**Random Walks on the Click Graph---** Nick Craswell and Martin Szummer

검색 엔진은 어떤 문서들이 어떤 쿼리에 대해 클릭 되었는지 기록할 수 있고, 이 query-document 쌍은 ‘soft'한 relevance judgement로 사용된다. true judgement와 비교하면 클릭 로그는 noisy 하고 sparse한 relevance information을 준다. 이 논문에서는 Markov random walk model을 커다란 클릭 로그에 적용해서 주어진 쿼리에 대한 문서들의 확률적 순위를 제공하려고 한다. 이 모델은 쿼리에 대해 아직 클릭 되지 않은 관련 있는 문서들을 검색하게 하고 효율적으로 이 문서의 순위를 매긴다.

사용자가 검색한다는 관점에서 클릭 되었지만 관련 없는 문서는 클릭 데이터 상에서 노이즈를 구성한다. 관련 있지만 클릭 되지 않는 문서들은 클릭 데이터에서 sparsity를 구성한다. 클릭 데이터 상에서 노이즈를 줄이기 위한 시도로는 사용자 행동에 대한 추가적인 정보를 이용하는 방법이 있다. 예를 들어, 문서 클릭 뒤에 사용자의 브라우징 패턴을 고려하는 것이 있다. 이 방법은 관련되지 않은 클릭 데이터를 찾음으로써 노이즈를 줄일 수 있다. 이 논문은 sparsity 문제에 초점을 맞춘다.

클릭 데이터는 쿼리와 문서, 두 가지 타입의 노드를 갖는 bipartite 그래프로서 표현된다. 쿼리와 문서 쌍에 대한 클릭이 있을 때 엣지가 연결된다. 엣지는 모든 사용자의 전체 클릭 수에 따라 weight를 갖는다.

query - document : search. 쿼리가 주어지면 관련 있는 문서를 찾는다. (document-query와 함께 반복적으로 번갈아 일어남)

query - query: suggestion. 사용자가 좀 더 원했을 만한 다른 쿼리를 찾는다.

document - query: annotation. 주어진 문서에 대해 관련된 쿼리를 붙여놓는다.

document - document: relevance feedback. 사용자와 관련된 문서 외에 추가로 관련된 문서를 찾는 것.