 제시되지 않았다. 본 논문에서는 유저가 여러 사이트를 탐색하지 않고도 자신이 원하는 정보를 얻을 수 있도록 트리 구조가 최소한의 거리로 제공된 페이지들을 반환하도록 하며, 모바일 검색에 있어서 구조의 필요성을 고려한 가장 효율적인 기법을 제안하고자 한다. 구체적으로는 모바일 검색의 효율성을 높이기 위한 단계적 구조, 즉 특별한 트리 구조를 사용하여 검색을 한다. 대표적으로 트리 구조를 만들기 위한 두 가지 방법이 있는데, 최단트리(Rooted Shortest Tree)와 최적 나눔트리(Optimum Branching Tree)가 그것이다.

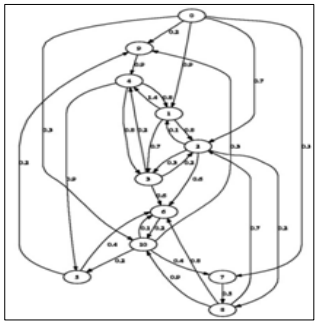
3. 실험

본 장에서는 두 가지 예제 홈페이지를 가지고 실험하였다. 첫 번째는 이해를 돕기 위한 간단한 가상 홈페이지를 통해서 실험하였고, 두 번째는 실제 운영 중인 스탠포드 대학(stanford.edu)의 홈페이지를 이용하였다. [그림1]은

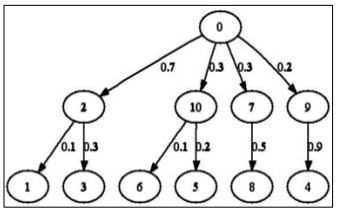
본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 육성·지원사업(NIPA-2010-C1090-1031-0002)의 연구결과로 수행되었음.

가상 홈페이지를 웹 그래프를 표현한 것이다. [그림1]을 가지고 각각 (a)*RST*와 (b)*OBT*를 구하면 [그림2,3]과 같이 두 가지 형태의 트리가 그려진다. [그림1]의 모든 링크, 즉 모든 아크의 합은 각각 3.6, 3.3으로 *OBT*가 더 작다.

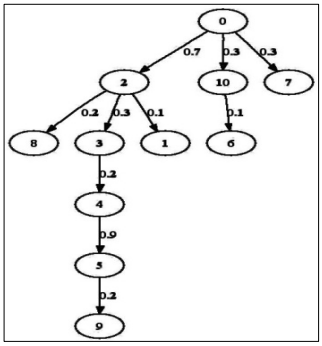
하지만 *OBT*의경우 9번 페이지와 같은 깊은 차수의 페이지들이 생성하게 된다. 만약 9번 페이지에 가고 싶다면 *RST*를 이용하면 한 번 클릭으로 도달할 수 있지만, *OBT*의 경우 5번 클릭으로 도달 할 수 있다. 전체적으로 *RST*의 경우 2번 클릭으로 모든 페이지에 도달 할 수 있게 된다. 그 이유는 *RST*의 경우 메인 페이지로부터 검색해 나간다고 했을 때, 빠른 시간에 검색할 수 있도록 비용이 최소가 되는 페이지들을 루트와 가까운 위치에 배치한다. 반면 *OBT*의 경우 총 아크의 값을 줄이기 위해 상대적으로 많은 페이지를 거쳐서 도달할 수 있는 페이지가 존재하게 된다. [그림4]는 Top-k 페이지 당 누적 비용을 나타낸 그림이다.



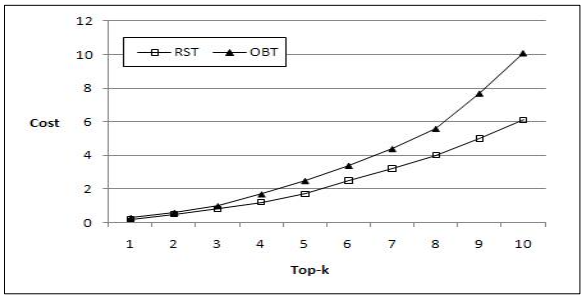
[그림1] 가상 웹 그래프



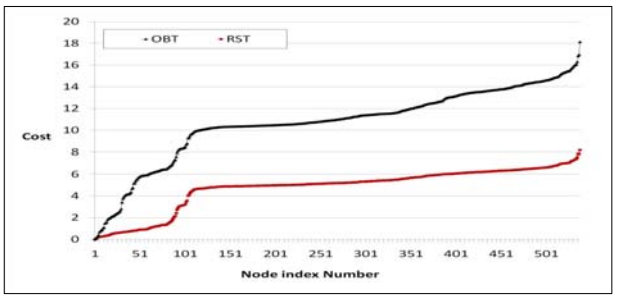
[그림2] RST



[그림3] OBT



[그림4] TOP-K 페이지 당 비용



[그림5] 사례 사이트의 OBT, RST 누적비용 비교

여기서 사용되는 비용이란 최악의 경우 사용자가 모든 페이지를 한 번씩 클릭해 나간다고 했을 때, 누적된 비용을 의미한다. 즉 각 페이지를 검색해 나갈 때 마다 각 비용은 누적되고 각 페이지를 검색하는 데 소모되는 시간이라고 가정했을 때 단계를 거듭할수록 *RST*가 더욱 효율적임을 알 수 있다. 두 번째 실험은 사례 사이트로서 스탠포드대학의 실제 웹사이트 정보를 이용하여 실행한 실험 결과이다. [그림5]는 사례 사이트의 *RST*와 *OBT*의 누적 비교를 나타낸 그림이다. 누적 비교를 보면 알 수 있듯이 깊이에 따른 Cost 차이가 점점 가중되는 것을 확인할 수 있으며, 그 내용은 앞서의 방법에서 설명된 바와 같다.

4. 결론

본 논문에서는 효과적인 모바일 검색을 위하여 그래프 기법에 기반한 구조적 검색의 필요성을 설명하면서, 가장 중요한 두 가지 트리 기법을 제시하고 그중에서 *RST*트리는 *OBT*트리보다 항상 효율적인 결과를 제공함을 증명하였으며, 실험을 통해 그것을 확인하였다. 따라서 *RST* 트리를 이용해, 모바일 사용자에게 검색 가이드를 제안하고, 적은 수의 액세스를 통해 제한된 리소스 환경에서도 보다 빠르게 원하는 페이지에 접근 할 수 있음을 알 수 있다.

현재 모바일 시장은 애플의 앱스토어와 아이패드 등장으로 인해 크게 성장하고 있으며, 국내의 경우도 모바일 시장은 더욱 더 팽창될 전망이다. 이러한 환경에서 각종 모바일 장치들이 보다 효과적으로 활용 될 수 있도록 다양한 방법들이 제안되는 시점에서 본 논문에서 제시한 트리 구조를 통한 웹 검색은 매우 시의성이 있다 할 것이다. 향후 추가적인 인덱스 등을 통해 훨씬 고효율 시스템이 가능할 것이며, 현재 키워드 기반의 인덱스 그리고 그래프 인덱스 등을 통해 성능을 높일 수 있는 연구가 후속될 것이다.

5. 참고문헌

[1] J. Antoniuk, A. Nascimento. A light weight PDA-friendly collection fusion technique. *SIGIR*,439-440,2003.
[2] D. Lizorkin, et al. Accuracy estimate and optimization techniques for SimRank computation. *VLDBJournal*,19(1),45-66,2010.
[3] D. Metzler, et al. Building enriched document representations using aggregated anchor text. *SIGIR*:219-226,2009.
[4] E. Glover, *etal.* Inferring hierarchical descriptions. CIKM 2002, pp.507-514.