

블로그 공간의 내부 구조 분석

하지운*, 배덕호*, 김상욱*, 백석철**

*한양대학교 전자컴퓨터통신공학과

** NHN(주)

e-mail: oneofus@agape.hanyang.ac.kr

Analyzing the Inner Structure of the Blogosphere

Jiwoon Ha*, Duck-Ho Bae*, Sang-Wook Kim*, Seok-Chul Baek**

*Dept. of Electronics and Computer Engineering, Hanyang University

** NHN Corp.

요 약

최근, 블로그 이용자의 수가 크게 증가함에 따라 블로그 공간에 대한 분석의 중요성이 점차 증가하고 있다. 본 논문에서는 블로그 공간의 보우타이 구조 분석을 통해 블로그 공간의 내부 구조의 특성을 파악한다. 더 나아가, 웹 공간과의 비교를 통해 블로그 공간만의 고유의 특성을 도출한다.

1. 서론

블로그는 사용자가 자신의 글을 온라인상에 저장할 수 있는 개인 홈페이지의 일종이다. 블로거는 블로그의 이용자를 의미한다. 포스트는 블로거가 작성하여 블로그에 저장한 글을 의미한다. 블로거는 다른 블로거가 작성한 포스트를 대상으로 여러 가지 행동들을 취할 수 있는데, 이러한 행동들은 스크랩, 트랙백, 댓글이 존재한다[1]. 본 논문에서는 블로거들과 포스트들, 그리고 작성된 포스트에 대한 행동들로 구성된 공간을 블로그 공간(blogosphere)이라 한다[2].

최근, 블로그 공간의 규모가 커짐에 따라 블로그 공간에 대한 분석의 중요성 또한 증가하였다. 그 동안 웹 공간에 대한 분석은 많이 이루어진 반면, 블로그 공간에 대한 분석은 거의 이루어지지 않았다. 블로그 공간에 적합한 응용 알고리즘 개발을 위해서는 블로그 공간의 내부구조에 대한 심도 있는 분석이 선행되어야 한다.

본 논문에서는 블로그 공간의 보우타이 구조(bow-tie structure)[3]에 대해 분석한다. 더 나아가, 웹 공간과의 특성 비교를 통해 블로그 공간만의 고유의 특성을 도출한다.

2. 보우타이 구조 (Bow-tie Structure)

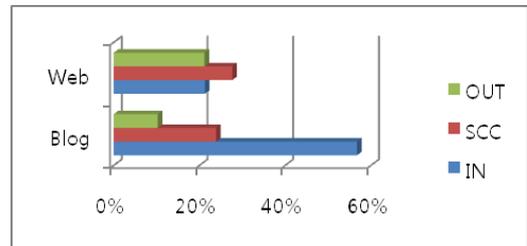
참고 문헌 [3]에서는 웹 공간을 거시적인 관점에서 나타내는 모델로 보우타이 구조(bow-tie structure)를 제안하였다. 참고 문헌 [3]에 따르면 웹 공간의 보우타이 구조는 SCC가 전체 노드의 약 28%를 차지하고, IN, OUT, Tendrils가 각각 전체 노드의 약 21%를 차지한다.

본 논문에서는 국내에서 가장 큰 블로그 서비스에서 2006년 4월부터 수개월간 수집하여 익명으로 처리한 블로그 공간 데이터를 대상으로 보우타이 구조를 구성하였다. 그림 1은 블로그 공간과 웹 공간의 보우타이

구조 각 컴포넌트의 비율을 나타낸다. 그 결과, 웹 공간의 보우타이 구조는 IN, SCC, OUT의 비율이 서로 비슷한 반면, 블로그 공간의 보우타이 구조는 IN이 OUT에

비해 약 5.5배가 크게 나타났다.

블로그 공간에서 보우타이의 각 컴포넌트의 성향은 다음과 같이 사료된다. OUT에 속하는 블로거들의 대부분은 다른 블로거의 포스트에 행동을 취하기보다 직접 포스트를 작성하는 경향을 가지며, IN에 속하는 블로거들의 대부분은 포스트를 작성하기보다는 다른 블로거들이 작성한 포스트에 행동을 취하는 경향을 갖는다. SCC에 속하는 블로거들은 자신이 포스트를 작성하기도 하며, 다른 블로거들이 작성한 포스트에 행동을 취하기도 하는 경향을 갖는다. 블로그 공간은 웹에 비해 자신에게 필요한 정보를 담아내기 위한 목적을 갖는 블로거들의 비중이 높으므로, 기본적으로 다른 블로거의 포스트에 행동을 취하는 블로거들이 많으며, 이로 인해 IN의 비중이 커진 것으로 생각된다. 이와 더불어 블로그 서비스 업체에서 제공하는 편리한 스크랩 기능 또한 IN의 비중이 커지는데 영향을 준 것으로 보인다[4].



(그림 1) 웹 공간[3]과 블로그 공간의 보우타이 구조.

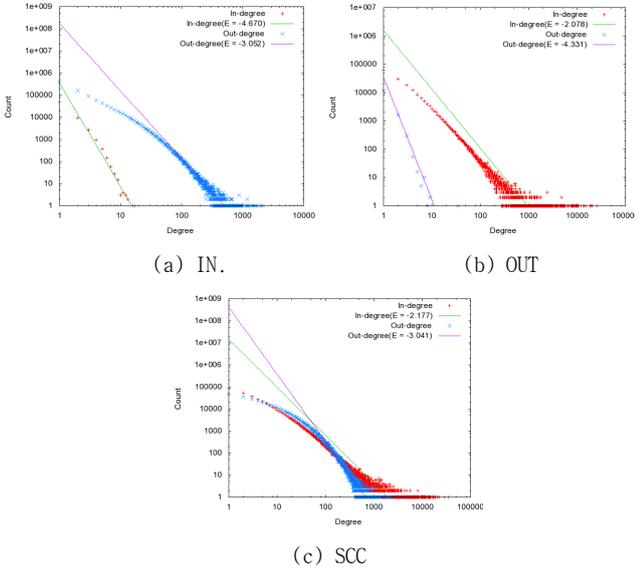
보우타이 구조 각 컴포넌트를 구성하는 블로거들의 성향을 고려하여 마케팅 등의 비즈니스 전략을 수립할 수 있다. 특정 컴포넌트에 속한 블로거들을 대상으로 개인화된 전략을 적용하면 전체 블로그 공간의 블로거들에 동일한 비즈니스 전략을 적용하는 것보다 더 높은 효과를 기대할 수 있다.

3. 컴포넌트 별 degree 분포

각 컴포넌트에 속한 블로거들의 행동 성향을 파악하기 위해 각 컴포넌트의 degree 분포를 측정한다. 그림 2는 각

컴포넌트의 degree 분포를 나타낸다.

IN의 블로거들은 주로 다른 블로거의 포스트를 스크랩하고, OUT의 블로거들은 다른 블로거들로부터 스크랩을 많이 받음을 알 수 있다. 또한, SCC의 블로거들은 스크랩을 하는 것과 받는 것이 비슷한 비율임을 알 수 있다.



(그림 2) IN, OUT, SCC의 degree 분포.

4. IN, OUT 의 내부 구조

참고 문헌 [5]에서는 IN, OUT의 내부에 보우타이 구조가 재귀적으로 나타나는지 분석하였다. 보우타이 구조의 각 컴포넌트 내부에 보우타이 구조가 존재하기 위해서는 충분한 규모의 SCC, WCC가 존재해야 한다.

본 논문에서는 블로그 공간의 IN, OUT 내부의 SCC의 크기, WCC 개수와 크기를 측정하고, 이를 웹 공간과 비교하였다. 표 1은 측정한 결과를 나타낸다. 표 1에서 보는 것과 같이 블로그 공간도 웹 공간과 마찬가지로 IN, OUT 내부에서 보우타이 구조가 존재하기 위한 충분한 크기의 SCC, WCC가 존재하지 않음을 알 수 있다.

<표 1> IN, OUT의 내부 구조

	Web [5]		Blog	
IN				
nodes in IN	14,400,000	11%	962,019	56.80%
max SCC	5,876		2	
number of WCCs	3,680,000		40,163	
max WCC	197,500	1.30%	766	0.08%
OUT				
nodes in OUT	53,300,000	39%	176,073	10.39%
max SCC	9,349		7	
number of WCCs	25,400,000		5,764	
max WCC	14,940,000	28.01%	5,955	3.38%

반면, 웹 공간과 블로그 공간 간의 차이점도 발견할 수 있다. 블로그 공간이 웹 공간에 비해 전체에서 IN이 차지하는 비율이 매우 큼에도 불구하고 IN 내부에서 가장 큰 WCC(max WCC)가 매우 작고, OUT 또한 내부에서 가장 큰 WCC의 크기가 웹 공간에 비해 매우 작음을 알 수 있다. 블로그 공간은 웹 공간에 비해 IN과 OUT 내부 구성원 간의

활발한 소통은 이루어지지 않음을 추론할 수 있다.

5. SCC의 내부 구조

SCC의 노드 중 IN, OUT과 특정 관계를 갖고 있는 노드를 통해 SCC의 내부 구조를 파악할 수 있다. 참고 문헌 [5]에서는 entry points, exit points, bridges, connectors, petals의 다섯 가지 노드를 통해 SCC의 내부 구조를 분석하였다. Entry points는 IN의 노드 중 한 개 이상의 노드들로부터 링크를 받은 노드를 뜻한다. Exit points는 OUT의 노드들 중 한 개 이상의 노드를 링크한 노드를 뜻한다. Bridges는 entry points와 exit points의 역할을 모두 하는 노드를 뜻한다. Connectors는 SCC의 노드 중 받은 링크와 한 링크가 각각 한 개인 노드를 뜻한다. Connectors 중 링크를 한 노드와 링크를 받은 상대 노드가 같은 노드인 경우를 petals라고 한다.

본 논문에서는 블로그 공간의 SCC의 내부 구조를 파악하고, 이를 웹 공간과 비교하였다. 표 2는 측정한 결과를 나타낸다.

<표 2> SCC의 내부 구조

	Web [5]		Blog	
SCC				
nodes in SCC	44,700,000	33%	427,101	24.07%
entry points	2,600,000	5.87%	285,671	66.89%
exit points	29,600,000	72.03%	292,050	68.38%
bridges	2,000,000	4.58%	204,127	47.79%
connectors	2,960,000	6.63%	14,213	3.33%
petals	1,400,000	3.14%	940	0.22%

표 2로부터 블로그 공간의 entry points, exit points, bridges의 비율이 웹 공간에 비해 매우 높음을 알 수 있다. 이는 SCC의 대부분의 노드들이 IN, OUT과 직접적으로 연결되어 있음을 나타낸다. 검색, 블로그 포털 등의 외적인 요인으로 인해 SCC과 IN, OUT의 연결성이 높아진 것으로 추론할 수 있다.

6. 결론

본 논문에서는 블로그 공간의 보우타이의 각 컴포넌트 내부의 특성 분석을 통해 블로그 공간의 특성을 파악하였다. 더 나아가, 이를 웹 공간의 특성과 비교하여 두 공간의 공통점과 차이점을 밝혔다.

감사의 글

본 연구는 NHN(주)의 지원을 받았습니다. 그러나, 본 논문에서 제시된 의견이나 결론, 또는 권고 등은 온전히 저자(들)의 것이며, 반드시 지원회사의 입장을 대변하는 것은 아닙니다.

참고문헌

[1] NHN(주), 블로그 홈, <http://blog.naver.com>.
 [2] S. Herring et al., "Conversation in the Blogosphere: An Analysis "From the Bottom Up",," In *Proc. of the 38th Annual Hawaii Int'l. Conf. on System Sciences*, pp. 107b, 2005.
 [3] A. Broder et al., "Graph Structure in the Web," *Computer Networks*, Vol. 33, No. 1, pp. 309-320, 2000.
 [4] J. Ha, D. Bae, and S. Kim, "Analyzing Characteristics of a Blog Network," In *Proc. of Korea Computer Congress*, Vol. 36, No. 1(C), pp. 171-174, 2009.
 [5] D. Donato et al., "Mining the Inner Structure of the Web Graph," *Journal of Physics A: Mathematical and Theoretical*, Vol. 41, pp. 12-17, 2008.