

대학 홈페이지 하이퍼링크 중심도와 방문자수 관계 분석

권혜정*, 전미현**, 배태웅***, 김규태****
*고려대학교 정보전산처, **고려대학교 대학교육개발원
고려대학교 입학처, *고려대학교 전기전자공학부
e-mail : elphis@korea.ac.kr

A Correlation between the Centrality of the Hyperlinks and the Number of Visits in the Homepages of the University

Hyejung Kwon*, Mihyun Jeon**, Tae-Woong Bae***, Gyu-Tae Kim****
*Office of Information & Computing, Korea University
**Center for Teaching & Learning, Korea University
***Korea University Admission Office
****School of Electrical Engineering, Korea University

요 약

대학 내 기관별 웹사이트의 방문자 패턴을 분류하고, 하이퍼링크에 따른 네트워크 분석을 통해 방문자수와의 연관성을 연구하였다. 웹사이트 언어에 따라 영문 홈페이지는 공통적으로 낮은 방문 패턴을 보였으며, 단과대학이 개별 학과에 비해 높은 방문자수를 보였다. 웹사이트 하이퍼링크의 중심도(Centrality) 분석에서는 사이중심도(Betweenness centrality)를 정도(degree)로 정규화한 값이 방문자수의 상한 값에 대응되는 현상이 발견되었다. 홍보효과를 위해 홈페이지 내용, 구조 설계뿐만 아니라 웹 상에 적절한 노출 전략도 도움이 됨을 알 수 있었다.

1. 서론

학령인구의 감소로, 학생을 유치하고 정원을 충원해야 하는 문제가 대학에게는 과거보다 더 중요해진 상황이다. 이러한 상황에서 대학은 오프라인상에서 대면 홍보를 하고, 온라인상에서 홈페이지를 통하여 각 기관이 전달하고자 하는 정보를 전달하고 이미지를 구축한다. 홈페이지의 방문자수를 파악하고, 이용률을 높이는 것이 학교를 홍보하는 데 중요한 요소가 되었다. 하지만 단과대학, 학과, 산하기관 등 다양한 특성을 가진 기관들이 있어 학교 전체 홈페이지를 통한 홍보효과를 위해 어떻게 홈페이지가 설계되어야 할 지에 대한 체계적인 연구는 많지 않은 실정이다. 학교본부에서 개별 기관의 홈페이지 제작을 지원해주는 학교가 늘고 있는 현 시점에서 홍보 극대화를 위한 설계 방향을 가늠하는 것은 효율적인 홍보 정책을 위해서도 중요하다.

한 연구는 하이퍼링크 연결이 전문가의 사이트 품질 판단과 중요한 관계가 있음을 보여준다. 이 연구는 웹사이트로 들어오는 하이퍼링크의 수를 사이트 신뢰성의 지표로 간주하고, 하이퍼링크 네트워크에서 사이트로 들어오는 중심성이 방문 행동 및 웹 사이트의 신뢰도와 유의미한 관련이 있음을 발견했다. 또한 웹사이트가 다른 사이트와 하이퍼링크 네트워크를 형성할 때, 웹마스터는 그 사이트에 평균보다 높은 신뢰성을 가지기를 요구한다는 것을 발견했다. 따라서

신뢰할 만한 웹사이트일수록 그 웹사이트로 향하는 하이퍼링크를 더 많이 가지게 된다. 들어오는 링크가 많을수록 웹사이트 방문자가 많아지고 그 반대로 마찬가지이다. 단순한 링크수가 아닌 링크수를 웹사이트 내 페이지수로 정규화한 값을 각 웹사이트별 지명도로도 제시하고 있다 [1]. 신뢰도나 지명도는 방문자가 가지는 믿을 수 있는 사이트에 대한 방문 동기를 부여할 수 있다. 또 한 연구에서는 국회의원과 서울시 의회의원들이 운영하는 웹사이트의 아웃링크를 분석하여 국회의원 당선 여부 등에 따라 관리 정도가 차이가 있음을 밝히고 있다 [2]. 이는 웹사이트의 외부로의 링크가 각 사이트의 철학과 지향점에 따라 다르며 방문자를 고려함에 있어 어떤 부분을 중요시 하는가를 의미한다고 하겠다.

기존 연구들이 웹사이트의 연결성을 네트워크 분석을 통해 밝히고 있지만, 방문자 수와 같이 실질적인 효과로 정량적으로 연계 지은 연구는 많지 않다. 최근 구글 애널리틱스와 같이 방문자에 대한 데이터도 다양하게 얻을 수 있게 됨으로써 네트워크 분석을 통해 방문자 패턴에 미치는 영향도 살펴볼 수 있게 되었다. 홈페이지의 방문 수에 영향을 끼치는 요소가 검색 엔진 노출, 관련자 수, 콘텐츠, 관련 SNS 활성화 정도 및 홈페이지 링크의 연결 정도 등 다양하기에 실제 방문자 수와 같은 효과는 다양한 요인들이 복합적으로 나타날 것임을 충분히 예상할 수 있다. [3]

본 연구에서는 이 요소 중 홈페이지간의 하이퍼링크 관계(제휴네트워크)가 홈페이지 방문자수 와 어떤 관계가 있는지 분석하였다.

2. 연구방법

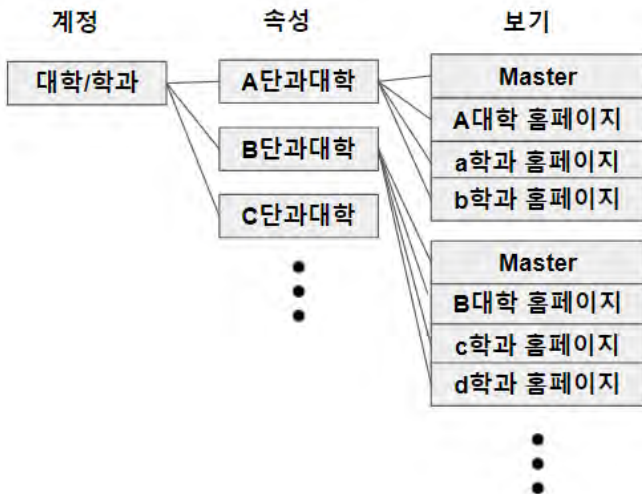
2.1 분석 대상

연구 대상으로 삼은 대학교 산하의 홈페이지는 본부, 대학, 학과, 대학원, 부속기관, 기타 기관에 걸쳐서 700 여개가 존재한다. 대학 산하 기관 중 단과대학과 학과에 교내 구성원의 대부분을 차지하는 학생과 교원이 소속되어 연구 및 교육의 역할을 담당하고 있다. 따라서 약 12 개 단과 대학과 약 58 개 학과의 국문, 영문 홈페이지를 본 연구의 분석대상으로 한정하였다.

본 연구를 위해 각 단과대학, 학과 홈페이지에 구글 애널리틱스 분석 스크립트를 삽입하였다. 구글 애널리틱스에서 얻을 수 있는 정보는 방문자수, 체류시간, 방문 페이지, 접속 기기, 접속 국가, 접속 경로, 접속 시간, 이탈율 등 다양하다. 홈페이지 구축 효과는 홈페이지의 노출 정도와 이용률로 측정할 수 있으므로, 분석 가능한 정보 중 방문자 수 데이터를 분석의 대상으로 삼았다.

2.2 분석 과정

2.2.1 홈페이지 분석 환경 구축 및 방문자수 추출 과정



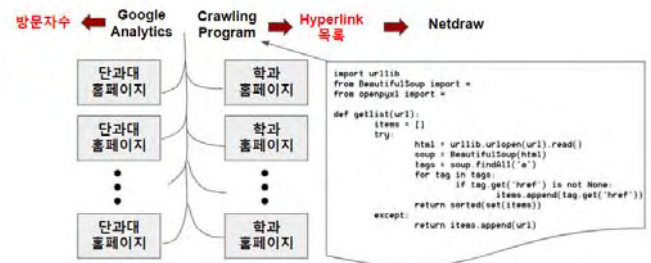
(그림 1) 대학/학과별 홈페이지 분석환경 구성 체계

학교 내 단과 대학과 학과 홈페이지 별로 방문자수를 얻기 위해서 구글 애널리틱스 스크립트를 각 홈페이지 소스에 삽입하였다. 구글 애널리틱스는 계정(최대 100 개), 속성(최대 50 개), 보기(최대 25 개)의 구조로 이루어져 있다. 각 기관별로 계층적 체계를 구성하기 위하여, '대학/학과' 계정을 만들고, 속성에는 각 단과대학의 리스트가 위치하도록 하였다. 향후에 홈페이지를 분석할 때, 홈페이지 간 방문 경로도 분석이 가능하게 하기 위하여 속성마다 Master 보기를

추가하여 단과대학, 학과 홈페이지의 데이터를 모두 볼 수 있도록 하였다. 그리고 보기에서 단과 대학 내 실제 개별 홈페이지의 데이터를 볼 수 있도록 구성하였다.(그림 1)

모든 학과의 방문자수를 추출하기 위하여 구글 스프레드시트에 구글 애널리틱스 부가기능을 설치하였다. 이 부가기능에서 레포트 설정 파일 생성 기능을 활용하여 레포트 설정 파일을 생성하였다. 각 뷰의 ID를 삽입하고 추출하고자 하는 정보에 방문자수를 설정하고 레포트 생성 기능을 이용하여 레포트를 생성하였다. 그리고 구글 스프레드 시트의 탭 별 셀 참조 기능을 이용하여 홈페이지 별 방문자수 리스트를 추출하였다.

2.2.2 홈페이지 별 하이퍼링크 추출 과정



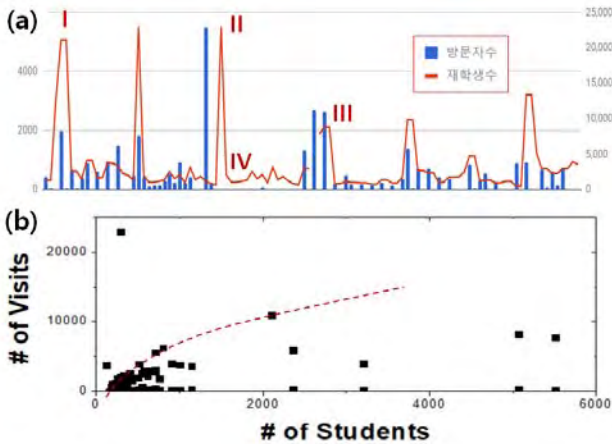
(그림 2) 대학/학과별 하이퍼링크목록 추출 과정

각 홈페이지에서 링크를 하고 있는 하이퍼링크 목록은 개별 사이트에 접속하여 <a> 태그를 찾아 목록을 만드는 파이썬 프로그램을 통해 추출하였다. (그림 2) 이때 <a> 태그에 연결된 링크가 상대 혹은 절대 주소로 표현된 것이 같은 링크의 다른 경로를 가리키고 있을 경우는 주소를 축약하여 하나의 링크로 대응하였다. 2 edge 모드로 표현하기 위해 인용하고 있는 웹페이지와 인용되고 있는 웹주소로 나누어 추출하였다. 하이퍼링크를 네트워크도로 표현하고 개별 사이트의 중심성을 계산하기 위해서 netdraw 프로그램을 활용하여 네트워크도를 생성하였다. Netdraw 로 분석할 수 있도록 2 edge 모드의 엑셀 자료를 netdraw의 vna 파일 형식으로 변환하는 매크로 프로그램을 작성하여 python 으로 추출한 하이퍼링크 목록을 일괄 변환하였다.[4]

3. 연구결과

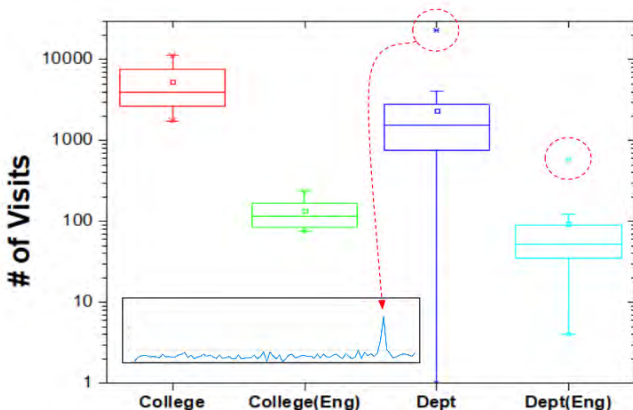
그림 3 에 기관별 재학생 수와 기관별 웹사이트 방문자수를 각각 (a) 기관별 (b) 재학생 수에 따른 방문자수 상관관계 그래프로 나타내었다. 재학생 수가 많은 기관에 대해 웹사이트 방문자 수가 많은 경우와 적은 경우를 각각 (I 유형), (II 유형)으로 표현하였다. (그림 3a) 재학생 수가 적은 기관에 대해서도 방문자수가 많은 경우 (III 유형), 방문자 수가 적은 경우 (IV 유형)으로 찾을 수 있다. 방문자 수가 많은 홈페이지

이지가 효용성이 높다고 생각한다면 특별히 (II 유형)에 속하는 기관은 좀 더 홈페이지 활용도에 대해 관심이 필요할 것이다. 상관관계 그래프(그림 3b)에서 계산된 상관관계수는 $R^2=0.04767$ 로 0 에 상관관계가 크지 않다고 볼 수 있지만, 점선으로 표시한 것처럼 소속 재학생 수가 많을수록 소속 홈페이지 방문자 수에 기여하는 부분적인 경향도 찾을 수 있다.



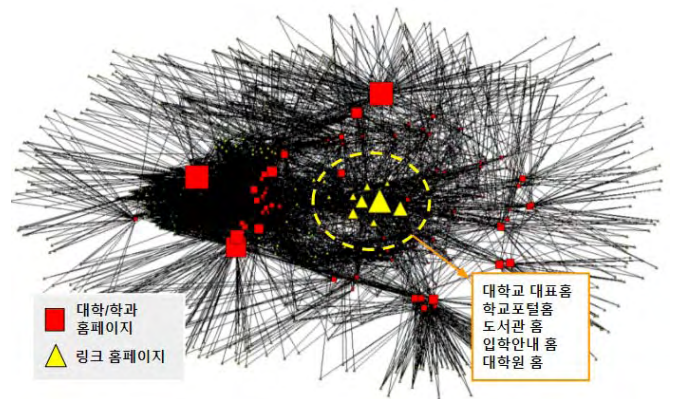
(그림 3) 기관별 재학생수와 웹사이트 방문자수 관계도 (a) 기관별 (b) 재학생수-방문자수 상관관계도

그림 4는 구글 애널리틱스로 추출한 방문자수를 단과대학과 학과, 그리고 국문과 영문 홈페이지로 구분하여 상자도로 나타낸 것이다. 단과대학이 학과에 비해 방문자수가 많은 것은 국문이나 영문 모두에 나타나 있다. 영문 홈페이지는 국문 홈페이지에 비해 40 배 가량 방문자수가 적었다. 동그라미 점선으로 표시한 학과사이트는 학과에 특정한 분포에서 많이 벗어나고 있다. 이 경우 구글트렌드에서 해당 검색어가 급격히 증가한 시점과 정확히 일치하여 일반적인 방문자 패턴에서 벗어나고 외부에서 직접 검색해서 방문자가 해당 사이트에 들어오는 특수한 상황이었음을 확인하였다. 영문 홈페이지에도 타 학과 홈페이지 보다 높은 방문자수가 나타난 것이 관측되었으나, 구글트렌드에서는 뚜렷한 변화를 보이지 않았다. 구글 애널리틱스 방문자패턴에서도 급격하지 않고 평균적으로 방문자수가 높아 뉴스화 되었거나 노출되어 방문자수가 많은 것은 아닌 것으로 판단된다.



(그림 4) 웹사이트 특성 별 방문자수 패턴. 삽입 그림: 구글트렌드에서 특정 학과로 검색한 검색 패턴

그림 5는 개별 대학/학과 홈페이지에서 링크하고 있는 웹사이트를 2-edge 모드 네트워크도로 표현한 것이다. degree 1에 해당하는 노드는 제거하여 사이트 간 연관성이 있는 노드에 초점을 두었다. 노드의 크기는 사이 중심성(betweenness centrality)에 비례하도록 하여 하이퍼링크를 많이 인용하고 있는 홈페이지를 손쉽게 찾아볼 수 있다. 네트워크도의 중심에는 예상대로 학교 대표홈페이지, 포털 홈페이지, 도서관, 입학, 대학원 홈페이지와 같이 개별 대학/학과에서 많이 인용하는 홈페이지가 위치한 것을 볼 수 있다. 대학/학과 홈페이지에서 사이중심도가 높은 학과의 경우 링크하고 있는 웹사이트가 특히 많아 홈페이지의 효용성을 높이고자 하는 특별한 노력이 있음을 엿볼 수 있었다.



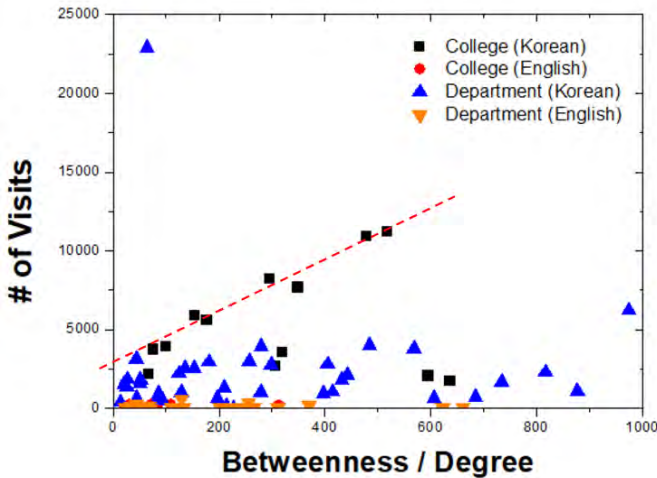
(그림 5) 사이중심도에 따라 크기가 주어진 교내 홈페이지의 하이퍼링크 네트워크도

그림 6은 사이 중심도(Betweenness Centrality)를 정도(Degree)로 나눈 값과 구글 애널리틱스로 분석된 각 사이트 방문자수를 나타낸 그래프이다. 여러가지 중심도 지수가 방문자 수와 특별히 연관관계가 보이지 않았지만, 사이중심도를 정도(Degree)로 나눈 것은 그림과 같이 방문자 수의 상한선을 설정하는 형태로 나타났다. 정도(Degree)가 높을수록 관계하는 구성원의 수가 많을 것이라는 가정을 해본다면 사이중심도를 정도로 나눈 값은 정규화된 방문자수와와의 연계가 가능함을 그림에서 확인할 수 있다.

$$\# \text{ of Visits} = 16.8 \times (\text{Betweenness}/\text{Degree}) + 2626.3 \text{ (식 1)}$$

식 1)에 따라 그림 6)에 표시된 점선은 아무리 링크를 홈페이지에 많이 더한다 하더라도 더 이상 방문자수가 늘지 않는 최대치가 있다는 것을 의미한다. 단과대학이 각 학과의 링크를 포함하는 것을 감안한다면 단과대학으로 들어와서 각 과로 가는 비율이 높은 것을 나타낸다고 할 수 있다. 즉 하이퍼링크가 잘 설계되어 웹페이지에 반영될 경우, 하위 기관 웹사이트 방문시 허브가 되는 상위 기관도 같이 방문자수가 늘

어남을 의미한다. 최대 방문자 수의 상한선을 하이퍼링크 설계에 따라 도달할 수 있는 최대 효용 지수로 참조해 볼 수 있을 것이다.



(그림 6) Betweenness/Degree 와 웹사이트 방문횟수와 의 관계: 단과대학, 학과별 방문횟수

영문사이트의 경우는 단과대학이나 학과나 상관없이 거의 최저치 바닥권을 형성하고 있는데 영문 사이트 링크를 통해 타 사이트로 가는 유도 효과가 거의 없다는 것을 의미한다. 한 학과 홈페이지는 방문자수 상한선 보다 월등히 큰 방문자수가 관측되었는데, 이는 방문자수를 산정하는 기간 중 특별히 언론의 주목을 받은 과정이 있음에 기인한 것으로 판단하였다. 제휴네트워크가 인지도와 신뢰도를 높여 방문자수에 영향을 줄 수 있다는 기존 연구결과를 고려해 본다면 [3], 본 연구결과는 정량적인 증거로서 의미가 있을 것이다. 이처럼 허브가 되는 단과대학별 링크를 잘 설계함으로써 방문자 입장에서 효율적인 기관 홈페이지로서 접근성을 높이는데 참고해볼만한 지표로서 활용될 수 있을 것이다.

4. 결론 및 제언

본 연구에서는 대학내 홈페이지 특성 별 방문자수 범위, 홈페이지의 하이퍼링크와 방문자 수의 관계에 대해 살펴보고, 사이 중심도 (Betweenness Centrality)를 정도(Degree)로 정규화한 값이 홈페이지 방문자 수와 관련이 있는 지수가 될 수 있음을 확인하였다. 그 결과, 홈페이지 방문자는 단과대학 국문 홈페이지, 학과 국문 홈페이지, 단과대학 영문 홈페이지, 학과 영문 홈페이지 순으로 숫자가 많음을 확인 할 수 있었고, 구글 애널리틱스 분석을 통하여 단과대학의 홈페이지를 경유하여 하위 학과에 방문하는 방문자수가 많음을 확인 할 수 있었다. 홈페이지 간의 링크 관계에서는 예상대로 학교 대표홈페이지, 포털 홈페이지, 도서관 홈페이지, 대학원 홈페이지 등이 가장 많은 링크를 받고 있음을 확인 할 수 있었다. 또한 정도 (Degree)로 정규화한 사이 중심도(Betweenness

centrality/ degree)가 증가할 때 홈페이지 방문자수가 증가할 수 있음을 확인하였지만 링크를 더 늘려도 더 이상 방문자수가 증가하지 않는 상한도 존재함을 알 수 있었다.

본 연구의 결과 중 특정 학과가 언론에 노출된 특정시점에 방문자수가 급증하였다는 사실은 홈페이지 기획 및 운영 방향에 시사점을 가져다 주었다. 효율적으로 홈페이지 방문자수를 늘리기 위해서는 내실 있는 홈페이지 콘텐츠의 배치와 더불어 신뢰성 있는 연관 기관에서 해당 홈페이지로 향하는 하이퍼링크를 효율적으로 구성하여야 함을 알 수 있었다. 하지만 링크만으로는 더이상 방문자수가 증가하지 않는 한계치가 존재한다. 단, 언론에 노출되는 이벤트가 있는 경우 방문자수가 이 한계치를 월등히 뛰어넘는 것을 볼 때, 언론에 노출되는 이벤트를 적절히 계획하는 것도 홈페이지 방문자수를 늘리는 방법임을 알 수 있었다.

영문 홈페이지의 링크 정도는 국문 홈페이지와 비슷한데 영문 홈페이지의 방문자수가 국문 홈페이지의 방문자수보다 현저히 떨어지는 것은 영문홈페이지 링크 구성에 변화가 필요함을 의미한다. 영문 홈페이지의 이용률을 높이기 위해서는 외국대학과 MOU 를 할 때 외국대학 홈페이지에 우리대학 영문홈페이지에 링크를 달아주는 등 외부적인 노력도 도움이 될 것으로 판단된다.

참고문헌

- [1] 박한우, 정보화정책 제 10 권제 4 호, 2003 년 겨울, p.36~52, "한국 웹사이트들 사이의 하이퍼링크를 통한 제휴 네트워크 구조에 대한 탐사적 연구".
- [2] 박윤희, 사회과학연구, 2010, 36(2), p139~160, "의원 웹사이트의 네트워크 분석에 관한 연구".
- [3] Han Woo Park, and Mike Thelwall, Journal of Computer-Mediated Communication, July 2003, Hyperlink Analyses of the World Wide Web: A Review.
- [4] Netdraw 형식: <https://gephi.org/users/supported-graph-formats/netdraw-vna-format/>