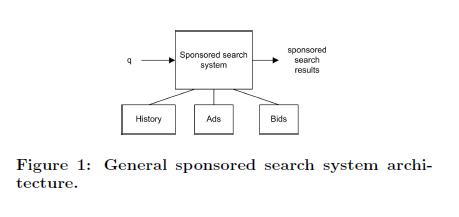
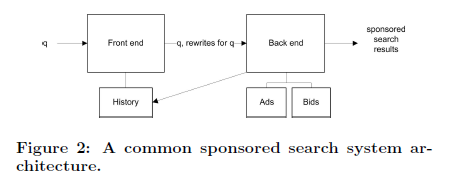
****

**Simrank++: query rewriting through link analysis of the click graph ---** Ioannis Antonellis, Hector Garcia Molina, and Chi Chao

이 논문은 sponsored search를 위한 query rewriting 문제를 다루고 있다. 기본적으로 과거 사용자의 쿼리에 맞추어 클릭 되었던 광고를 기록한 historical click graph를 다시 작성한다. 먼저 query q가 주어졌을 때, 우리는 처음 q와 비슷한 쿼리(사용자가 관심 있어 할 광고들을 가지고 있는) 들을 찾아내기 위해 Sim-rank를 고려한다. 이 논문에서는 기존의 Sim-rank가 적절하게 쿼리의 유사성을 찾는 데에는 부적절하다고 보고 두 개의 강화된 버전의 Sim-rank를 제시한다. 하나는 클릭 그래프의 edge에 weight를 준 것이고 다른 하나는 “evidence”를 이용하는 것이다.

기존의 sponsored search에서는 사용자의 쿼리와 관련된 광고들이 웹 서치 결과 상단이나 옆 쪽에 나타났다. 이 광고들의 위치는 쿼리에 대한 랭킹 스코어와 관련이 있다. Sponsored search system이 아래 그림에 나타나 있다. 시스템이 이용 가능 한 광고와 bids로 이루어진 데이터 베이스에 접근한다. 그리고 각 bid는 쿼리 q와 광고, 가격 p로 이루어져 있다. 각 bid에 대해서 가격제시자는 광고가 사용자가 입력한 쿼리에 잘 들어맞는 경우 가격을 지불한다.

  
많은 쿼리들의 경우, 직접적인 bids가 없기 때문에 이 경우 sponsored search system은 쿼리를 제시한 사용자가 흥미 있어 할 만한, 다른 광고를 찾는다. 그래서 이제 문제가 되는 것은 들어온 쿼리에 대해 사용자가 관심있어 할 만한 다른 광고들을 찾는 일이다.



일반적인 sponsored search system은 두 부분으로 이루어져 있다. Front end 부분은 사용자가 입력한 query q를 받고 주기적으로 다시 작성된 쿼리(ex) 카메라를 쿼리로 입력한 경우, 디지털 카메라, 배터리 등이 유사한 쿼리가 된다) 들을 내뱉는다. 그러면 Back end가 다시 작성된 쿼리를 보고 해당되는 결과를 도출한다. 이렇게 sponsored search system이 두 부분으로 나뉘어지며 back-end의 복잡성을 감소시켜 빠르게 bids를 바꿀 수 있다.

그러나, front end 부분에서 쿼리를 변형할 때, 크게 두 가지의 문제가 있다. 첫 번째는, 쿼리 자체가 짧기 때문에 그로부터 충분한 정보를 얻어내지 못해 유사한 다른 쿼리들을 많이 찾을 수 없다는 것이고 두 번째 문제는, bid 데이터베이스에 상대적으로 적은 쿼리들이 존재하기 때문에, 우리가 텍스트적으로 관련 있는 것들을 많이 찾아도 그걸로는 충분하지 않다는 점이다. 따라서 이 논문에서는 쿼리를 다시 작성하는 문제에 집중한다. Back-end 는 click 그래프(click을 기록 - 사용자가 쿼리를 입력했을 때 광고들에 의해 생성된 click) 를 생성하는 데 이 click 그래프는 weighted bi-partite graph이다.(쿼리가 한쪽에, 다른 쪽에 광고가 있는) 저자가 제기하는 방식은 유용한 rewritten query를 찾기 위해서 클릭 그래프 상의 연결을 분석하는 것이다. 그리고는 direct, indirect의 연결을 모두 살핀다. 즉, Simrank++이라는 기존의 Simrank에서 확장된 버전을 제시하고 이것을 기존의 query rewriting 기술과 비교한다.

Contribution:

1. Sponsored search 환경에서의 query rewriting 프레임 워크 제시
2. SimRank가 쿼리와 ads(유사성 점수) 간에 정확한 관계를 이끄는 데 실패한 사례 보기
3. SimRank의 2가지 확장 제시 – 클릭 그래프에서의 엣지의 weights을 고려한 것, 여러 쿼리들 간에 유사성을 보여주는 “evidence”를 고려한 것
4. Yahoo!에서의 실제 클릭 그래프를 이용해서 제시된 query rewriting 기술 평가