****

****

**Link Analysis for Web Spam Detection ---** Luca Becchetti, Carlos Castillo, Debora Donato, Ricardo Baeza-YATES, and Stefano Leonardi

Web spam : 페이지의 랭킹을 속여 올리려는 모든 시도들이 Web spam이다. 크게 content(keyword) spam과 link spam의 두 그룹으로 나뉜다.

spam page : spamming에 사용되었거나 다른 스팸 페이지로부터 상당한 양의 score를 받고 있는 페이지.

content spam : 페이지의 내용에 변화가 있는 것. 수많은 키워드들을 삽입하는 방법이 대표적

link spam : 사이트의 링크 구조에 변화가 있는 것.

link farm : 뺵빽하게 연결된 페이지의 집합으로써 페이지 랭크 알고리즘을 속이려는 것이 목적.

이 논문에서 소개하는 spam detection 알고리즘의 타겟은 그 페이지가 갖고 있는 대부분의 링크가 link farm에 참여함으로 만들어진 것들이다. (topological spamming)

link farm에 참여한 페이지들은 in link의 degree가 높으나 그래프의 나머지 부분들과는 관계를 잘 맺고 있지 않다. link farm은 광고를 삼으로써 nonspam site로부터 링크를 얻거나 합법적인 목적으로 사용되고 있던 도메인을 사곤 한다.