



빅데이터 속에서 표현된

사회혁신 키워드 분석

- ▶ 빅데이터 활용의 중요성
- ▶ 빅데이터 분석 방법론에 대한 검토
- ▶ 분석에 사용한 빅데이터 소스와 분석방법
- ▶ 주요 혁신 키워드에 대한 빅데이터 분석 결과
- ▶ 온라인 빅데이터 분석에서 얻은 시사점

1. 빅데이터의 활용의 중요성

- 지금까지 오픈랩(Open Lab)과정에서 탐색하고 발견해왔던 사회혁신 키워드들이 과연 어느 정도 객관성을 가지고 있는지, 나아가 다른 방식으로 탐구했을 때와 달리 어떤 특징과 장점들이 있는지를 간접적으로 알아보는 하나의 방법으로, 최근 기술적으로 각광을 받고 있는 빅데이터 활용을 고려해 볼 수 있음. 따라서 여기에서는 사회혁신 키워드에 대해서 어떻게 빅데이터를 활용할 수 있을지, 그리고 구체적으로 온라인 SNS 빅데이터를 기반으로 정보 분석을 했을 때 어떤 결과들을 나타나는지를 제한된 수준에서나마 확인해보려고 함.
- 정보기술의 발전, 초고속 네트워크의 글로벌 연결망 확대, 스마트 폰으로 상징되는 모바일 기기의 폭발적 증가와 인터넷 접속 등은 어마어마한 양의 디지털 데이터가 축적되는 새로운 시대를 열었음. 이러한 현상을 ‘빅데이터’로 부르고 있으며 이를 분석하는 ‘데이터 과학’이라는 새로운 학문분

야도 생각하고 있음.

- 2010년 구글 전 CEO 에릭 슈미츠 회장은, 현재 인류는 과거 인류역사 전체가 생성한 데이터보다 더 많은 데이터를 ‘단 하루 만에’ 만들어내고 있다고 주장함.(인류 문명이 시작된 후 2003년까지 축적된 데이터를 모두 합치면 약 50억 기가 바이트가 된다고 추정함. 그런데 최근에는 한 해에 약 3조 기가 바이트의 데이터 생성되는 것으로 알려져 있음¹⁹¹⁾)
- 빅데이터는 기존의 온라인 거래 데이터에 더하여, 모바일과 인터넷 등을 통해서 발생하는 시민들의 모든 활동들이 문자와 음성, 이미지와 동영상 등을 통해 디지털로 거대하게 축적되면서 만들어지고 있음. (포털 사이트를 통한 검색정보, SNS를 통한 대화와 관계망 정보, 스마트 폰 위치 서비스를 통한 동선 정보 등)

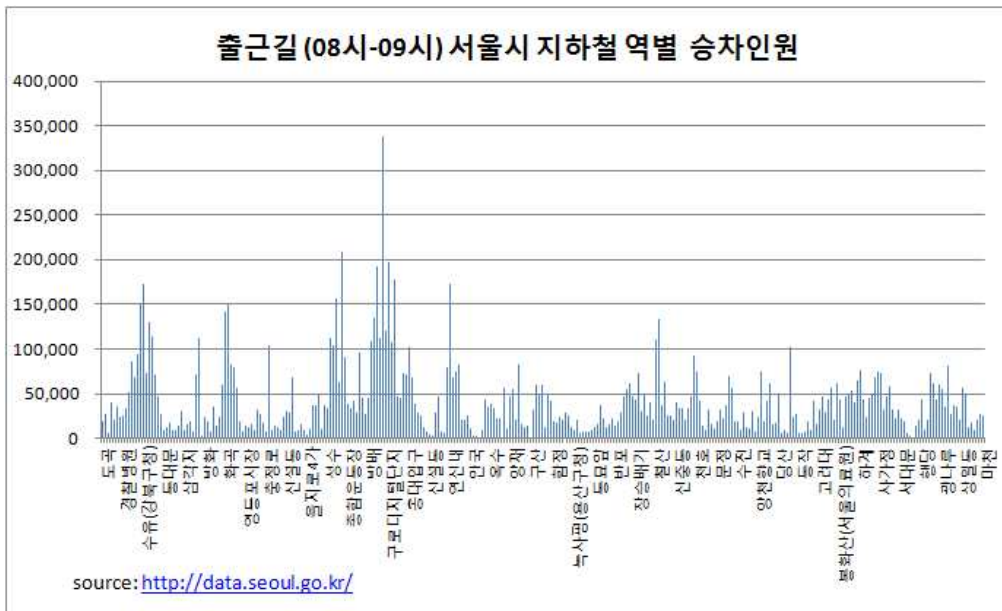


그림 6-1: 예시 - 서울시 공공 빅데이터 서비스 사례

- 서울시에서도 이미 서울 열린 데이터 광장(<http://data.seoul.go.kr/>)을 개설하여 공공 빅데이터 정보서비스를 초보적으로 시작함.(예를 들어, 교통카드 정보를 이용해서 서울시 모든 전철역의 모든 승객 흐름에 관한 데이터

191) 빅토어 마이어 외(2013), 『빅데이터가 만드는 세상』, 21세기 북스

를 빅데이터로 추적하여 공개하고 있음. 위의 표는 서울시가 공개한 데이터를 이용해서 2015년 1월 간단한 데이터 평균을 구해본 간단한 사례임.)

- 지난 몇 번의 선거 국면에서 온라인 여론 동향을 민첩하게 포착할 때, 사회적 재난이나 대형 이슈가 터졌을 때 여론의 향방을 모니터하려고 할 때 등, 다소 짧은 시간대의 여론 변화의 민감성을 포착하기 위해 주로 사용했음. 그러나 일정한 기간 대중의 인식이나 관심분야의 안정적 흐름이나 시계열적 변화 방향을 추적하려는 노력은 적었던 것이 사실임.
- 따라서 다수 시민이 관심을 갖는 사회적 문제들에 대해서, ‘시민 스스로 해법을 모색하거나 여론을 모아가는 과정에 대한 탐구의 중요한 방편’으로서 온라인상의 빅데이터 분석이 가능하며 의미를 지니게 됨
- 결국 빅데이터 속에 표현된 시민들의 생각과 아이디어 수집은, 시민들의 지향이 갈수록 복잡하고 다양화되는 현대사회 환경에서, 표본 여론조사 등 전통적인 샘플링 조사 방법이 가진 한계를 근본적으로 뛰어넘어 더욱 섬세하게 다차원적/복합적 여론을 얻을 수 있는 중요한 기회를 제공해주고 있음을 알게 됨.



그림 6-2: 사회혁신 영역에 빅데이터 기술을 적용해 보는 이유

- 따라서 사회관계망서비스(SNS)를 중심으로 거대하게 축적되고 있는 빅데이터를 사회분석과 시민인식조사의 영역에서 활용하는 것은 다른 모든 분야에서와 마찬가지로 사회혁신 활동과 혁신 정책 수립에도 매우 중요한 가치를 지닌다고 할 수 있음. 여기에서는 기왕에 오픈랩(Open Lab)에서도 출된 사회혁신 키워드의 몇몇 중요 문맥에 대해서, 온라인 빅데이터 분

석을 시도해봄으로써, 두 결과를 상호 비교해보고 각각의 특징과 장단점을 평가해보려고 함.

2. 빅데이터 분석 방법론에 대한 검토

“일단 세상이 데이터화되고 나면 정보의 잠재적 용도를 제한하는 것은 기본적으로 개인의 창의력뿐이다.”

1) 빅데이터와 상관성

- 빅데이터에서 ‘빅’은 절대적인 개념이 아니라 상대적인 개념이며, 특히 무작위 ‘샘플링’의 수준을 뛰어넘는 거의 해당 영역 데이터 전부를 포괄해야 하는 것으로 정의되고 있음. 이론 인해, 각종 여론조사나 표본 추출을 통한 사회조사 방법과 질적으로 다른 사회조사 분석의 가능성을 열어주고 있음. “많은 데이터를 가진 간단한 모델이 적은 데이터를 가진 정교한 모델보다 뛰어나다.”
- 또한 빅데이터는 거대한 데이터 속에서 탄탄한 상관성과 패턴을 보여준다는 점에서 기존 사회조사 데이터와 확실히 구별됨. 즉 빅데이터는 사회와 사건의 정확한 인과성(causality)를 보여주지는 못하지만, 과거에는 전혀 불가능했던 광범한 영역과 이질적인 관계에서의 강력하고 신뢰할 만한 상관성(correlation)과 일정한 패턴(pattern)을 보여주게 됨. 결국 빅데이터는 엄청난 양의 데이터에 수학을 적용하여 확률을 추론하려는 노력이라고 할 수 있음. “비인과적 분석은 점점 더 많은 분야에서 느린 인과적 사고보다도 더 유용하고 효율적인 툴이 되고 있다.” “데이터만 충분하다면 숫자들은 스스로 입을 연다. 페타바이트는 이런 말을 가능하게 한다. 상관성이면 충분하다.”¹⁹²⁾

192) 빅토어 마이어 외(2013), 『빅데이터가 만드는 세상』, 21세기 북스

2) 빅데이터 기반 분석의 정확성

- 빅데이터가 가진 잠재력을 가장 잘 보여주고 있는 하나의 사례는 구글이 선보였던 ‘미국내 독감발생 추이 예측’이라고 할 수 있음. 미국 질병관리국(CDC) 역학 예방팀은 독감 예측을 위해 통상 3000여개 진료소, 3000여 명의 외래 진료환자 진단의사, 122개 도시의 인구통계 사무소, 각 주 보건국 관계자들로부터 모든 자료를 동원하는 거대한 작업을 수반함. 그러나 구글은 자신들이 보유한 검색 빅데이터를 기반으로, 검색사용자들이 평상시 사용한 검색어들 가운데 독감 징후와 관련 있는 검색어를 추출함으로써, 미국 질병관리국이 예측한 독감추이와 거의 유사한 결과를 유도해내고 있음. (이와 비슷한 사례는, 구글이 온라인 쇼핑거래정보 빅데이터를 이용하여 온라인 상품 물가지수를 공개하는 구글 물가지수(Google Price Index)임. 각 국가 통계청에서 정기적으로 발표하는 소비자 물가지수(CPI) 역시 엄청난 인력과 자원이 동원되고 있음.)

3) 다양한 빅데이터 소스

- 우리나라의 포털 업체들 역시, 구글 정도의 수준은 아니더라도 빅데이터를 점차적으로 활용하고 있고 앞으로도 계속 확대할 예정에 있음. 예를 들어, 국내 포털업체 네이버는 광고주 등을 위해 검색어와 연관된 트렌드 서비스를 제공함. 이 서비스를 통해 ‘안전’이라는 키워드의 사용빈도를 뽑아보면, 지난 2014년 세월호 참사를 전후하여 안전 이슈가 사회적으로 급격히 높아졌고, 이 추세는 그해 10월까지도 이어졌음을 볼 수 있음. (아래 그림 참조)

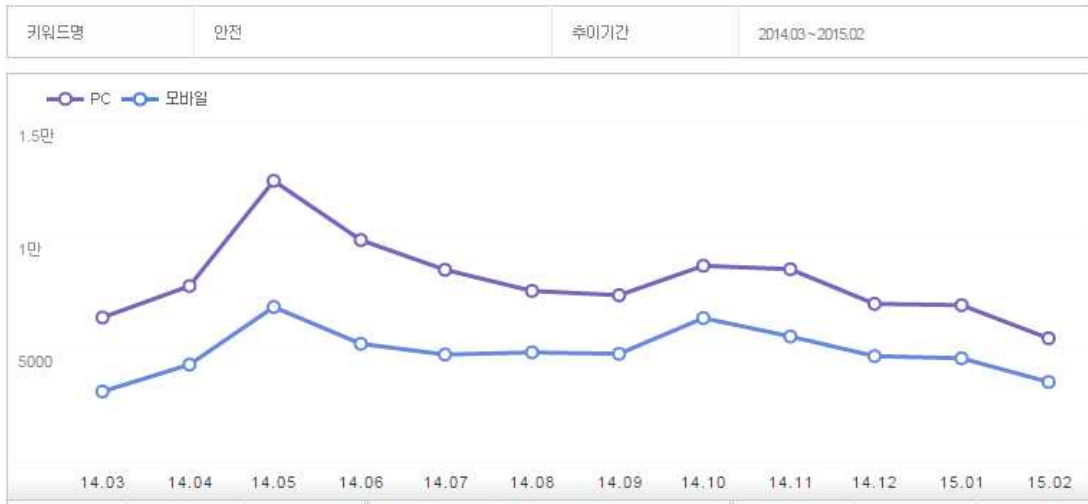


그림 6-3: 예시- 네이버에서 분석한 ‘안전’ 키워드 사용 추이(2014.3~2015.2)

- 물론 시민들의 사회적 활동 및 여론 분석이 포털 사이트 검색 데이터로만 가능한 것은 아님. 앞에서 예시한 시간대별 지하철 승하차 인원 추이는 교통카드 사용으로 발생한 빅데이터를 이용한 것임. 또한 구글에서 구축하고 있는 프로젝트인 ‘엔그램뷰어’(ngrams.googlelabs.com)는, 세계에서 출간된 모든 책을 디지털로 변환하는 것임. 구글이 지난 2010년 당시 이 프로젝트를 공개할 때에, 1500~2008년 출간된 520만 권 책을 스캔하여 단어 사용 빈도를 따져 데이터베이스(DB)를 구축했음.(전세계 책의 4% 정도). 이는 약 5000억 개의 단어가 망라된 것인데, 책에 사용된 단어를 통해서도 변경 추이를 파악할 수 있음.

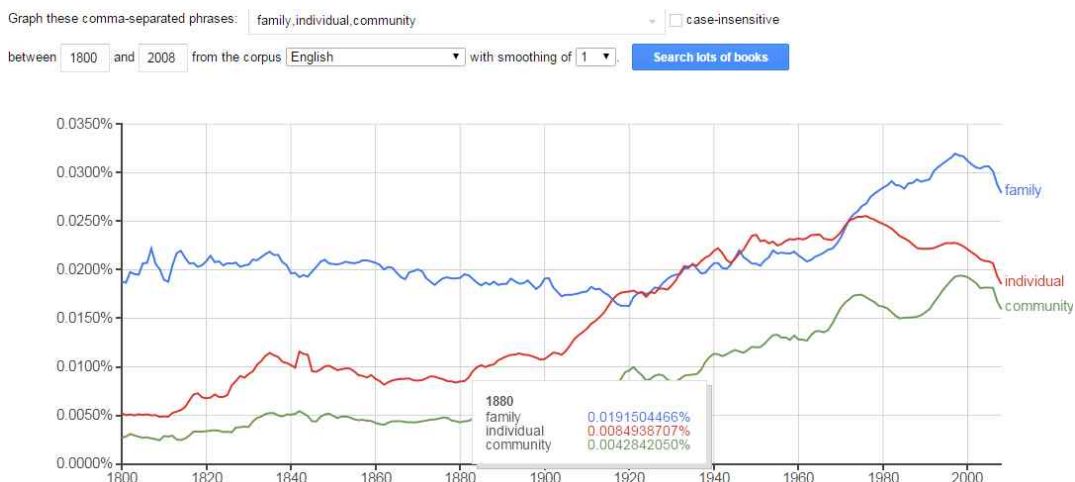


그림 6-4: 구글 엔그램뷰어로 분석해본 가족/개인/공동체 키워드 빈도 추이

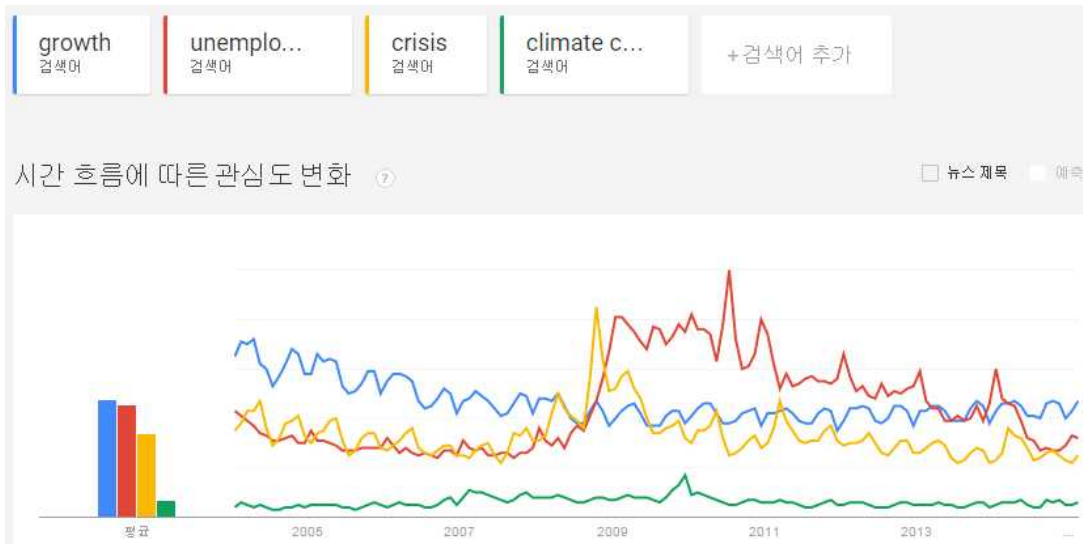


그림 6-5: 2008년 경제위기 전후로 성장/실업/위기 키워드의 사용빈도 변화

- 특히, 단기적인 여론동향을 벗어나서, 다소 긴 호흡의 차원에서 시민인식 구조 변화의 추세와 방향을 분석하고자 할 때 빅데이터는 매우 중요한 근거를 제공해줄 수 있음. 앞서 예시한 구글 독감 트렌드와 동일한 방법론을 사용하여 분석하는 구글 트렌드는 2004년부터 현재까지 약 10여 년 동안의 여론 흐름을 확인할 수 있는 지표임. (*국내 포털 업체들은 최대 7년 정도의 시계열 서비스를 제공하고 있으나, 정확성이 명확히 파악되지는 않았음)

4) 빅데이터를 이용한 시민의식 분석방법과 사례

- 빅 데이터를 기업이나 상업적인 분야 이외에 정치, 사회 분야에 적용하고 활용도를 제고한 경우는 아직 많지는 않으나 점점 확대되고 있는 추세임. 이와 관련하여 가장 전형적이면서 대표적인 사례가 “2012년 미국 대선에서의 오바마 대통령 선거운동”이라고 할 수 있음.
- “오바마 캠프는 미국 대통령 역사상 처음으로 내부 CTO와 대규모 테크놀로지팀을 두고 목적의식적으로 빅데이터를 통합하고 분석함. (심지어 오바마 캠프는, 후원금 모금에 참여한 사람들의 과거 데이터와 기부 패턴을 검토하여 후원금을 얼마나 요청할지를 도출해내기도 했음.

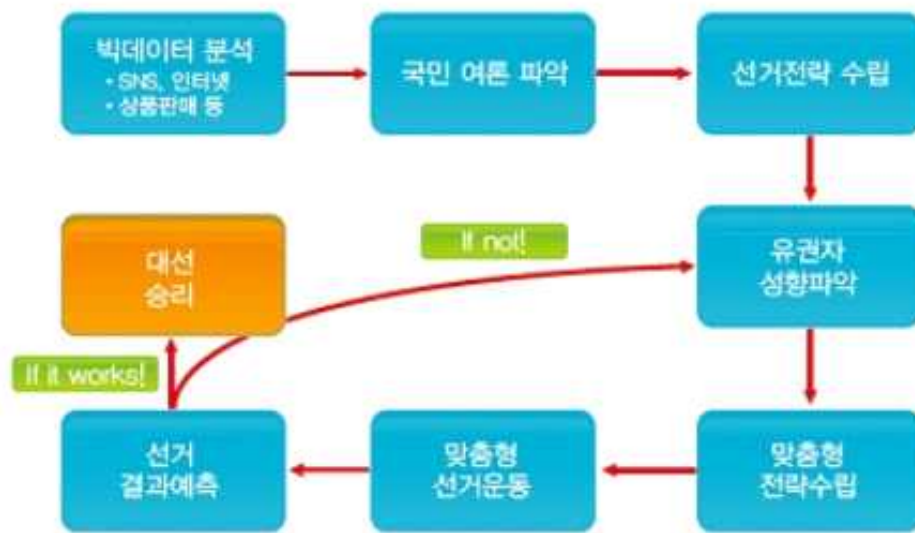


그림 6-6: 빅데이터를 활용한 선거운동 사례 (출처: 정보화진흥원)

- 즉, 과거 기부자들의 후원금 중 최고금액을 알아본 후, 과거의 최고 후원 금액에 약간씩 변동을 준 금액을 요청하는 실험을 하였음. 예를 들어 이전에 35달러를 후원금으로 낸 사람들을 대상으로 33, 34, 35, 36, 37, 38달러를 요청하는 이메일을 각각 보냈음. 그 결과 그런 방식의 후원금 요청 메일을 단순히 10달러, 30달러, 50달러 식으로 요청하는 것보다 모두 좋은 결과를 얻었다고 함.)¹⁹³⁾
- 공공정책 수립의 근거가 되는 시민생각이나 여론이 담겨 있는 빅데이터로서, 현재 단계에서 가장 일반적으로 활용되는 데이터는 SNS 데이터라고 할 수 있음. 특히 인터넷과 SNS 활용률이 매우 높은 우리나라에서 SNS 온라인 빅데이터 분석을 중심으로 시민들의 인식구조나 변화 추이를 탐색해보는 것은 매우 의미가 있을 것으로 판단됨.
- 예를 들어, 한 빅데이터 전문 업체의 프로그램 통해 서울시에서 추진 중인 주요 정책의 하나인 ‘도시 재생’을 분석해보면 아래 그림과 같은 결과를 얻을 수 있음. 아래 예시는 기간이 짧아 데이터 왜곡이 발생할 여지가 있

193) 고태석(2013), 『빅데이터, 승리의 과학』, 이지스퍼블리싱

다는 점을 감안하더라도, 시민들에게 도시 재생이 ‘다양하다.’, ‘도움이 되다.’와 같은 긍정적인 의미로 이해되고 있을 뿐 아니라, ‘불량’과 같은 극단적인 부정적 맥락으로도 받아들일 수 있다는 시사점을 주고 있음. (서울시가 추진하는 또 하나의 정책 키워드로서 ‘마을’을 검색해보면, 전통적인 과거 지향 개념인 마을의 의미인 ‘조용하다’, ‘아름다운’과 더불어 미래지향적 개념인 ‘부활’, ‘꿈꾸다’와 같은 의미로 이해되고 있다는 점에서 흥미로운 대목이 발견됨.)



그림 6-7: www.daumsoft.com에서 제공하는 social metrics 이용한 분석

- 위의 사례는 과거의 전통적인 샘플링 분석에서는 찾아보기 어려운, 시민들이 가지고 있는 생각들을 복합적으로 보여주거나, 더 나아가 시민들 속에 잠재되어 있는 사회혁신의 단초와 소망을 읽어낼 수 있는 기회를 제공해주고 있음.

3. 분석에 사용한 빅데이터 소스와 분석 방법

- 이번 분석에 사용한 빅데이터는 기본적으로 다음 소프트가 제공하는 ‘Social Metrics’라는 데이터와 틀을 기본으로 했고, 부수적으로 네이버 트렌드(trend.naver.com)나 구글 트렌드 등을 활용했음.
- ‘Social Metrics’는 트위터, 블로그, 뉴스를 주요 빅데이터로 구축하고 있고 기본적으로 2007년부터 데이터를 쌓아오고 있지만 완전한 데이터는 2013~2015사이라는 점을 주요 특징으로 하고 있음. 또한 엑셀형태로 분석한 범위의 원데이터를 풍부하게 제공해주고 있음. 여기에 더해 연관어 분석, 탐색어 추이, 여론 동향 등을 다양하게 분석하게 해주는 다양한 그래픽 틀을 제공해준다는 강점을 가지고 있으며 서비스 경험이 오래되어 신뢰도가 상대적으로 높았기 때문에 활용대상으로 선정했음.

	주요 내용
분석 소스	국내 온라인 사용자들의 SNS(트위터, 블로그, 뉴스 등)과 포털 서비스의 검색어
분석 기간	2012년 12월 이후의 데이터를 주로 활용하되, 필요하면 2007년 이후 데이터 조회
분석 지역	한국에서 발생한 데이터로 제한하는 것을 원칙으로 하되, 해외 유사사례를 조사하여 비교하는 방법 검토.
분석 영역	오픈랩 프로젝트에서 도출된 사회혁신 키워드와 연관된 검색키워드로 활용하고, 빅데이터 분석과정에서 연관 키워드관의 구체적인 관계성을 찾고 확장.
분석 내용	빈도분석, 연관어 분석, 의미분석을 중심으로 하되 시계열 자료 여건이 허락하는 선에서 추세분석, 예측분석 부가.

표 6-1: 분석 대상과 범위 개요

- 온라인 빅데이터 분석에서 유의해야 할 사항을 덧붙여 둔다면, 우선 앞서 지적한 것처럼 빅데이터는 근본적으로 데이터양이 많다는 뜻이 아님. 빅데

이터는 일반적으로는 당연히 양이 많을 수밖에 없지만, 원래의 의미는 추출(extraction), 또는 요약(Aggregation)되는 데이터가 아닌, 해당 사건과 관련하여 발생한 ‘거의 전체 데이터’를 직접 대상으로 하함. 또한 매우 정제되어 있는 데이터가 아니라 로그 데이터처럼 실시간으로 쌓여가는 원 데이터를 그대로 거칠게(?) 수집한 것에서 분석함.

- 따라서 필연적으로 추출된 데이터의 정확성 보다는 거의 전부를 포함한 데이터의 패턴을 발견하는 것이 목적임. (앞으로 분석과정에서 의뢰로 조회결과가 많지 않게 나올 수도 있다는 점을 고려해야 함. 실제로 전체 데이터가 그 정도에 그치기 때문임.)

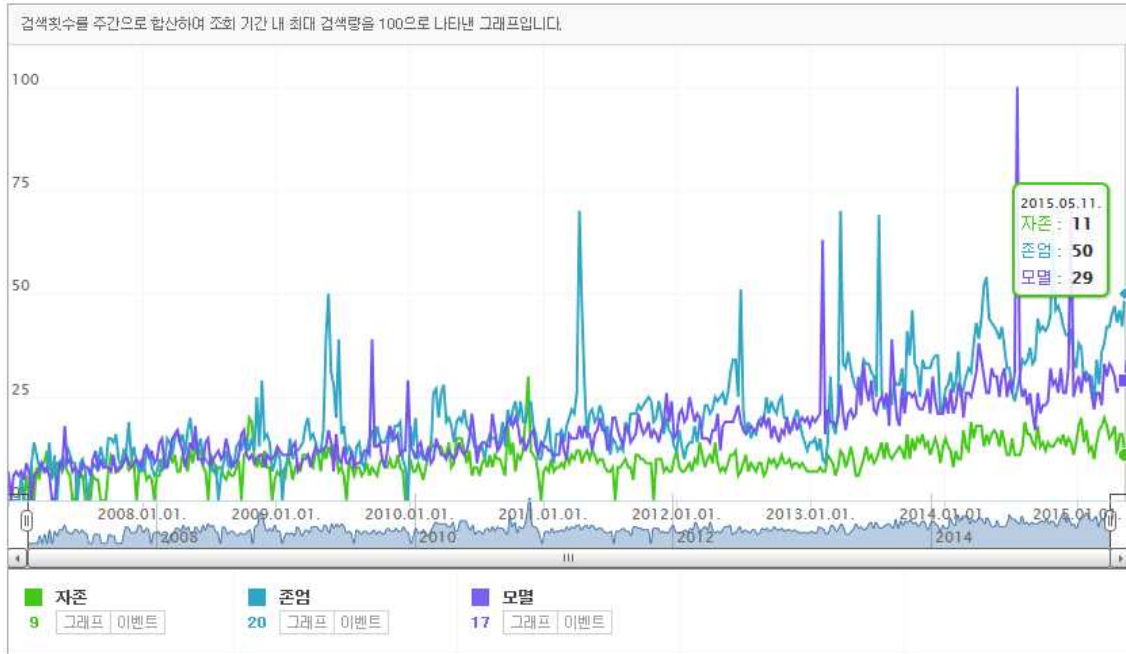
- 또한 빅데이터는 그 자체로 정제되어 있지도 않고 의미(semantic)에 따라 정확히 분류되어 있지도 않기 때문에 그 자체로는 아무것도 말해주지 않음. 따라서 일정한 맥락에 따라 분석가가 분석(analyzing)하고 마이닝(mining)해야 함. 그 과정에서 잡음(noise)을 제거하면서 기대하는 맥락(context)을 추적해야 비로소 가치 있는 정보를 얻게 되는 것임.(예를 들어, 빅데이터에서 ‘개인화’라는 키워드를 검색하게 되면 전혀 의미와 맥락이 다른 다양한 문장들이 통째로 뒤섞여 검색되기 때문에 아무런 의미나 시사점을 파악하기 어려움. 개인화(individualization)은 공동체와 비교되는 개인의 정체성을 표현하는 개념으로써 사회혁신 키워드에서 다루려고 하는 개념임. 그러나 우리말로 동일하게 번역되는 개인화(personalization)은 통상 마케팅이나 IT에서 사용되는 맞춤형 서비스 형태를 지칭하는 용어로 광범위하게 사용되고 있음. 따라서 단순히 개인화라고 하는 키워드만으로는 의미있는 빅데이터 분석이 가능하지 않음.)

- 따라서 빅데이터 분석에서 단순히 낱개 어휘를 사용한 분석은 명백히 제한성을 갖게 되며 단어와의 연관되어 포함하거나 배제해야 할 단어를 추가로 조합해야 함. 나아가서 범위를 좁혀가면서 계속 데이터를 추적해야 함.(이른바 드릴링 다운 기법의 사용)

- 또한 의미가 명확히 드러나기 위해서 대조적인 개념과 비교하면서 분석할 수도 있어야 하며, 그럼에도 불구하고 잡음을 완전히 제거하는 것이 불가능하므로 예외적 현상이 도드라지는 영역의 데이터에 대해서는 원래 데이터의 문장 자체를 탐색할 수 있어야 함.(다행스럽게 아래에 사용한 Social Metrix)는 이와 같은 기능을 제공하고 있음.)
- 마지막으로 데이터 분석의 기본은 대개 ‘시계열에 따라 정보가 어떻게 변해 가는가’하는 점을 추적하는 것임. 따라서 최대한 긴 시계열에 걸쳐 동일한 밀도로 누적된 데이터가 있어야 정확한 추세(Trend)를 분석할 수 있음. 하지만 너무 시계열이 길어지면 데이터 폭발(Data Explosion)이 일어나 조회 성능에 심각한 영향을 주게 되므로 trade-off관계를 따져서 사용할 필요가 있음.(국내 온라인 빅데이터를 제대로 축적한 시계열은 불행이 길지 않기 때문에 불가피하게 2010년대 이후의 짧은 시계열, 길다고 해도 2007년 이후 정도에 만족해야 함.)

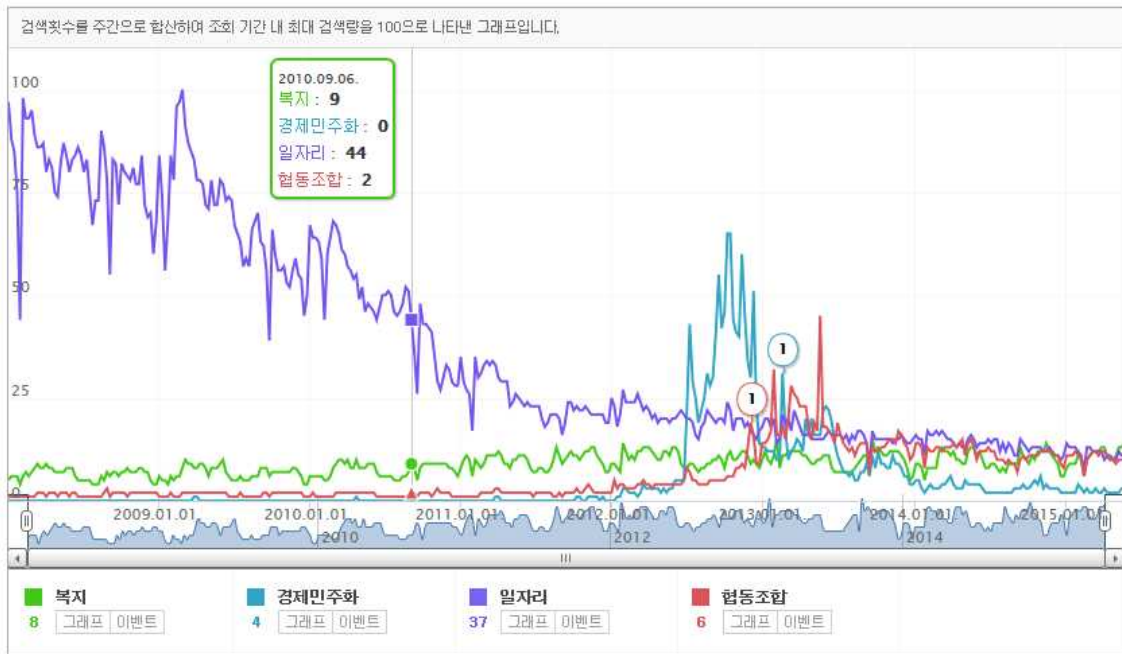
4. 주요 혁신 키워드에 대한 빅데이터 분석 결과

1) 주요 키워드에 대한 장기 시계열 분석



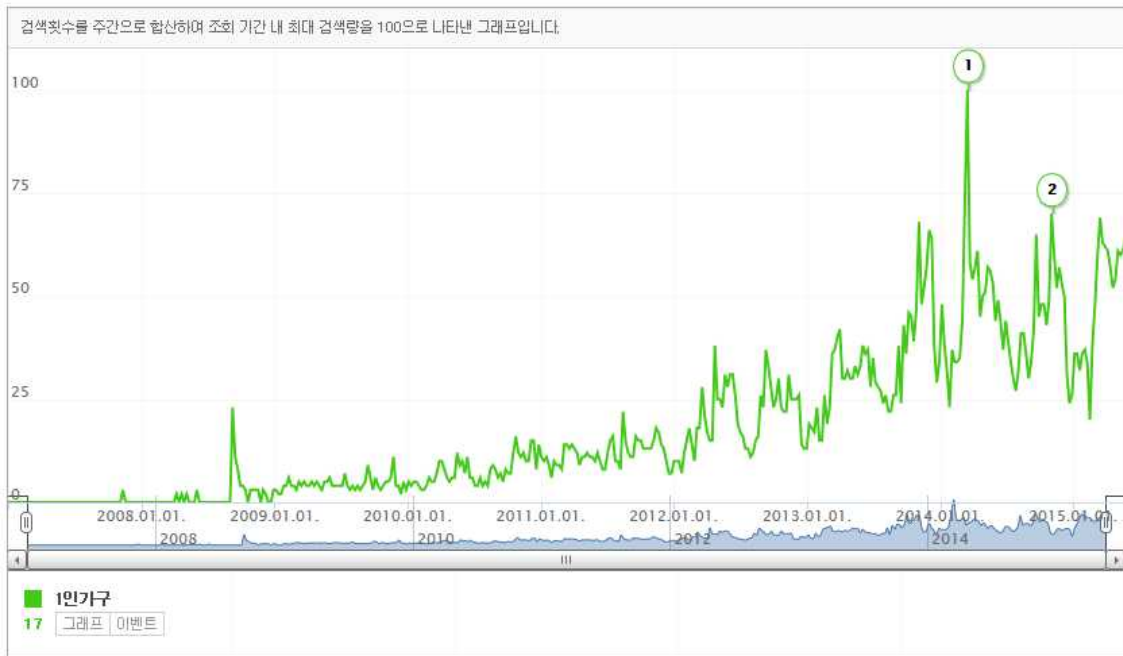
[주제어: 자존/ 존엄/ 모멸, 기간: 2007.1~2015.5, 데이터: 네이버 검색어]

- 개인의 자존과 존엄을 표현하는 키워드에 대한 장기 시계열을 보면, 의외로 이들 어휘는 꾸준하고 지속적으로 우리의 일상에서 사용하는 개념임을 알 수 있음.
- 그러나 이들 개념의 사용(특히 모멸)은 꾸준히 증가하고 있는데, 이는 점점 더 '일상용어'에서 '사회적 문제를 내포한 의제'로 전환되어 가고 있다는 추측을 해보게 함.(다른 방식으로 추가 조사가 필요한 대목임.)
- 또한 모멸과 같이 부정적인 용법으로 자존이나 존엄(예를 들어 자존감의 상처 등)을 사용하고 있다면 대처가 필요한 의제가 될 수 있다는 것을 암시함.



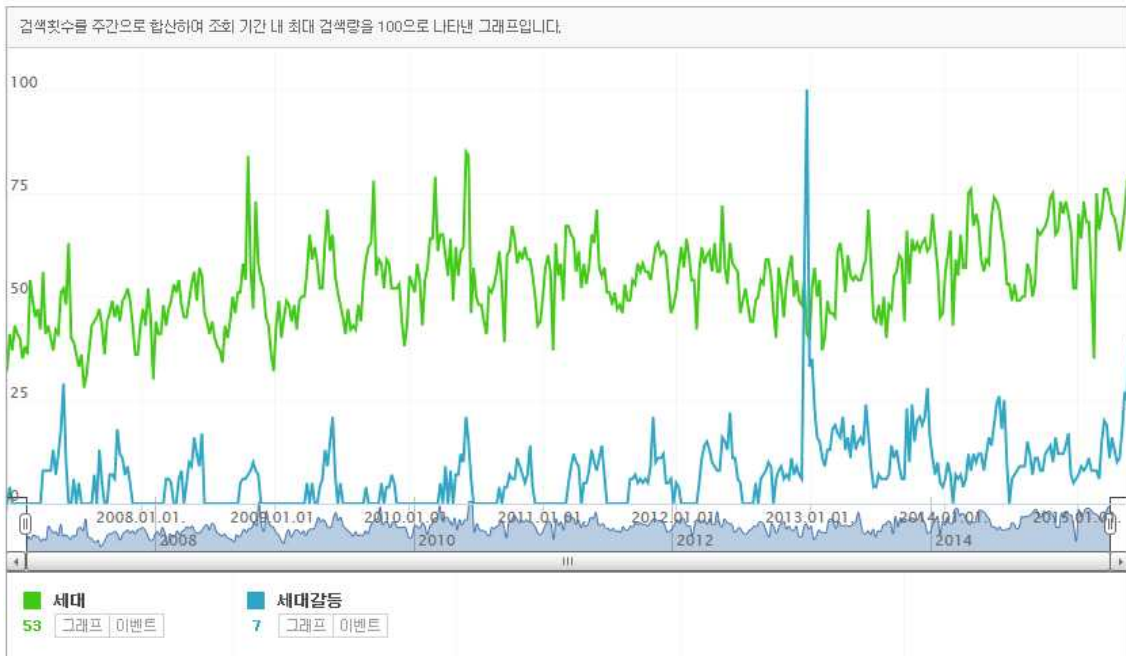
[주제어: 복지,경제 민주화,일자리,협동조합 기간: 2008.1~2015.5, 데이터: 네이버 검색어]

- 우리사회의 정책 어젠다 가운데 가장 대표적인 복지와 경제 민주화, 일자리의 시계열 추이를 분석해보면, 서로 대비되는 뚜렷한 징후를 볼 수 있음. 일자리는 글로벌 경제위기 초반기(2008~2010)에 집중적으로 의제화 되었다가 서서히 약화되고 있지만 여전히 상당한 지속성을 갖는다는 것임.
- 경제민주화 키워드가 가장 극적인데, 2012년까지는 거의 등장조차 미미하던 경제 민주화는 대선기간인 2012년에 최대화두가 되었음. 그러나 2013년까지 근근이 이어져오던 경제 민주화 이슈는 2014년부터 주요 의제의 무대 뒤로 퇴장하는 분위기임.
- 경제 민주화 대신에 온라인에 강력하게 등장한 키워드는 바로 ‘협동조합’임. 이제 협동조합은 복지나 일자리 만큼 온라인 공간에서 일상화 되었음.
- 복지라는 짧은 두 단어는 지속적으로, 그리고 점차 증가하는 방향으로 온라인에서 거명되는 개념임.(복지가 어떤 표현과 결합하여 맥락이 바뀌는지는 여기서는 파악하기 어렵고 좀 더 분석을 필요로 함.)



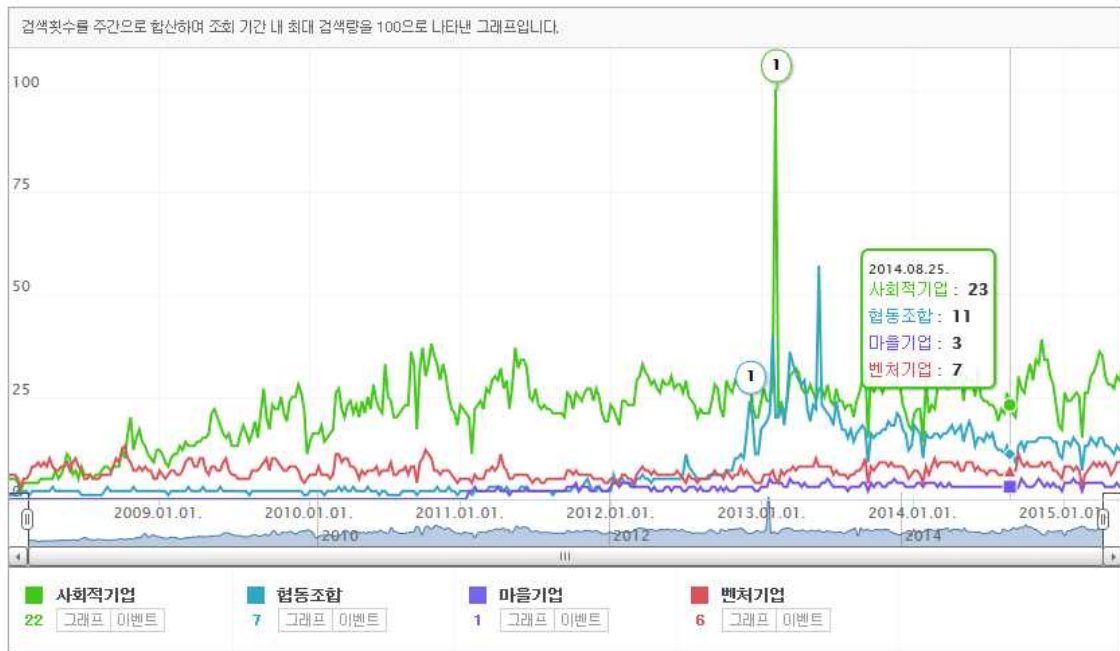
[주제어:1인 가구, 기간: 2007.1~2015.5, 데이터: 네이버 검색어]

- 개인화를 알아 볼 수 있는 대체 변수로서 '1인가구'의 장기 추이를 검색해 본 결과 꾸준히 상승추세를 보이고 있고, 특히 2012년 이후 검색어에서 1인가구 키워드가 나타나는 빈도는 매우 가파르게 증가하고 있음.
- 이정도 추이면 1인가구라는 키워드는 우리사회의 주요 의제로서 굳어져가고 있다고 판단됨.



[주제어: 세대, 세대 갈등, 기간: 2007.1~2015.5, 데이터: 네이버 검색어]

- 세대라는 키워드는 적어도 주어진 검색기간 안에서는 큰 편차 없이 온라인 공간에서 일상적으로 사용하는 어휘가 되었음.
- 그러나 세대갈등은 (키워드가 가지고 있는 이벤트적인 성격이 있다는 것을 감안해도) 2013년 이후부터 빈도수가 높아지고 있고, 더욱이 여전히 변동성이 있다고는 하지만 상당히 지속성이 높아져가는 것을 발견할 수가 있음.
- 반면에 문제 해결을 암시하는 ‘세대정의’나 ‘세대연대’, ‘세대협력’등은 아예 온라인 데이터에서 등장조차 하지 않음.
- 세대의제를 부정적 갈등 요인에서 긍정적 에너지로 변화시키는데 사회혁신의 미션이 있을 수 있다고 가정해 볼 수 있음.



[주제어: 사회적기업, 협동조합, 마을기업, 벤처기업, 기간: 2007.1~2015.5, 데이터: 네이버 검색어]

- 적어도 온라인에서 사회적 기업은 2009년부터, 협동조합은 2013년부터 벤처기업이라는 키워드를 앞질렀고 이런 추세는 지금까지 지속되고 있음.
- 2007년부터 사회적 기업이 육성된 점 등이 감안되어 협동조합보다는 사회적 기업이 훨씬 더 많이 등장하고 있고, 2013년에 잠깐 협동조합이 높은 빈도를 보였지만 사회적 기업의 안정적인 빈도와는 비교가 안 됨.
- 2012년부터 처음으로 온라인상에 등장한 마을기업은, 비록 사회적 기업이나 협동조합에 비해서 빈도수가 턱 없이 낮지만 지속적으로 발견되고 언급된다는 점을 주목할 필요가 있음.(물론, 의제를 한 단계 더 확산할 계기가 필요해 보임.)
- 여기서 맥락을 정확히 파악하는 것은 어렵지만, 벤처기업은 현재 온라인상에서 거명되고 있는 주요 키워드라고 볼 수는 없음.



[주제어: 도시개발, 도시재생, 기간: 2007.1~2015.5, 데이터: 네이버 검색어]

- 오랜 동안 개발신화에 익숙해졌고, 선거철만 되면 개발의제가 언제나 뜨거운 의제의 중심이었던 우리나라 조건에서, 2013년 이후부터는 도시재생이라는 키워드가 도시개발 키워드를 압도하면서 온라인상에서 검색되고 있음. (물론 도시 재생이라는 키워드가 과연 사회혁신적인 의미로 사용되고 있는지, 아니면 무늬만 바뀐 도시개발이 되었는지는 추가적인 분석이 필요한 대목임)
- 그러나 도시개발이라는 키워드 역시 (상당히 중립적인 용어로 사용될 수도 있지만) 꾸준히 어느 정도 수준을 유지하면서 지속되고 있어서, 어떤 퇴행적 계기가 작동하면 부상할 가능성도 배제하기 어려움.



[주제어: 재개발, 마을공동체, 기간: 2007.1~2015.5, 데이터: 네이버 검색어]

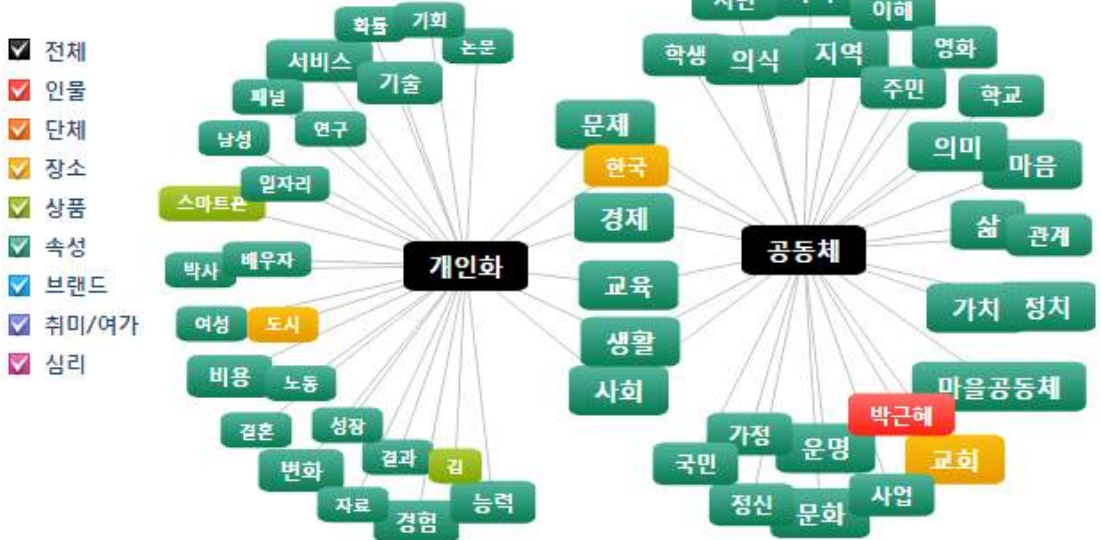
- 2012년부터 갑작스럽게(?) 등장하기 시작한 ‘마을공동체’라는 키워드는 대단히 안정적인 지속성을 가지면서 온라인에서 거명되고 있어서 주목됨. (어떤 동기와 용도로 거명되고 있는지는 뒤에 다른 방식의 분석으로 보충하겠음.)
- 반대로 우리사회에서 언제나 뜨거운 이슈였던 재개발은, 뉴타운 광풍이 있었던 2008~2009년에는 대단히 높은 비중으로 언급되더니, 2012년 이후 절반 이하 수준으로 찾아들었음. 그러나 적어도 마을공동체라는 키워드보다 높은 수준으로 계속 유지되고 있어, 의제과위의 잠재력을 완전히 상실했다고 평가할 수 없음. 다시 말하면, 부동산 거품 등이 재연되는 식으로 계기가 주어지면 언제든 부활(?)할 수 있다고 해석하는 것이 합리적일 것임.

2) 몇 가지 키워드에 대한 의미 분석과 비교분석

탐색어 : 개인화, 공동체 기간 : 2014-12-01 ~ 2015-05-31 구버전

연관어 ✕

매체선택 전체 트위터 블로그 커뮤니티 뉴스 연관어 수 30



No.	연관어	분류	탐색건수
1	서비스	속성	250
2	사회	속성	239
3	기술	속성	173
4	생활	속성	157
5	경험	속성	132

No.	연관어	분류	탐색건수
1	사회	속성	16331
2	지역	속성	13108
3	의식	속성	10003
4	삶	속성	8876
5	마을공동체	속성	7951

[주제어: 개인화와 공동체, 기간: 2014.12~2015.5, 데이터: 다음소프트]

- 그 자체로는 매우 독립적인 어휘인 ‘개인화’와 ‘공동체’를 단순 분석해보면, 압도적으로 공동체에 대한 탐색건수가 컸음. (더욱이 개인화의 경우 서비스나 기술과 연관어가 발견되는 것을 보면 상당부분 individualization의 의미가 아니라, personalization의 의미로 사용했기 때문에 잡음이 크다는 사실까지 감안해야 함.)
- 그런 점에서 보면, 우리나라는 지금도 의연히 온라인상에서조차 각종 공동체(가족, 교회, 학교, 지역, 국가 등)라는 개념이 일상의 주변에 있다는 것을 암시함.
- 다만 공동체 가운데 ‘마을 공동체’가 상당한 순위(5위)로 올라왔던 점이 눈에 띈다고 할 수있음.



[주제어: 1인가구, 기간: 2014.12~2015.5, 데이터: 다음소프트]

- 개인화를 상징할 대체 키워드로서 1인가구를 살펴보면, 연관어 측면에서 (그림의 좌측) 우선 ‘증가’, ‘식사’와 같이 예상 가능한 키워드들이 눈에 띄지만, 다른 한편으로는 가구와 생활, 주택 등 매우 현실적인 물질적 조건과 양식에 대한 연관어가 주류를 이루고 있음. 이는 1인 가구 현상이 정착되었다기보다 생활양식이 만들어져가는 단계라고 파악됨. (이는 앞서의 시계열 분석과도 일치함.)
- 1인 가구를 긍정, 부정의 의미 측면에서(그림의 오른쪽) 살펴보면, ‘외롭다’, ‘좌절하다’, ‘어렵다’, ‘이혼’ 등 상당정도 부정적인 의미들이 많이 개입되어 있음. 이는 1인가구가 꼭 자발적인 선택에 의한 것은 아니라는 점을 암시함.
- 그러나 다른 편에서 ‘대세’, ‘예쁜’, ‘기대’와 같은 긍정어도 적지 않은 걸 보았을 때, 1인 가구 현상을 온통 부정적인 경향이자 비자발적으로 떠밀려서 선택되었다고 해석하는 것이 무리임을 말해줌.



[주제어: 모멸, 잉여, 기간: 2012.12~2013.5(좌측)와 2014.12~2015.5(우측), 데이터: 다 음소프트]

- 자존과 존엄은 잡음이 매우 많은 키워드여서 대체 키워드로 부정적 의미를 담고 있는 ‘모멸’과 ‘잉여’를 선택해 보았음.
- 두 가지 키워드 모두 온라인상에서 폭발적인 빈도를 보이는 것은 아니라 하더라도, 연관어 종류가 매우 다양한 것을 보아서 무게 있는 의미를 가지고 소통되고 있는 것은 확실하며, 특히 주목할 것은 2013년 상반기에 비해 2년 후인 2015년에 두 키워드 모두 탐색 빈도가 높아지고 있다는 사실임. (틀림없이 모멸이나 잉여 등의 의제들이 상승하는 추세라고 이해되지만, 그러나 이 정도의 데이터 값으로 더 이상의 해석을 하기는 어려울 것임.)

탐색어 : 여가 | 기간 : 2012-12-01 ~ 2013-05-31 | 기간 : 2014-12-01 ~ 2015-05-31

금/부정 분석 현황

매체선택 전체 트위터 블로그 커뮤니티 뉴스 연관어 수 30 트위터 블로그 커뮤니티 뉴스 연관어 수 30

- 전체
- 긍정
- 부정
- 중립
- 기타



No.	연관어	분류	탐색건수	연관어	분류	탐색건수
1	즐거움	긍정	5628	즐거움	긍정	8495
2	다양한	중립	2955	다양한	중립	4634
3	스트레스	부정	1437	스트레스	부정	2231

[주제어: 여가, 기간: 2012.12~2013.5(좌측)와 2014.12~2015.5(우측), 데이터: 다음소프트]

- 시간권리의 대체 키워드로서 ‘여가’를 검색해 본 것인데, 일반적인 의미에서 긍정적 의미들로 이뤄져 있음.(부정적 언어로 표시된 ‘스트레스’도 실제 문맥에서는 ‘스트레스를 풀기위한 여가’와 같은 용법으로 사용되었으므로 긍정적 의미를 가진다.)
- 다만 2013년 상반기에 비해 2015년 상반기에 여가라는 키워드의 사용빈도가 다소 올라갔다는 정도가 발견됨.



[주제어: 마을 복지 기간: 2012.12~2013.5(좌측)와 2014.12~2015.5(우측), 데이터: 다음 소프트웨어]

- 마을 복지는 마을공동체라는 키워드가 온라인상에 등장하던 초기 2013년에는 거의 의미 없는 수준으로만 온라인상에서 언급됨(좌측)
- 하지만 2015년 이르면, (비록 언급 빈도수가 많은 편은 아니라 하더라도) 충분히 의미와 문맥을 가지고 신중하게 언급되기 시작함. 마을복지가 제대로 온라인 키워드로 등장하고 있다고 해석할 수가 있음.

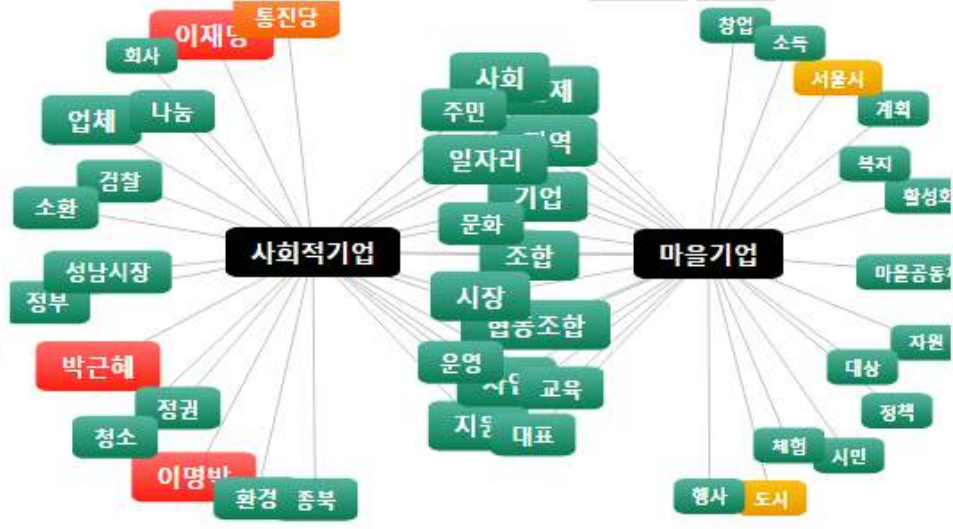
탐색어 : 사회적기업, 마을기업 | 기간 : 2014-12-01 ~ 2015-05-31

구버전

연관어

매체선택 전체 트위터 블로그 커뮤니티 뉴스 | 연관어 수 30 | 확인

- 전체
- 인물
- 단체
- 장소
- 상품
- 속성
- 브랜드
- 취미/여가
- 심리



No.	연관어	분류	탐색건수
1	기업	속성	8356
2	이재명	인물	5228
3	지역	속성	4819
4	지원	속성	4740
5	이명박	인물	4593

No.	연관어	분류	탐색건수
1	기업	속성	1649
2	조합	속성	1085
3	지역	속성	1080
4	협동조합	속성	997
5	사회적기업	속성	911

[주제어: 사회적 기업, 마을 기업 기간: 2014.12~2015.5, 데이터: 다음소프트]

- 유사하면서도 경쟁적인 키워드인 사회적 기업과 마을기업을 분석해봄. 두 개의 키워드 모두 공통적으로 일자리, 사회, 주민 등의 연관어를 공유하고 있고 전체 연관어의 절반이상을 공유하고 있는 점을 볼 때 온라인에서는 사회적 기업과 마을 기업을 하나의 부류로 인지하고 있다고 해석됨.
- 차이점은 사회적 기업이 주로 지자체와 연결되어 있는데 반해서 마을기업은 지자체 연관보다는 자율적 기업의 연관어들과 상당히 어울리고 있다는 점임. 이는 마을기업이 지자체의 정책적 지원을 적게 받는다고 해석할 수도 있고, 마을기업이 독립성을 지닌다고 해석할 수도 있음.

검색어: 젠트리피케이션 기간: 2015-01-01 - 2015-05-31

연관어

매제선택 전체 트위터 블로그 커뮤니티 뉴스 연관어 수 30

- 전체
- 인물
- 단체
- 장소
- 상품
- 속성
- 브랜드
- 취미/여가
- 심리



No.	연관어	분류	참석건수
1	지역	속성	5
2	도시	장소	4

[주제어: 젠트리피케이션 기간: 2014.12~2015.5, 데이터: 다음소프트]

- 2013년 상반기에는 전혀 온라인상에서 발견할 수 없던 키워드 ‘젠트리피케이션’이 2015년에는 비록 적은 빈도수지만 발견되고 있음.
- 특히 연관어들이 매우 다양한 내용으로, 비교적 정확한 문맥으로 연결되어 있다는 점을 볼 때 젠트리피케이션이 사회적 의제로 부상될 조짐이 확인 하다고 해석할 수 있음.

5. 온라인 빅데이터 분석에서 얻은 시사점

- 이상으로 다분히 실험적인 수준에서 사회혁신 키워드와 관련된 어휘들에 대해 온라인 빅데이터 분석을 시도해봄. 이 결과에서 가장 먼저 지적해야 하는 것은 온라인상에서 ‘사회혁신’이라는 키워드의 흔적이 거의 없다는 것임. 때문에 그래픽으로 보여줄 것도 없었음. (매년 9월 서울시에서

사회혁신관련 국제 컨퍼런스를 할 때에만 잠깐 나타나는 키워드) 이는 사회혁신 자체를 온라인에서 담론화 시켜야 할 필요가 있다는 것을 의미함.

- 가족에 대한 온라인 분석이 절실했는데, 가족은 워낙 광범위한 용도로 사용되고 있기 때문에 목적에 부합하는 방식으로 분석하기가 매우 난해하고 다양한 수준의 잡음이 개입됨. -여기서는 하지 못했음. 대신 1인 가족이나 고립, 잉여와 같은 대체 변수를 찾아서 분석해본 결과 사회혁신 키워드에서 발견했던 것과 유사한 결과를 얻었음.(이는 자존과 존엄에 대해서도 마찬가지로 있음.)
- 가족 이상으로 ‘공동체’나 ‘관계’에 대한 분석역시 대단히 광범위한 용도로 사용되어 쉽게 마이닝 하기가 만만치 않음. 여기서는 주로 마을 공동체를 대상으로 삼아 분석을 시도했음. 마을 공동체는 온라인상에서 ‘대세’라고까지 할 수는 없지만 안정적으로 논의되는 주제 가운데 하나로 자리 잡았음.
- 이는 도시와 관련된 혁신의제들인 ‘도시 재생’도 마찬가지임. 추세적으로, 그리고 양적으로 도시 재생이라는 키워드가 도시개발이나 뉴타운, 재개발이라는 키워드를 추월해가고 있음. 다만 그 어휘를 사용하면서 얼마나 혁신적인 의미를 담아내는지는 미지수임. 확실한 것은 최근 등장한 이슈인 젠트리피케이션까지 온라인상에 등장하면서 도시재생이 풍부해질 개연성을 보이고 있다는 점임.
- 다만 앞서 전제했던 것처럼, 데이터 활용대상이나 적용된 방법론이 매우 기초적인 수준에서 머물러 있다는 점을 인정할 수밖에 없음. 이와 관련된 최초의 시도라는 점에서 만족해야 한다는 뜻임. 이후에 더 활용도를 높여 나가면서 더 개선된 분석 방법론을 개발하고 축적할 필요가 있음.