

해시태그를 이용한 게시물 탐색 최적화 기법

최동하¹, 이예형², 김태현³, 허원재⁴, 김영종[✉]

*승실대학교 소프트웨어학부

✉승실대학교 소프트웨어학부

carmania71@naver.com, kanto00@naver.com, th2252@gmail.com, hky0105@naver.com

✉youngjong@ssu.ac.kr

Post Search Optimization Technique Using Hashtags

Dongha Choi¹, Yehyeong Lee², Taehyeon Kim³, Wonjae Heo⁴,

Youngjong Kim[✉]

*School of Software, Soongsil University

✉School of Software, Soongsil University

요 약

기존에 제안된 웹 검색 전략에서 벗어나, 해시태그로 카테고리를 반영하여 더 빠르고 효율적인 게시물 검색 전략을 제안하고, 수학적으로 접근하여 비교한다.

1. 서론

현재의 인터넷 환경은 일반 사용자들이 능동적으로 정보 생산에 참여하는 웹 2.0이다. 사용자들이 정보를 생산하고 개방하며 공유하는 영역이 점점 늘어나며 웹에 저장되는 정보의 양은 더욱 많아지고 있다. 그에 따라 이 많은 양의 정보를 처리하는 방법이 중요해졌고, 사용자가 원하는 정보를 정확하고 빠르게 찾는 방향으로 검색 알고리즘이 연구되어 왔다. 하지만 홍수처럼 쏟아지는 정보 속에서 사용자가 원하는 정보를 빠르고 정확하게 찾는 데에는 한계가 있다. 이러한 상황 속에서 본 논문에서는 태그를 사용하여 게시물을 검색하는 방법을 제안하고 기존의 게시물 검색 방식과 비교하여 성능을 논한다.

2. 관련 연구

제안 방식인 해시태그 탐색을 알아보기에 앞서 기존의 웹 검색 전략에 관한 연구들을 살펴본다. Marchionini(2006)는 탐색적 검색 프로세스에 주목하여, 사용자가 취하는 탐색적 검색의 전략에 대해 밝혀내었다[1]. 연쇄적으로 나타나는 웹 검색 행동에 주목하여, 검색어 뭉치의 개념을 만들고 사용자의 검색 행동으로부터 검색어를 고찰해낸 연구도 있다.[2] 기존의 정보 검색 모델보다 정보 검색자의 실제 행동에 훨씬 더 가깝고, 결과적으로 효과적인 인

터페이스 설계에서 검색자의 사고를 더 잘 안내하는 'berrypicking'이라고 하는 온라인 및 기타 정보 시스템에서의 새로운 검색 모델에 대해 논의하기도 하였다.[3] 이처럼 웹 검색 전략 연구들은 사용자 입장에서 검색을 이해하고 검색 인터페이스를 향상시키려는 노력들이 이어져 왔다.

3. 해시태그를 이용한 검색(제안방식)

태그는 간단하면서도 웹 페이지 혹은 게시글의 주제를 함축적으로 담고 있어 검색 과정에 온전한 지표를 제공해주고, 사용자에게도 더 만족스러운 검색 결과의 도출을 유도해 준다. 이는 기존의 키워드 검색과 달리 다중 분류 구조에 있어서 확장성을 제공하여 검색의 효율을 높여준다.

다음 그림의 상황을 가정해보자. 해시태그를 사용하여 검색된 정보를 저장하는 데이터의 무방향성 그래프는 $G = (V, M)$ 이고, 여기서 V 는 전체 데이터의 집합, M_n 은 카테고리별로 분류된 각 n 번째의 데이터 집합, 즉 하나의 해시태그이다. T 는 모든 해시태그의 집합이며, $M \subset T$ 이다.

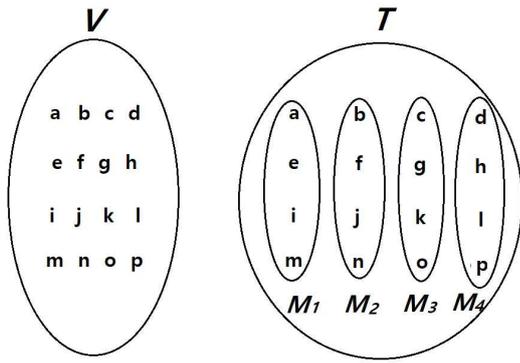


그림 1. 예시

참고문헌

- [1] Mather, B.D. Making up titles for conference papers. Ext. Abstracts CHI 2000, ACM Press (2000), 1-2.
- [2] 이중식. 웹검색 행태 연구. 한국문헌정보학회지, 45(2), 209-228. (2011).
- [3] Bates, M. J. The design of browsing and berrypicking techniques for the online search interface. Online review, 13(5), 407-424. (1989).

사용자가 게시물 k 를 찾기 원한다고 할 때, 기존의 검색 전략으로는 탐색 알고리즘에 따라 전체 게시물 집합 V 의 원소들과 비교하여 결과를 도출한다. 반면에 제안된 검색 전략을 활용하면, 집합 T 의 k 가 포함된 해시태그 M_3 에 먼저 접근하고 그 내에서 검색하므로 더 빠르고 효율적이다.

4. 결론

웹 서비스의 접속 장벽이 낮아지고 사용자들이 점점 늘어남에 따라 웹에서 생산되는 정보의 양이 급속도로 증가하고 있다. 사용자가 웹을 통해 검색하는 시간과 비용을 단축해 주기 위해서 사용자가 원하는 정보를 정확하고 빠르게 검색해주는 방법을 고안해야 한다.

본 논문에서는 게시판의 각 게시물을 해시태그로 분류하여 검색하였다. 기존의 웹 검색 전략에서 벗어나 해시태그로 태그 카테고리를 반영하여 더 빠르고 간단한 검색 전략을 제공하고자 했고, 집합을 이용한 수학적 접근을 통해 이 전략을 확인하였다.