

논문 2011-48CI-6-10

한국 블로그 공간의 시간의 흐름에 따른 특성 변화

(Characteristics of Korean Blogosphere over Time)

하 지 운*, 배 덕 호*, 김 상 욱**

(Jiwoon Ha, Duck-Ho Bae, and Sang-Wook Kim)

요 약

블로그가 웹 공간에서 중요한 정보 전달 매체로 떠오름에 따라, 블로그 공간의 현상들이 중요한 사회 현상으로 취급되게 되었다. 이러한 블로그 공간의 출현은 기업들에게 블로그 공간을 대상으로 하는 새로운 비즈니스 모델 수립의 기회를 제공한다. 블로그 공간은 시간이 경과함에 따라 지속적으로 변화한다. 블로그 공간에서의 성공적인 비즈니스 정책 수립을 위해서는 시간의 흐름에 따른 블로그 공간의 변화에 대한 이해가 필요하다. 본 논문에서는 한국 블로그 공간 고유의 특징들이 한국 블로그 공간의 특성에 미치는 영향에 중점을 두어 한국 블로그 공간의 변화에 대해 심도 있는 분석을 수행한다. 본 연구의 결과가 효과적인 알고리즘들을 개발 및 새로운 비즈니스 정책 수립에 도움이 될 것으로 기대한다.

Abstract

As blogs become an important medium through which to communicate and exchange information on the World Wide Web, phenomena in the blogosphere are treated as important social phenomena. The advent of the blogosphere may provide opportunities for establishing new business models targeting online world. The blogosphere changes over time. To establish successful business policies in the blogosphere, the changes in the characteristics of the blogosphere should be understood. In this paper, focusing on the influence of convenient features of the Korean blogosphere, we analyze the changes of the characteristics of the Korean blogosphere over time. We expect that the results of these analyses would be helpful in developing effective algorithms and in establishing new business models.

Keywords : 사회연결망분석, 데이터마이닝, 블로그 공간

I. 서 론

블로그는 사용자가 자신의 글을 온라인상에 저장할 수 있는 개인 홈페이지의 일종이다. 블로거는 블로그의 이용자를 의미하며, 포스트는 블로거가 블로그에 작성

한 글을 의미한다. 블로거는 작성된 포스트에 대해 스크랩, 워인글, 댓글 달기 등 다양한 액션들을 수행할 수 있다. 댓글은 다른 블로그의 포스트에 자신의 의견을 남기는 기능이다. 워인글은 다른 블로그의 포스트에 링크를 남기고 자신의 블로그에 해당 포스트로의 링크를 포함하는 새로운 포스트를 작성하는 기능이다. 스크랩은 한국 블로그 공간에 한하여 사용되는 다른 블로그의 포스트를 자신의 블로그로 복사하는 기능이다. 본 논문에서는 블로거들과 포스트들, 그리고 작성된 포스트에 대해 블로거들이 행한 액션들로 구성된 공간을 블로그 공간(blogosphere)이라 한다^[1].

최근, 블로그 공간이 크게 활성화됨에 따라, 웹 공간에서 블로그 공간이 차지하는 비중이 높아지게 되었다. 이에 따라 블로그 공간에서 일어나는 현상들이 대중매

* 학생회원, ** 평생회원, 한양대학교 전자컴퓨터통신공학과

(Department of Electronics and Computer Engineering, Hanyang University)

※ 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것입니다(2008-0061006). 또한, 지식경제부 및 정보통신산업진흥원의 IT융합 고급인력과정 지원사업(NIPA-2011-C6150-1101-0001)과 두뇌한국 21 사업의 부분적인 지원을 받아 수행되었습니다.

접수일자: 2011년10월10일, 수정완료일: 2011년11월2일

체나 여론, 마케팅에 많은 영향을 미치는 중요한 사회 현상으로 취급되게 되었다. 이러한 블로그 공간의 출현은 기업들에게 새로운 비즈니스의 기회를 제공한다. 블로그 공간에서의 성공적인 비즈니스 정책 수립을 위해서는 블로그 공간의 특성과 블로그 공간을 이용하는 블로거들의 행동 패턴에 대한 올바른 이해가 선행되어야 한다.

블로그 공간은 시간이 경과함에 따라 지속적으로 변화한다. 블로그 공간에 신규 블로거들이 유입되기도 하며, 기존의 블로거들이 탈퇴하기도 한다. 또한, 블로거들은 자신의 생각을 알리기 위해 새로운 포스트를 작성하거나, 관심이 있는 포스트들에 대하여 스크랩, 워인글, 댓글 달기 등 액션들을 수행하기도 한다. 이러한 블로그 공간의 변화를 분석함으로써 블로그 공간이 발전해 온 경향을 알 수 있으며, 나아가 향후 발전 과정을 예측할 수 있다.

한국 블로그 공간은 한국의 높은 인터넷 보급률, 최신 IT기술에 대한 높은 친화력, 그리고 사용자 친화적인 한국 블로그 공간 고유의 특징들을 바탕으로 블로그의 일반적인 역할인 일인 미디어로서의 역할 외에 소셜 네트워크적인 역할까지도 포괄하는 특징적인 공간으로 발전하였다. 또한, 한국 블로그 공간은 그 고유한 편의 기능들로 인해 웹 공간 또는 기존의 블로그 공간과 상이한 구조적 특성을 보인다^[2].

한국 블로그 환경의 고유한 특징에 주목하여 한국 블로그 공간의 구조적 특성을 분석하거나 블로그 공간에서의 정보 파급을 분석하는 연구가 이루어져왔다^[2~7]. 그러나 시간의 흐름에 따른 블로그 공간의 특성 변화에 대해 심도 있는 분석을 수행한 논문은 존재하지 않았다. 본 논문에서는 시간의 흐름에 따른 한국 블로그 공간의 변화를 관찰하고 한국 블로그 공간 고유의 특징이 시간의 흐름에 따른 변화에 미치는 영향을 분석한다. 첫째, 시간의 흐름에 따른 Bow-tie 컴포넌트의 비율 변화를 통해 블로그 연결망의 연결성이 시간의 흐름에 따라 점점 높아짐을 보인다. 둘째, 블로그 공간에서의 포스트 생성 및 스크랩 경향과 포스트 생성에 대한 경향을 분석하여 요일 별 포스트의 생성에는 특정한 경향이 나타나지 않지만 시간 별 포스트 생성 특정한 경향이 존재함을 밝힌다. 셋째, entropy plot 분석을 통해 블로그 공간에서의 블로거들의 행동이 uniform하게 일어남을 밝힌다.

본 논문의 구성은 다음과 같다. 제 II장에서는 본 연

구와 관련된 기존의 연구들을 기술한다. 제 III장에서는 실험을 사용된 블로그 데이터를 설명하고, 블로그 데이터로부터 블로그 연결망을 구성하는 방법에 대하여 설명한다. 제 IV장에서는 시간의 흐름에 따라 한국 블로그 공간의 특성이 변화하는 모습에 대해 분석한다. 제 V장에서는 시간의 흐름에 따른 포스트 생성 및 스크랩 경향을 분석한다. 끝으로, 제 VI장에서는 본 연구의 결론을 내린다.

II. 관련 연구

본 장에서는 웹 공간과 블로그 공간을 대상으로 한 기존 연구에 대해 간략히 설명한다. 기존의 웹 공간에 대한 연구들은 웹 페이지를 노드로, 웹 페이지들 간의 하이퍼링크를 에지로 웹 연결망을 구성한 후, 구성된 웹 연결망을 대상으로 특성을 파악하기 위한 다양한 실험들을 수행하였다^[8~10]. 블로그 공간을 대상으로 하는 연구들은 포스트를 노드로 하고 두 포스트 사이의 하이퍼링크를 에지로 하는 포스트 연결망을 구성하거나, 블로그를 노드로 하고 블로그의 포스트와 다른 블로그의 포스트 사이의 하이퍼링크가 존재하는 경우 이를 두 블로그 간의 에지로 간주하는 블로그 연결망을 구성하고, 이에 대한 분석을 수행하였다^[11~13].

참고 문헌 [9]은 대량의 웹 연결망 데이터를 대상으로 하는 실험을 통해 웹 연결망의 페이지의 degree의 분포와 WCC, SCC의 크기의 분포가 power-law 분포를 따르는 것을 밝히고 average connected distance, diameter, 임의의 두 노드 간의 경로 존재 확률 등을 통해 웹 연결망의 연결성을 분석하였다. 또한, 웹 연결망을 거시적으로 모델링한 Bow-tie 구조를 제안하고 크게 SCC, IN, OUT, Tendrils의 4가지 컴포넌트로 구성되는 Bow-tie 각 컴포넌트의 비율을 밝혔다. 그림 1은 웹 공간의 Bow-tie 구조를 나타낸 것이다. SCC(strongly connected component)의 모든 노드들은 SCC의 다른 모든 노드들로 도달 가능한 방향성 경로(directed path)를 가진다. IN의 모든 노드들은 SCC로 도달 가능한 경로를 가지고 있지만, SCC에서 도달 가능한 경로는 가지고 있지 않다. OUT의 모든 노드들은 IN과는 반대로 SCC에서 도달 가능한 경로는 가지고 있지만, SCC로 도달 가능한 경로는 가지고 있지 않다. Tendrils는 IN 또는 OUT에 연결되어 있는 컴포넌트로, Tendrils에 속한 노드들은 SCC로 도달하거나 SCC로부

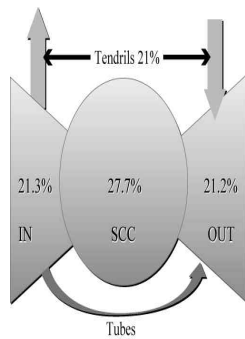


그림 1. 웹 공간의 Bow-tie 구조^[9]
Fig. 1. Bow-tie structure of the Web^[9].

터 도달할 수 있는 경로가 존재하지 않는다. 특히, Tendrils 중 SCC를 거치지 않고 IN에서 OUT으로 직접 도달 가능한 경로를 갖는 노드들의 집합을 Tubes라 정의한다. SCC는 전체 WCC의 약 25%를 차지하고, IN, OUT, Tendrils는 각각 전체 WCC의 약 20%를 차지하는 것으로 나타났다.

참고 문헌 [12]에서는 블로그와 포스트의 in-degree와 out-degree가 모두 power-law 분포를 따르는 것을 밝히고, 더 나아가 블로그 당 포스트 수, 블로그 간의 링크 수 등의 분포도 power-law 분포를 따르는 것을 밝혔다. 또한, 블로거의 out-degree와 in-degree 간의 관계를 분석하여 블로거에 대한 관심과 그 블로거의 활동 간에 주목할 만한 관련성이 없음을 밝히고, 블로그 연결망에서의 정보의 생성 및 파급에 주중에 활발하게 일어나고 상대적으로 주말에 적게 나타나는 패턴이 존재함을 밝혔다. 또한, 포스트가 생성된 후 포스트에 대해 생성되는 링크 수가 시간에 따라 어떻게 변화하는가를 측정하여 포스트 생성 후 24시간 내에 대부분의 링크가 생성되고, 시간의 흐름에 따라 생성되는 링크의 수가 감소함을 보였다.

참고 문헌 [2]에서는 거시적인 관점에서 한국 블로그 공간의 구조를 분석하기 위해, 블로그 연결망에서 Bow-tie 구조를 구성하고, 구성된 블로그 연결망의 Bow-tie 컴포넌트 비율이 웹 연결망의 Bow-tie 컴포넌트 비율과 차이가 있음을 밝혔다. 또한, Bow-tie의 IN/OUT 내부 구조를 분석하여 IN/OUT 내부에 작은 Bow-tie 구조가 생성되지 않음을 밝혔다.

III. 블로그 데이터

본 연구에서는 한국 블로그 공간의 특성을 파악하기

위해 한국의 최대 블로그 서비스 중 하나인 네이버 블로그에서 수개월간 수집하여 익명으로 처리한 데이터를 사용하였다. 사용된 데이터는 네이버 블로그의 일부로서 블로거 약 170만 명, 포스트 약 1억 개를 포함하고 있다. 또한, 시간에 따른 변화를 분석하기 위해 전체 데이터를 7개의 시간 구간으로 분할하였다.

기존의 웹 연결망에 관한 연구들은 웹 페이지를 노드로 하고, 블로그 연결망에 관한 연구들은 블로그 또는 포스트를 노드로 하고, 노드들 간의 하이퍼링크를 에지로 하여 웹 연결망과 블로그 연결망을 구성한다. 그러나 한국 블로그 공간의 포스트들은 하이퍼링크를 거의 가지고 있지 않기 때문에 하이퍼링크를 이용하여 블로그 연결망을 구성할 경우, 블로그 공간 전체를 연결하는 연결망을 구성할 수 없다. 하나의 웹 페이지가 다른 웹 페이지에 하이퍼링크를 연결하는 것은 해당 웹 페이지의 내용에 관심이 있다는 것을 의미한다. 한국 블로그 공간에서는 하이퍼링크가 거의 존재하지 않지만, 웹 페이지의 하이퍼링크와 유사하게 관심을 나타내는 것으로 스크랩, 트랙백, 댓글 달기와 같은 액션들이 존재한다. 코멘트는 블로거들의 해당 포스트의 내용에 대한 관심을 표명하기보다는 가벼운 인사의 형태로 사용하는 경우가 많다. 반면, 스크랩과 트랙백은 해당 포스트의 내용에 대한 관심과 지지가 큰 경우에 행하게 되는 액션이다. 그러나 실제로 트랙백은 한국 블로그 공간에서 거의 사용되지 않는 것으로 나타났다. 따라서 본 연구에서는 스크랩을 이용하여 분석을 수행하였다. 즉, 액션의 주체인 블로거를 노드로, 블로거가 행한 스크랩을 에지로 간주하여 두 노드들을 연결한다.

그림 2는 블로그 연결망의 구성 방법을 예로 나타낸 것이다. 왼쪽 그림에서 실선은 해당 블로거가 해당 포스트를 작성했음을 나타내고, 점선은 해당 블로거가 해당 포스트를 스크랩 했음을 나타내는 것이다. 왼쪽 그

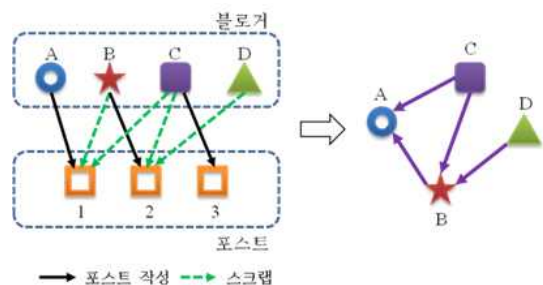


그림 2. 블로그 연결망 구성
Fig. 2. Building a blog network.

림에서 블로거 A가 작성한 포스트를 블로거 B가 스크랩 하였으므로, 블로거 B에서 블로거 A로의 방향성 에지가 생성된 것이다.

IV. 한국 블로그 공간의 특성 변화 분석

본 장에서는 시간의 흐름에 따른 블로그 공간의 특성을 분석한다. 이를 위해, 구간 별 데이터를 이용하여 블로그 공간이 시간에 따라 변화해온 과정을 분석한다. 또한, 한국 블로그 공간 고유의 특성이 이러한 변화에 미치는 영향을 분석한다.

시간의 흐름에 따라 한국 블로그 공간의 Bow-tie 구조가 변화하는 모습을 파악하기 위해 각 구간의 데이터를 누적해가며 시간의 흐름에 따른 Bow-tie 컴포넌트의 노드 수 변화와 상대적 비율 변화를 측정하였다. Bow-tie 구조 전체의 약 90%를 차지하는 핵심적인 요소인 IN, SCC, OUT을 대상으로 실험을 수행하였다. 그림 3(a)는 시간의 흐름에 따른 블로그 연결망의 Bow-tie 컴포넌트의 노드 수 변화를 나타낸 것이다. 가로 축은 각 구간, 세로축은 Bow-tie 구조 각 컴포넌트

내의 노드 수를 나타낸다. 그림 3(b)는 블로그 연결망의 Bow-tie 컴포넌트의 상대적 비율 변화를 나타낸 것이다. 가로 축은 시간의 흐름에 따른 각 구간, 세로축은 Bow-tie 구조에서 각 컴포넌트가 차지하는 비율을 나타낸다.

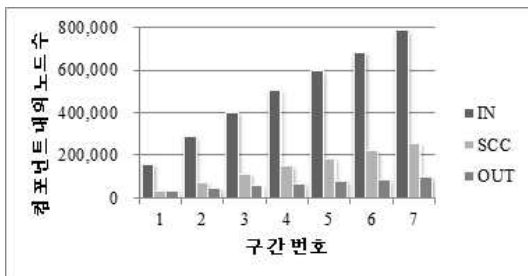
그림 3(a)에서 보이는 것과 같이, 시간의 흐름에 따라 IN 내의 노드 수는 빠르게 증가하고, SCC 내의 노드 수도 점진적으로 증가하고 있으며, OUT 내의 노드 수는 아주 느리게 증가하고 있음을 알 수 있다. IN 내의 노드들은 주로 다른 블로거의 포스트를 스크랩하는 블로거들로, 이는 새롭게 유입된 블로거들의 특성으로 볼 수 있다. 한국 블로그 서비스 업체들이 제공하는 편리한 블로그 생성 및 스크랩 기능이 신규 블로거들이 쉽게 진입할 수 있는 기회를 제공하고 있음을 알 수 있다.

IN 내의 노드들의 수가 가장 빠른 속도로 증가하고 있는 것으로 보이는 것과 달리, Bow-tie 컴포넌트의 비율의 측면에서는 SCC의 비율이 점점 증가하고 있음을 알 수 있다(그림 3(b)). SCC에 속하는 노드들은 블로그 공간에서 서로 영향을 주고받으며 활발하게 활동하는 블로거들을 의미한다. 따라서 블로그 공간이 성장함에 따라 활발하게 활동하는 블로거들의 비율이 높아지고, 일방적으로 스크랩을 하기만 하거나 스크랩을 받기만 하는 블로거의 비율은 상대적으로 줄어드는 것을 알 수 있다. 이는 한국의 블로그 공간이 양적 성장과 더불어 질적 성장도 함께 하고 있음을 보여주는 것이다.

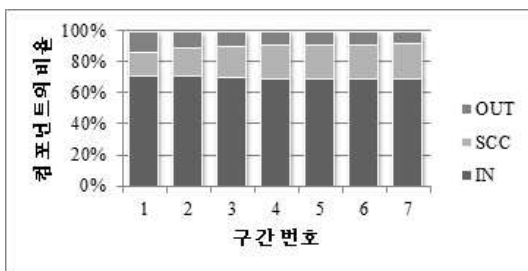
V. 포스트 생성 및 스크랩 경향 분석

한국 블로그 공간에서의 포스트 생성의 전체적 경향을 파악하기 위해 구간 별 전체 포스트 수의 변화를 측정하였다. 새로운 포스트가 생성되는 경향을 관찰하기 위해, 요일 별, 시간대 별 생성되는 평균 포스트의 수를 측정하였다. 그림 4는 요일 별 평균 포스트 생성 수를 나타낸 것으로, 가로축을 요일, 세로축을 각 요일에 생성된 평균 포스트 수로 표시하였다. 실험 결과, 요일 별 평균 포스트 생성 수는 약 19만 건으로 각 요일 별로 거의 차이를 보이지 않았다. 즉, 다른 블로그 공간에서 발견되는 주중, 주말 패턴^[12]이 나타나지 않음을 알 수 있었다.

그림 5는 시간 별 평균 포스트 생성 수를 나타낸 것으로, 가로 축을 시간, 세로축을 각 시간에 생성된 평균 포스트 수로 표시하였다. 실험 결과, 새벽과 아침 시간



(a)



(b)

(a) Bow-tie 컴포넌트 내의 노드 수 변화
 (a) Number of nodes in each component over time.
 (b) Bow-tie 컴포넌트의 비율 변화
 (b) Proportion of each component over time.

그림 3. 시간의 흐름에 따른 블로그 연결망의 Bow-tie 구조 변화

Fig. 3. Bow-tie structure of blog network over time.

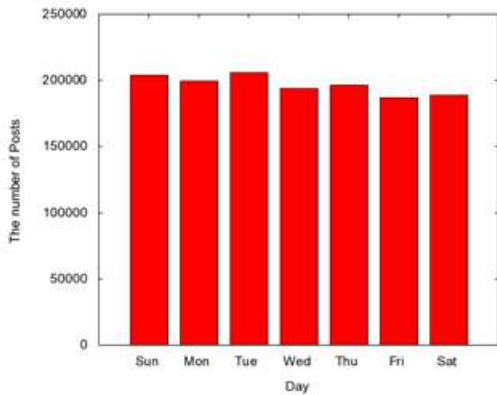


그림 4. 요일 별 포스트 생성 수
Fig. 4. Number of posts in each day of the week.

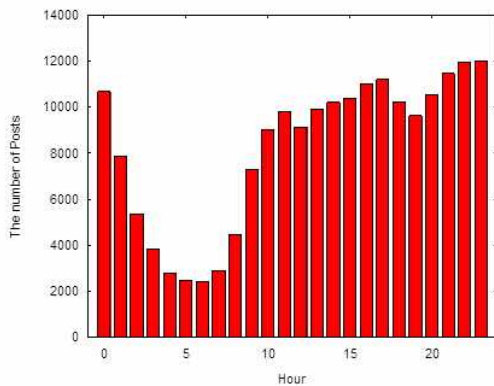


그림 5. 시간 별 평균 포스트 생성 수
Fig. 5. Number of posts in each hour.

대 3시부터 8시 사이에 다른 시간대에 비해 블로그 활동이 거의 없음을 알 수 있다. 또한, 퇴근 시간대인 18시부터 19시 사이에 블로그 활동이 다소 감소하는 경향이 보인다.

참고 문헌 [11, 13~14]는 entropy plot^[15]을 통해 블로그 연결망의 burstiness를 측정하였다. Burstiness란 특정 현상이 어떤 구간에 얼마나 집중적으로 발생하는지를 의미하는 것으로, 해당 현상이 특정 구간에 많이 집중될수록 해당 현상은 burst하다고 표현한다. Entropy plot은 resolution W 에 대한 특정 현상의 entropy $H(W)$ 의 분포로 나타낸다. Resolution이란 전체 구간을 몇 개의 세부 구간으로 나누었는가를 의미하고, Resolution W 에서의 entropy $H(W)$ 는 식 (1)과 같이 정의된다^[15]. 이 때, $p_{i,w}$ 는 전체 구간에 대한 i 구간에서의 현상의 수의 비율을 의미한다. 즉, $p_{i,w} = n_{i,w}/n$.

$$H(W) = - \sum_i p_{i,w} \log_2(p_{i,w}) \quad (1)$$

Entropy plot은 $H(W)$ 를 세로축으로, $\log_2(W)$ 를 가로

축으로 하는 그래프로 정의한다. Entropy plot의 자료들이 선형 함수의 모습으로 그려지는 경우, 해당 현상은 self-similar하다고 정의한다^[15]. 특정 현상이 self-similar하다는 것은 전체 구간에서의 해당 현상의 패턴이 세부 구간에서도 동일하게 나타남을 뜻한다.

또한, entropy plot의 요소들에 대한 선형 함수 회귀선을 그렸을 때, 그 선형 함수의 기울기 β 를 information fractal dimension이라 한다^[15]. 기울기 β 를 통해 해당 현상의 burstiness를 알 수 있다. β 가 1이라면 해당 현상이 uniform하게 일어남을 의미하고, β 가 0이라면 해당 현상이 정확히 동일한 시각에 일어났음을 의미한다.

한국 블로그 공간에서 포스트의 생성의 burstiness를 확인하기 위해 포스트 생성 수의 entropy plot을 구하고 information fractal dimension을 측정하였다. 실험 결과, 그림 6과 같이 포스트 생성 수의 entropy plot은 선형 함수의 형태로 나타났다. 이는 블로그 공간에서의 포스트 생성이 self-similar한 특성을 가짐을 의미한다. 또한, 포스트 생성 수의 information fractal dimension 값은 기존의 블로그 공간의 information fractal dimension 값인 0.88^[11], 0.72^[14]에 비해 0.997로 1에 더 가까운 값을 나타냈는데, 이는 한국 블로그 공간에서의 포스트 생성

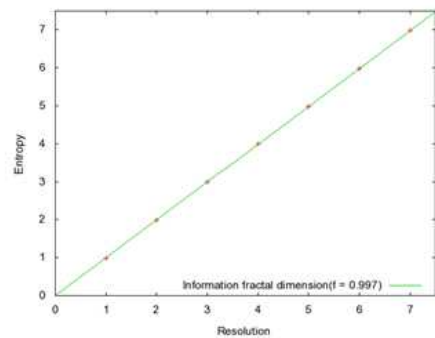


그림 6. 포스트 생성 수의 entropy plot
Fig. 6. Entropy plot of number of posts.

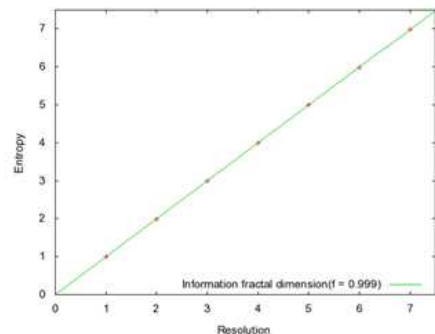


그림 7. 스크랩 수의 entropy plot
Fig. 7. Entropy plot of number of scraps.

이 다른 블로그 공간에서의 포스트 생성보다 더 uniform 하게 나타남을 의미한다.

포스트의 생성이 uniform하게 일어난다는 것은 일별 생성 포스트 수가 서로 유사함을 의미한다. 이는 사회적 이슈 및 현상과 관계없이 포스트의 생성 수가 거의 일정함을 뜻한다. 이는 한국 블로그 공간의 블로거들은 자신이 관심 있는 고유한 주제에 대한 연재 형식으로 포스트를 생성하거나 일상에 관한 내용으로 포스트를 생성하는 것으로 생각할 수 있다. 이는 곧 한국 블로그 공간에서의 블로거들의 활동이 외부적 요인으로 인해 크게 변동하지 않고 안정되어 있음을 의미한다.

그림 7은 한국 블로그 공간에서의 스크랩 발생의 entropy plot과 information fractal dimension을 측정함으로써, 스크랩 발생도 포스트의 생성과 마찬가지로 self-similar한 특성과 uniform하게 일어나는 특성을 가지고 있음을 알 수 있다.

VI. 결 론

본 논문에서는 한국 블로그 공간의 시간의 흐름에 따른 특성 변화에 대한 심도 있는 분석을 수행하였다. 분석을 위해 블로그를 노드로, 블로거가 다른 블로거의 포스트에 관심의 의미로 행한 스크랩을 에지로 블로그 연결망을 구성하였다.

첫째, 시간의 흐름에 따른 블로그 공간의 특성 변화를 분석하였다. 시간의 흐름에 따른 블로그 연결망의 Bow-tie 컴포넌트의 노드 수 및 비율의 변화, average connected distance 변화를 분석하여 블로그 연결망의 연결성이 시간의 흐름에 따라 점점 높아짐을 보였다. 이는 기존 논문에서 사회연결망에서 시간의 흐름에 따라 치밀화(densification)가 발생한다는 주장과 부합한다^[16]. 블로그 연결망의 연결성이 높아지는 데에는 한국 블로그 서비스 업체에서 제공하는 스크랩기능, 검색 기능, 블로그 서비스의 초기 화면에 인기 포스트나 양질의 포스트를 노출시키는 기능 등 블로그 서비스를 편리하게 사용할 수 있는 환경적 요소가 영향을 미친 것으로 생각할 수 있다.

둘째, 블로그 공간에서의 포스트 생성 및 스크랩 경향을 분석하였다. 구간 별 포스트 생성 패턴과 스크랩 발생 패턴을 분석하여 다른 블로그 공간에서 발견되는 주중, 주말 패턴이 나타나지 않음을 밝혔다. 또한, 시간대 별로 블로거들의 활동량에 차이가 나타남을 밝혔다.

셋째, 블로그 공간에서의 포스트 생성 및 스크랩 경향을 분석하였다. 구간 별 포스트 생성 패턴과 스크랩 발생 패턴을 분석하고, 이들의 entropy plot을 구하였다. 이를 통해 한국 블로그 공간에서의 포스트 생성과 스크랩 발생은 기존의 연구들에서 분석 대상으로 삼은 블로그 공간에서의 포스트 생성과 링크 생성에 비해 더 uniform함을 밝혔다. 이는 특별한 사회적 이슈나 이벤트가 있는 경우가 아니더라도 블로거들이 매우 활발하게 블로그 공간을 사용하고 있음을 의미한다.

본 연구의 결과가 한국 블로그 공간의 고유한 특성을 이해하고, 이를 이용한 효과적인 알고리즘들을 개발하는데 도움이 될 것으로 기대한다. 또한 한국 블로거들의 성향을 이용하여 성공적인 비즈니스 정책을 수립하는데 도움이 될 것으로 기대한다.

참 고 문 헌

- [1] S. Herring et al., "Conversations in the Blogosphere: An Analysis "From the Bottom Up"," In *Proc. the 38th Annual Hawaii Int'l. Conf. on System Sciences*, pp. 107b-107b, January 2005.
- [2] J. Ha et al., "Analyzing a Korean Blogosphere: A Social Network Analysis Perspective," In *Proc. ACM Symposium on Applied Computing*, March 2011.
- [3] Y. Kwon et al., "The Information Diffusion Model in the Blog World," In *Proc. ACM KDD Workshop on Social Network Mining and Analysis*, June 2009.
- [4] Y. Kwon, S. Kim, and S. Park, "An Analysis of Information Diffusion in the Blog World," In *Proc. ACM CIKM Workshop on Complex Networks in Information and Knowledge Management*, pp. 27-30, November 2009.
- [5] S. Lim et al., "Determining Content Power Users in a Blog Network," In *Proc. ACM KDD Workshop on Social Network Mining and Analysis*, June 2009.
- [6] S. Lim et al., "Construction of a Blog Network Based on Information Diffusion," In *Proc. ACM Symposium on Applied Computing*, March 2011.
- [7] 송석순, 윤석호, 김상욱, "블로그 공간에서의 링크 기반 클러스터링 방안," *대한전자공학회논문지-CI*, 제46권, 3호, 42-49쪽, 2009년 5월
- [8] R. Albert, H. Jeong, and A. Barabasi, "Diameter of the World Wide Web," *Nature*, Vol. 401, pp. 130-131, 1999.

[9] A. Broder et al., "Graph Structure in the Web," *Computer Networks*, Vol. 33, No. 1, pp. 309-320, 2000.

[10] D. Donato et al., "The Web as a Graph: How Far We Are," *ACM Trans. on Internet Technology*, Vol. 7, No. 1, pp. 4-28, 2007.

[11] M. Gotz et al., "Modeling Blog Dynamics," In *Proc. Int'l. Conf on Weblogs and Social Media*, May 2009.

[12] J. Leskovec et al., "Cascading Behavior in Large Blog Graphs: Patterns and a model," In *Proc. SIAM Int'l. Conf on Data Mining*, April 2007.

[13] M. McGlohon, L. Akoglu, and C. Faloutsos, "Weighted Graphs and Disconnected Components: Patterns and a Generator," In *Proc. ACM Int'l. Conf on Knowledge Discovery and Data Mining*, pp. 524-532, August 2008.

[14] M. McGlohon et al., "Finding Patterns in Blog Shapes and Blog Evolution," In *Proc. Int'l Conf on Weblogs and Social Media*, March 2007.

[15] M. Wang et al., "Data Mining Meets Performance Evaluation: Fast Algorithms for Modeling Bursty Traffic," In *Proc. IEEE Int'l Conf on Data Mining*, pp.507-516, February 2002.

[16] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations," In *Proc. ACM SIGKDD Int'l Conf on Knowledge Discovery in Data Mining*, pp. 187, August 2005.

— 저 자 소 개 —



하 지 운(학생회원)
2008년 한양대학교 정보통신대학
정보통신학부 컴퓨터전공
학사 졸업.
2010년 한양대학교 전자컴퓨터
통신공학과 석사 졸업.

2010년~현재 한양대학교 전자컴퓨터통신공학과
박사과정 재학 중.
<주관심분야: 사회연결망분석, 인터넷 포탈 데이
터 분석, e-비즈니스, 데이터 마이닝>



배 덕 호(학생회원)
2006년 한양대학교 정보통신대학
정보통신학부 컴퓨터전공
학사 졸업.
2008년 한양대학교 전자컴퓨터통
신공학과 석사 졸업.
2008년~현재 한양대학교 전자컴
퓨터통신공학과 박사과정
재학 중.

2010년~2010년 미국 Carnegie Mellon
University, Computer Science
Department, Visiting Scholar.
<주관심분야: 사회연결망분석, 인터넷 포탈 데이
터 분석, 논문 데이터 분석, e-비즈니스, 데이터
마이닝, 임베디드 DBMS, 플래시 메모리 DBMS,
이동 객체 DBMS>



김 상 옥(평생회원)
1989년 서울대학교 컴퓨터공학과
학사 졸업.
1991년 한국과학기술원 전산학과
석사 졸업.
1994년 한국과학기술원 전산학과
박사 졸업.

1991년~1991년 미국 Stanford University,
Computer Science Department,
방문 연구원.
1994년~1995년 KAIST 정보전자 연구소
전문 연구원.
1999년~2000년 미국 IBM T.J. Watson Research
Center, Post-Doc.
1995년~2003년 강원대학교 정보통신공학과
부교수.
2003년~현재 한양대학교 정보통신대학
정보통신학부 교수.
2009년~2010년 미국 Carnegie Mellon
University, Visiting Scholar.
<주관심분야 : 데이터베이스 시스템, 저장 시스
템, 트랜잭션 관리, 데이터 마이닝, 멀티미디어 정
보 검색, 공간 데이터베이스/GIS, 주기억장치 데
이터베이스, 이동 객체 데이터베이스/텔레매틱스,
사회 연결망 분석, 웹 데이터 분석>