

네트워크 분석을 통한 온라인 임신 Q&A 게시판의 내용적 특성에 관한 연구

A Study on the contents property of online pregnant electronic bulletin boards by using network analysis

권선영, 성균관대학교 문헌정보학과, freesson@naver.com
주서현, 성균관대학교 문헌정보학과, knockthebook@gmail.com

Sun young Kwon, Department of Library and Information Science, Sungkyunkwan University
Ju seo hyeon, Department of Library and Information Science, Sungkyunkwan University

국내에서 온라인 커뮤니티는 1990년대 말 인터넷과 개인 PC의 급속한 보급으로 인해 대중화되기 시작했다. 커뮤니티는 이용자 간 활발한 커뮤니케이션을 기반으로 구축되고 있으며 이러한 커뮤니케이션은 일상적인 언어적 맥락을 바탕으로 구성된다. 이러한 커뮤니티의 내용 분석은 연구자의 직관에 의해 수행되기 마련이나 연구대상의 양이 많아질 경우 연구자의 직관만으로는 이를 분석하는데 제약이 있을 수 있다. 네트워크 분석은 이와 같은 문제점을 해결 할 수 있는 방법 중 하나이다. 본 연구는 온라인 임신 Q&A 게시판의 내용적 특성을 살펴보기 위하여 네트워크 분석을 수행하였다. 먼저 수집한 데이터의 전체적인 내용적 특성을 살펴본 후, 게시물 작성자인 임신부의 임신 시기에 따른 게시글의 내용적 특성을 살펴보았다. 분석 결과 임신시기별로 핵심적인 단어의 빈도의 차이가 있었으며 네트워크의 구성 또한 다른 것을 확인하였다. 또한 성별에 따라서도 그 내용적 특성이 다른 것을 살펴 볼 수 있었다.

1. 서론

과거와 비교해 봤을 때, 오늘날의 데이터는 기하급수적으로 그 양이 증가하고 있기 때문에 이를 연구자의 직관만으로 분석하기에는 어려움이 따른다. 이와 같은 문제점을 위한 방안으로 다양한 분석 방법이 고안되었는데 네트워크 분석 또한 그 중 하나이다. 본 연구에서 연구대상으로 선정한 온라인 커뮤니티는 임신 Q&A 게시판을 포함한 육아정보 커뮤니티로써 육아, 임신 등 다양한 정보가 커뮤니티 구성원들에 의해 생성되고 공유되고 있다.

인터넷과 개인 PC의 보급은 이전 임신부들이 임신과 관련한 정보를 담당 의사 또는 전문 서적에 의존해야 했던 것과 달리 시간과

장소에 구애받지 않고 즉각적으로 필요한 정보를 획득할 수 있는 환경을 조성하였다. 그렇기 때문에 온라인상에서의 임신과 관련한 Q&A를 분석해 보는 것은 오늘날 임신부들이 갖는 질의와 관련한 일반적인 내용을 살필 수 있을 뿐만 아니라 더 나아가 이를 세분화하여 임신 시기에 따라 나타나는 내용적 특성의 차이와 태아의 성별에 따른 내용적 특성 또한 살필 수 있다.

2. 선행 연구

2.1 온라인 커뮤니티

온라인 커뮤니티란 인터넷 상에서 각종 모

임과 단체를 만들 수 있도록 지원하는 서비스(홍일여 2012)로 국내에서 온라인 커뮤니티는 1990년대 말 인터넷과 개인 PC의 급속한 보급이 이루어지면서 대중화되기 시작되었다. 이러한 온라인 커뮤니티의 활성화는 연구자들에 의해 다방면으로 연구되기 시작했으며 더 나아가 내용 분석 및 이용자의 정보탐색, 이용 동기, 상호작용 등과 같이 다방면으로 연구가 진행되어져 온 바 있다.

온라인 커뮤니티는 전통적인 커뮤니티와 특성적 차이를 갖는다. 전통적 커뮤니티는 공적인 동류의식을 지니며 주로 지리적 기반을 중심으로 면대면 상호작용이 이루어지는 특징이 있는데 반해, 온라인 커뮤니티는 사적인 이해나 관심을 중심으로 지리적 한계가 없으며 주로 CMC(computer-mediated communication)를 통해 상호작용을 한다(나정호 2007). 또한 온라인 커뮤니티는 전통적인 커뮤니티와 비교해 볼 때 시간과 공간의 제약을 넘어서 가상적인 커뮤니티(김여진 2001)라는 점에서도 그 차이를 살필 수 있다.

임신·육아와 관련한 커뮤니케이션에 대한 연구로는 온라인 초등교사 커뮤니티의 분석을 통해 초등학교 여교사의 임신, 출산 및 육아 문제를 살펴본 이승민(2007)의 연구가 있다. 이승민(2007)은 임신과 관련하여 인식하는 문제점을 임신 중에 생기는 학교 사정 문제, 관리자나 동료 교사와의 관계 문제, 임신 중 학부모와의 관계로 나눈 바 있다. 이 연구는 임신과 관련한 환경, 상황적 커뮤니케이션을 살펴본다는 점에서 주목할 만하다.

2.2 네트워크 분석

본 연구는 임신과 관련한 온라인 커뮤니티의 내용적 특성을 살펴보기 위하여 용어의 네트워크 분석을 활용하였다. 언어 네트워크 분석은 구성요소들 간의 관계를 찾아내 시스템

의 구조를 파악하는 사회네트워크분석을 언어 분석에 적용한 방법론으로 언어 간의 관계를 파악하는 데 유용(장정우 외 2012)하다.

박한우(2004) 등은 특정 개념에 대한 언어 정보는 그것이 음성언어이든 문자언어이든지간에 연결 네트워크(associative network)를 통해 인지구조 내에 구조화 시켜 놓은 의미에 대한 표상이며 결국 언어 네트워크 분석을 수행한다는 것은 집단 내의 개인처럼 인지구조라는 공간 내에서 벌어지는 인지 요소들 간의 사회 시스템 구조연결이라고 유추해 볼 수 있다고 언급하였다. 더불어 네트워크 분석은 특정 학문이나 대상의 경향을 분석하는 데 사용되기도 한다. 일반적으로 내용 분석에 근거한 텍스트 분석과 네트워크 텍스트 분석의 가장 큰 차이점은 전통적인 내용 분석에 근거할 때는 사전에 설정한 범주에 해당하는 연구 주제의 출현빈도를 분석하는 데에 초점을 두게 되는데 반해 네트워크 텍스트 분석은 특정 단어와 동시에 출현하는 단어가 무엇인가에 보다 무게 중심을 둔다는 점을 들 수 있다. 이러한 방법은 연구동향 분석에 있어 단순히 단어빈도만이 아니라 단어 간의 관계성을 파악한다는 점에서 특징을 가지고 있다고 할 수 있다(최영출 외 2011).

이세진 등(2012)은 인터넷 여성 커뮤니티의 네트워크 분석을 시도한 바 있는데, 내용 분석을 통한 이용자 간의 커뮤니케이션의 내용적 특성을 살펴보고 네트워크 분석을 통해서 커뮤니티 이용자들의 조직화되어 커뮤니티를 지속시키는 지의 구조적 측면을 살펴본 바 있다.

기존의 연구를 살펴보면 언어 네트워크 분석은 학문이나 대상의 경향을 분석 뿐 아니라 특정 언어 네트워크의 내용을 핵심어나 행위자등의 측면에서 다각도로 분석하는 등 다양한 시도가 이루어져 온 것을 알 수 있다.

3. 연구 방법

3.1 연구 방법

본 연구는 임신과 관련된 온라인 커뮤니티 Q&A 게시판의 내용적 특성이 어떠한지를 알아보는 것을 목적으로 한다. 이를 위해 특정 포털의 임신·육아 관련 커뮤니티들 가운데 가장 활발한 커뮤니케이션이 발생하고 있는 것으로 판단되는 한 커뮤니티를 선정한 후, 커뮤니티 내에서도 연구 주제와 밀접한 관련이 있는 임신 Q&A 게시판을 연구 대상으로 선정하였다. 이 게시판에는 하루에 평균적으로 500건에서 600건의 게시물이 업로드 되는데 유사한 형태의 질문, 내용을 분석하기에는 단문 형태의 질문, 과도한 인터넷 용어가 사용된 질문 등이 제외된 2일간의 데이터 1,352건을 대상으로 한다. 데이터를 선정한 후에는 핵심 단어를 추출하고 이들을 네트워크상에서 각 노드의 연결정도중심성, 근접중심성, 매개중심성, 위세중심성 측정하고 단어 간의 관계를 분석하여 임신 Q&A 게시판의 내용을 분석하였다.

3.2 자료 분석

본 연구에서는 내용적 특성을 살펴보기 위하여 동시출현 단어분석(co-word analysis)을 사용하였다. 동시 출현 단어분석은 용어 간의 관계를 살펴보기 위하여 특정 텍스트에서 단어 또는 구가 동시에 출현하는 패턴을 이용하였다. 이 방법은 용어가 특정 텍스트의 내용을 적절하게 표현하고 있거나, 어떤 용어 간의 연결에 의해 텍스트가 구성되어진다는 가정이 전제된다(Cambrosio et al 1993). 하지만 동시출현 단어분석을 하기 위해서는 용어의 정제가 필수적이다. 인터넷 상의 용어들은 학술용어와는 달리 정제되지 않은 상태인

경우가 대부분이며 이 경우 정확한 분석을 하기 어렵다. 예를 들어 ‘아이’란 용어는 ‘아기’, ‘애기’, ‘아이’, ‘애’로 제작기 표현되고 있으므로 이를 ‘아이’로 통일하여 분석하였다.

분석을 위하여 KrKwic 및 형태소분석기인 KLT2010을 사용하였으며, 네트워크 시각화 도구로는 Ucinet에 포함된 netdraw와 군집 시각화 도구로는 vosviewer를 사용하였다.

4. 분석 결과

4.1 핵심어 출현빈도 분석

동시출현 단어분석(co-word analysis)을 수행하기 위해서는 먼저 분석대상에서 핵심어를 추출하여야 한다. 분석대상인 총 1,352건의 게시물에서 추출된 단어는 42,214개였으며 이중 중복, 불용어, 오타자 등을 제외한 후 빈도를 측정하고 임신시기별로 출현빈도를 살펴 보았다(<표 1> 참조).

<표 1> 단어의 전체 출현빈도

단어	전체출현빈도	시기별 출현빈도			
		임신전	임신초기	임신중기	임신후기
아이	435	0	11	75	127
병원	351	67	156	52	56
입덧	179	5	135	23	0
걱정	169	14	69	39	0
들깨	135	0	43	16	51
생리	134	108	18	0	0
검사	117	16	48	42	9
증상	105	36	41	12	11
아기집	94	24	0	0	0
태동	83	0	0	51	28

용어를 살펴보면 각 시기에 따라 증가하는 용어, 감소하는 용어, 지속적으로 출현하는 용어 등이 있음을 알 수 있다. 이처럼 시기에 따른 용어의 출현패턴(출현빈도에 근거)은 시기별로 핵심어가 존재할 것을 예상해볼 수 있다. 이에 따라 분석을 위해 각 시기별로 게시물을 나누고 분석 대상 용어를 추출하였으며

그 대상용어 수는 아래와 같다(<표 2> 참조).

<표 2> 임신기간별 게시물 및 단어의 수

기간	게시물수	Term total	Term Unique	Term Analysis
임신전	355	7,224	3,295	97
임신초기	417	14,269	5,566	109
임신중기	231	7,317	3,501	100
임신후기	275	8,536	4,211	99
임신전체	1,352	42,214	16,573	200

4.2 전체 네트워크 분석

빈도 분석을 통해 임신 Q&A 게시물의 주요단어 가운데 출현빈도가 10회 이상인 단어들을 추출하였다. 그런 다음 연구목적과 관련된 주요단어 200개를 선정한 후 주요단어 간의 네트워크 분석을 수행하였다. 200개의 노드는 총 29,792의 링크로 연결되며 이에 대한 중심성 분석 결과는 다음과 같다(<표 3> 참조).

<표 3> 단어의 중심성 분석

단어	연결정도 중심성	근접 중심성	매개 중심성	위세 중심성
아이	8,575	95.960	159.370	0.386
병원	7,823	95.707	145.589	0.378
입덧	2,318	86.111	49.049	0.107
걱정	2,899	93.687	93.879	0.167
둘째	2,394	91.919	84.224	0.134

네트워크 집중도 = 11.966%

연결정도 중심성(degree Centrality)은 노드(단어)와 직접 연결되는 링크가 많을수록 높아진다. 분석 결과 ‘아이’, ‘병원’의 경우 다른 단어에 비해 상대적으로 연결정도 중심성의 값이 높았다. 이 단어들은 매개중심성, 위세중심성의 값에서도 상대적으로 높은 차이를 나타내는 반면, 근접중심성 결과는 큰 차이를 보이고 있지 않는 것을 볼 수 있다. 근접 중심성(closeness Centrality)은 높을수록 네트

워크 내에서 핵심단어와 가장 가까운 거리에 위치한다고 볼 수 있는데, 전체 네트워크에서의 출현빈도 10회 이상인 단어를 통해 분석을 수행할 경우 대체로 거의 모든 단어가 중심성이 높은 것으로 나타났다.

4.3 임신시기별 네트워크 분석

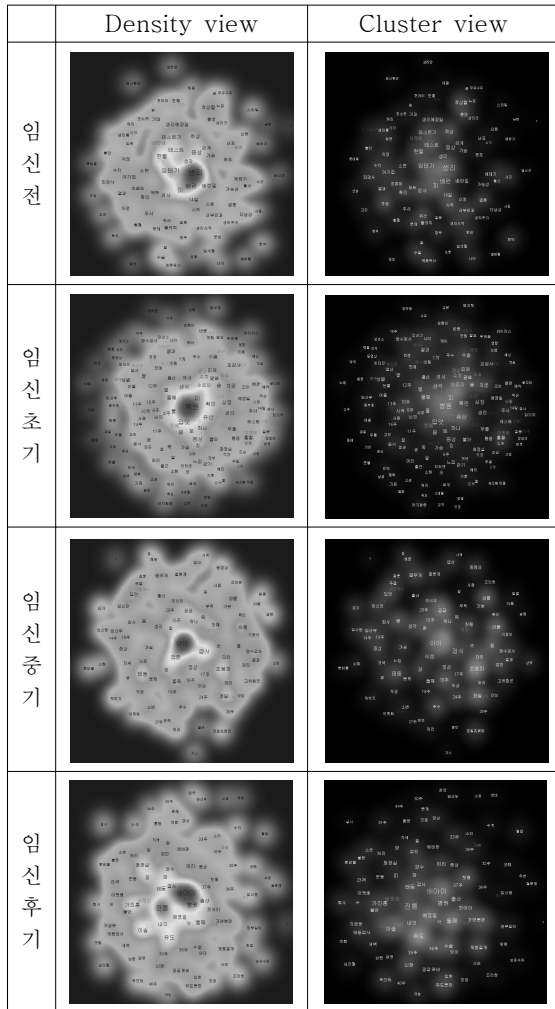
본 연구에서는 각 임신시기별로 게시물의 내용적 특성이 다를 것을 가정하고 이를 검증하기 위해 네트워크 분석을 수행하였다. 임신시기별로 연결정도 중심성 값을 기준으로 상위어들을 살펴보면 각 시기별로 핵심어들의 중심성 값은 차이가 있음을 알 수 있다.

<표 4> 임신시기별 핵심어 중심성 분석

시기	단어	연결정도 중심성	근접 중심성	매개 중심성	위세 중심성
임신전	생리	3,393	96.97	94.085	0.485
	병원	1,612	96.97	113.428	0.277
	입덧기	1,603	96	88.007	0.264
임신초기	생리	3,401	97	97.896	0.485
	병원	1,617	97	115.975	0.277
	배란	1,573	95.098	92.699	0.269
임신중기	아이	1,294	98.020	170.040	0.380
	병원	1,103	95.192	158.430	0.315
	태동	1,062	67.808	22.562	0.197
임신후기	아이	2,814	100	141.678	0.405
	진통	2,604	87.500	49.038	0.443
	배	2,492	95.154	89.732	0.219

4.4 네트워크 시각화를 통한 임신시기별 내용 분석

본 연구를 위한 네트워크분석 결과처럼 링크의 수가 많은 경우 netdraw와 같은 프로그램에서는 관계를 쉽게 살펴보기 힘든 점이 있으므로 밀도의 정도(density)를 비교적 쉽게 살펴볼 수 있는 vosviewer를 통해 살펴보도록 하였다. 각 시기별의 특성은 <그림 1>과 같다.

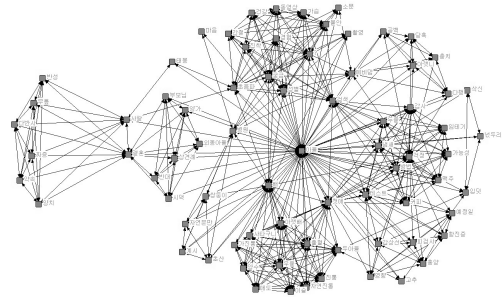


<그림 1> 임신 시기별
Density & Cluster view

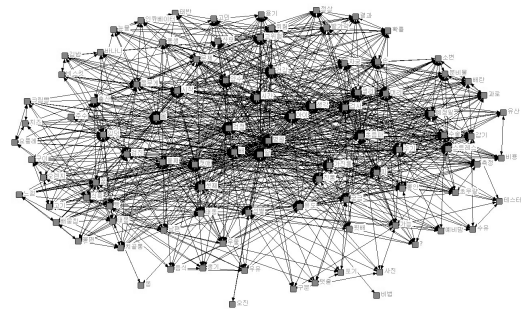
각 임신 시기 모두 핵심어를 중심으로 밀집되어 있는 것을 살펴볼 수 있는데, 이는 각 시기별로 게시물의 특정 주제를 중심으로 통해 활발한 커뮤니케이션이 발생할 때 나타나는 현상이라고 볼 수 있다. 각 시기별로 그 내용적 특성은 군집(클러스터링)의 형태와 규모를 통해 살펴볼 수 있는데 해당 군집의 단어를 살펴보면 ‘임신전’의 경우 임신증상, 가능성과 관련한 군집이 가장 큰(26개 단어) 반면 ‘임신후기’의 경우 ‘출산’과 관련한 군집이 가장 큰 것(28개의 단어)으로 나타났다.

4.5 네트워크 분석을 통한 아이의 성별에 따른 내용분석

본 연구에서는 태아의 성별에 따라 내용적 특성이 차이가 있을 것을 가정하였다. 태아의 성별을 언급한 게시물을 구분하고 각 그룹에서 핵심어를 추출하여 네트워크 분석한 결과는 <그림 2>, <그림 3>과 같다.



<그림 2> 아들을 임신한 경우의
네트워크 분석



<그림 3> 딸을 임신한 경우의
네트워크 분석

두 개의 네트워크를 비교해보면 아들의 네트워크 밀도는 0.3063으로 딸 네트워크의 0.7167에 비해 촘촘하지 않으며 군집형태(Clustering)도 차이가 있음을 볼 수 있다.

내용적으로 살펴보면 각 네트워크에서 공통적으로 출현하는 용어가 있는 반면, 네트워크 내에서만 출현하는 단어가 있음을 알 수 있는데 그 예는 <표 5>와 같다.

<표 5> 네트워크 구분에 따른 출현 단어

구분	개별 출현 단어	공통출현단어
아들 네트 워크	간절, 집착, 태몽, 외동아들, 양가, 시 택, 다행, 태몽	배, 성별, 궁금, 생리, 건강, 방 법, 몸, 냄새, 영향
딸 네트 워크	치즈, 요플레, 구이, 오렌지, 고기, 빵, 식육,	

5. 결 론

본 연구에서는 인터넷 임신 Q&A 게시판의 내용적 특성을 네트워크 분석을 통해 살펴보았다. 이는 연구자의 직관에만 의존하여 내용을 분석을 하던 한계에서 벗어나 계량적으로 내용 분석을 시도했다는 점에서 의의가 있다. 연구자는 계량적인 분석을 위해서 동시출현분석을 사용하였으며 네트워크 분석프로그램 및 시각화 도구를 통해 내용분석을 수행하였다.

결과를 종합해보면, 연구 대상인 게시물의 시기별로 네트워크 분석을 수행한바 임신 Q&A 게시판의 글은 대체로 밀도가 높으며, 단어 간의 관계를 살펴본 바, 각 시기별로 핵심어가 존재하고 핵심어를 중심으로 커뮤니케이션이 활발하게 일어나고 있음을 알 수 있었다. 또한 태아의 성별에 따라 게시물의 내용적 특성은 확연한 차이가 있음을 알 수 있다.

다만 본 연구는 두 가지 제약을 지니고 있다. 하나는 기간이 데이터 수집 기간이 상대적으로 짧기 때문에 분석된 데이터가 커뮤니티의 내용적 특성을 면밀하게 분석하여 해석을 제공한다고 볼 수는 없다. 또 하나로는 동시출현분석은 출현하는 빈도만을 나타내기 때문에 정의 영향인지, 부의 영향인지를 전혀 알 수 없다는 점이다. 즉 네트워크의 노드와 링크가 시각적으로 표현이 되고 군집이 되지만, 실제로는 문맥적 해석은 다시금 연구자에 의해 수행될 수밖에 없기 때문이다. 이와 같

은 부분들을 보완할 수 있는 방법론을 적용하여 연구를 확장한다면 보다 구체적인 분석이 가능할 것이다.

참고문헌

- 김여진. 2001. 『가상공동체의 커뮤니케이션 특성에 관한 연구』. 석사학위논문. 한국외국어대학교.
- 나정호. 2007. 『웹 커뮤니티 활동에 있어서의 성별 차이에 대한 연구 - 텍스트의 내용분석 : 온라인 커뮤니티 게시물 중심으로』. 석사학위논문. 한양대학교 경영대학원.
- 박한우, Leydesdorff, L. 2004. 한국어의 내용분석을 위한 KrKwic 프로그램의 이해와 적용: Daum.net에서 제공된 지역혁신에 관한 뉴스를 대상으로, 『한국자료분석학회지』. 6(5): 1377-1387.
- 이세진, 이정교. 2012. 온라인 여성 커뮤니티에서의 이용자 간 상호작용과 사회연결망에 관한 연구. 『社會科學研究』. 38(2).
- 이승민. 2007. 『초등학교 여교사의 임신·출산 및 육아문제에 관한 연구: 온라인 초등학교 커뮤니티 「인디스쿨」 분석을 중심으로』. 석사학위논문. 창원대학교.
- 장정우, 최경호. 2012. 언어네트워크분석을 이용한 통계법 내용분석. 『통계연구』. 17(2): 53-66.
- 최영출, 박수정. 2011. 한국행정학의 연구경향 분석: 네트워크 텍스트 분석방법의 적용, 『韓國行政學報』. 45(1): 123-139.
- 홍일여. 2012. 『영유아동 양육모의 인터넷 육아정보 커뮤니티 이용에 관한 연구』. 석사학위논문. 동국대학교.
- 한국어 분석 모듈 KLT-2010
 <<http://nlp.kookmin.ac.kr/HAM/kor/download.html>>
- Krkwic <<http://www.leydesdorff.net/krkwic/>>
 Ucinet ver.6.487
 <<https://sites.google.com/site/ucinetsoftware/home>>
- vosviewer ver.1.5.4 <<http://www.vosviewer.com/>>