

# 여론 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구

Spatiotemporal Analysis of  
Unstructured Big Data  
for Public Opinion Monitoring

황명화





# 여론 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구

Spatiotemporal Analysis of Unstructured Big Data  
for Public Opinion Monitoring

■ 연구진

연구책임 황명화 책임연구원

■ 외부연구진

이양원 국립부경대학교 교수

민보경 서울시립대학교 강사

■ 연구심의위원

사공호상 국토연구원 선임연구위원

임은선 국토연구원 연구위원

김대종 국토연구원 연구위원

김동한 국토연구원 책임연구원

# 발간사

2014년 학술논문 DB사이트 DB피아의 이용자들이 가장 관심을 많이 가진 3대 연구주제 중 하나가 ‘소셜네트워크서비스(Social Networking Service: SNS)’라고 합니다. 정보통신기술의 발달로 SNS를 이용한 사람들의 의사소통이 보편화되고, 이러한 변화상이 개인의 일상사와 더불어 사회 전반에 많은 영향을 미치고 있기 때문일 것입니다. SNS의 발달은 정책분야에서도 역시 많은 변화를 일으키고 있습니다. 청와대를 비롯한 많은 정부 부처와 기관에서 SNS로 대변되는 인터넷 소셜미디어를 활용하여 정책을 적극 홍보하고 국민과의 소통창구를 확대하고자 노력하고 있습니다. 하지만, 정부와 공공기관의 인터넷 소셜미디어 활용은 ‘관’에서 ‘민’으로의 일방적 정보전달 통로 수준으로만 유지되는 경향이 많아 아쉬움이 있습니다.

정책홍보와 소통창구로서의 역할 외에 인터넷 소셜미디어의 정책활용 방안은 무엇인가? 본 연구는 이 같은 의문으로 시작되었습니다. 스마트폰 1인당 1개 시대이며 스마트폰 보급률이 79.4%에 이르는 우리나라에서 인터넷 소셜미디어는 명실상부 국민여론이 형성되고 전파되는 새로운 공간입니다. 국민수요에 충실하고 국민이 체감할 수 있는 정책을 수립·집행하여 국민행복을 증대하고자 하는 우리나라의 현 국정기조에서 볼 때, 인터넷 소셜미디어라는 디지털 공간에서 형성되는 국민의 여론은 소중하고 가치있는 정보원이 아닐 수 없습니다.

인터넷 소셜미디어가 국민여론을 담고 있는 귀중한 자료원이라고 할 때, 이 자료원에서 생산되는 데이터를 어떻게 분석해야 국민과 지역의 정책수요를 보다 효과적으로 파악할 수 있을까? 본 연구가 던지는 두 번째, 그리고 핵심 질문입니다. 이 질문에 답하기 위해 본 연구는 국민여론은 시시각각 변화하며 이 변화 양상은 지역마다 장소마다 다르다는 점에 착안하여, 비정형 시공간분석 모델을 제안합니다. 인터넷 소셜미디어에서는 방대한 양의 데이터가 실시간으로 수집되어 비정형 빅데이터를 생산하기 때문입니다. 또한, 지역별로, 시기별로 변화하는 국민여론의 트렌드를 지도

분석과 같은 시각적인 방법으로 포착해낼 수 있다면, 국민이 그리고 지역에서 원하는 정책이 무엇인지 보다 시의적절하게 파악할 수 있기 때문입니다. 아무쪼록 이번 연구가 국민과 지역의 정책수요를 파악하는데 효과가 클 것으로 기대되는 소셜미디어 빅데이터의 시공간분석을 활성화하고 이의 정책 적용방안에 대한 논의까지도 이끌어 낼 수 있는 계기가 되기를 기대합니다. 끝으로 본 연구를 수행하는데 노력을 아끼지 않은 황명화 책임연구원의 노고를 치하하며 외부연구진으로 참여해 주신 이양원 교수님과 민보경 강사님께도 깊이 감사드립니다.

2015년 1월

국토연구원장 김 경 환

## 주요 내용 및 정책제안

### 본 연구보고서의 주요 내용은 다음과 같음

- ① 수요 맞춤형 정책발굴을 위한 비정형 빅데이터 분석의 정책활용은 꾸준히 증가하고 있지만 현재까지는 대부분 전국적 여론 트렌드의 수시 분석에 머물러, 지역별, 시기별 여론 트렌드의 세밀한 변화를 파악하기 어려움
- ② 인터넷 소셜미디어에 반영된 여론의 시공간적 변화 트렌드를 파악할 수 있는 새로운 분석방법론으로서, 비정형 빅데이터, 텍스트마이닝, 시공간분석 등을 융합활용하여 지도 기반으로 분석하는 비정형 빅데이터 시공간분석 모델 제안
- ③ 비정형 빅데이터 시공간분석 모델은 비정형 빅데이터의 수집, 비정형 빅데이터의 전처리, 비정형 빅데이터의 시공간분석의 세 단계로 구성
- ④ 비정형 빅데이터는 인터넷 뉴스미디어와 소셜미디어에 크롤링, Open API 등의 방법을 적용하여 수집할 수 있으며, 데이터의 품질개선과 공간분석을 위해 불필요 자료 제거, 지오코딩, 문장정제 등의 전처리를 거쳐야 함
- ⑤ 비정형 빅데이터의 시공간분석을 위해 여론의 내용과 형성주체에 초점을 둘 수 있으며, 구체적으로 단어사용패턴과 토픽/감성의 시공간 분석, 온라인 사회관계망의 지리적 분석 등의 기법을 활용할 수 있음

### 본 연구보고서의 정책제안은 다음과 같음

- ① 비정형 빅데이터를 활용한 여론 분석을 통해 국민수요와 지역 맞춤형 정책을 발굴하려는 정부/공공기관이 증가하고 있으므로, 공공부문에서 공동활용할 수 있는 여론 모니터링 체계를 구축하여 정책역량 강화 및 예산 중복투자 예방
- ② 비정형 빅데이터 시공간분석 모델을 「정책지도」, 「공간 빅데이터 체계 구축 사업」 등의 국가 및 지방 사업과 연계하여, 정책지도의 다양성을 증대하고 여론 관련 공간빅데이터의 활용성 제고
- ③ 비정형 빅데이터 시공간분석 모델의 정책활용을 위해 중앙정부 및 지자체 내 관련 조직(정책기획과, 국민소통과, 주민참여과 등) 간 협력체계를 구축하고, 정책수요 발굴시 비정형 빅데이터 활용을 권고하는 업무지침 마련

# 요약

## 1. 연구의 배경 및 목적

- 국민행복을 추구하는 국정기조에 따라 수요 및 지역맞춤형 정책의 중요성이 부각되면서, 국민의 정책수요를 시의성있게 파악할 수 있는 정책수단의 필요성이 증대하고 있음
  - 국민의 정책수요를 파악하기 위해 국민여론을 직·간접적으로 청취하고 이를 정책 형성에 반영하는 노력이 증대하고 있음
  - 특히, 정보통신기술의 발달에 따른 국민들의 의사소통방식 변화를 고려하고자 인터넷 소셜미디어로 수집된 비정형 빅데이터를 분석하여 국민여론의 트렌드와 그 동태를 모니터링하려는 움직임이 커지고 있음
  - 비정형 빅데이터 분석의 정책활용은 꾸준히 증가하고 있지만 현재까지는 대부분 전국적 여론 트렌드의 수시 분석에 머물러, 지역별, 시기별 여론 트렌드의 세밀한 변화를 파악하기 어려우며 이를 위한 새로운 분석 접근법 필요
- 본 연구에서는 인터넷 소셜미디어에 반영된 국민여론의 시공간적 변화 트렌드를 파악할 수 있는 새로운 분석방법론을 개념적 수준에서 개발하고 이의 정책적 함의를 밝히고자 함
  - 비정형 빅데이터, 텍스트마이닝, 시공간분석 등의 다양한 데이터와 분석기법을 융합 활용하여, 인터넷 소셜미디어에 반영된 국민여론의 시공간적 패턴을 지도를 표출할 수 있는 분석기법들을 살펴보고, 이들의 정책활용 방향을 제시함

## 2. 여론 모니터링과 비정형 빅데이터 분석

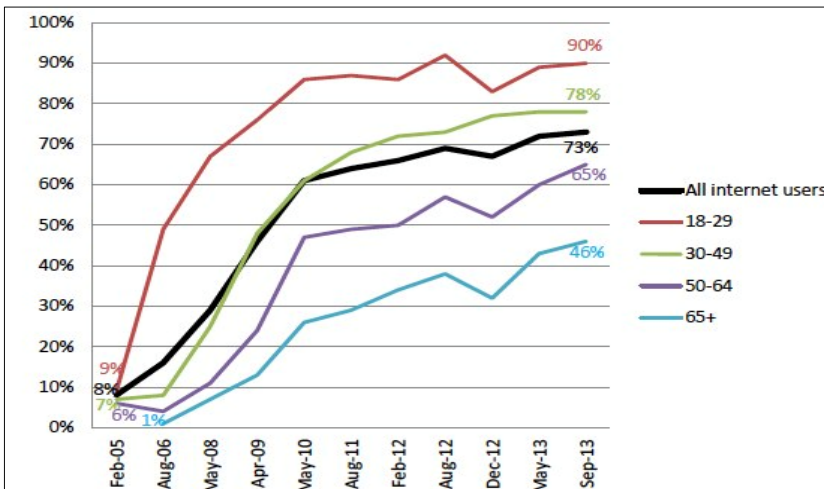
### □ 여론 모니터링의 개념과 방법

- 여론(Public Opinion)이란 대중(the mass) 혹은 공공 사안에 관여하는 개인들의 집단적 의견을 말함
- 여론 모니터링이란 여론이 어떠한지, 어떻게 변화하고 있는지 그 동향을 지속적으로 조사·분석하여 정보와 지식으로 변환하는 일련의 활동을 뜻함
- 여론 모니터링의 방법은 분석의 대상이 되는 여론의 속성이 무엇인가에 따라, 여론내용에 관한 분석방법들과 여론의 형성주체에 관한 분석방법들로 나뉨

### □ 인터넷 소셜미디어 기반의 여론 부상

- 1980~90년대의 PC시대에서 인터넷 시대와 모바일 시대를 거쳐 지금은 스마트 시대로 나아가고 있음
- 인터넷 패러다임의 변화가 스마트 폰 환경으로 전이되면서, 최근 블로그, 소셜네트워크서비스(SNS), 위키 등의 소셜미디어 활용이 20~40대의 젊은 층을 중심으로 급속히 증대하고 있으며 새로운 국민여론 형성의 공간으로 부상함

<인터넷 사용자 중 연령대별 소셜미디어 이용률의 변화>

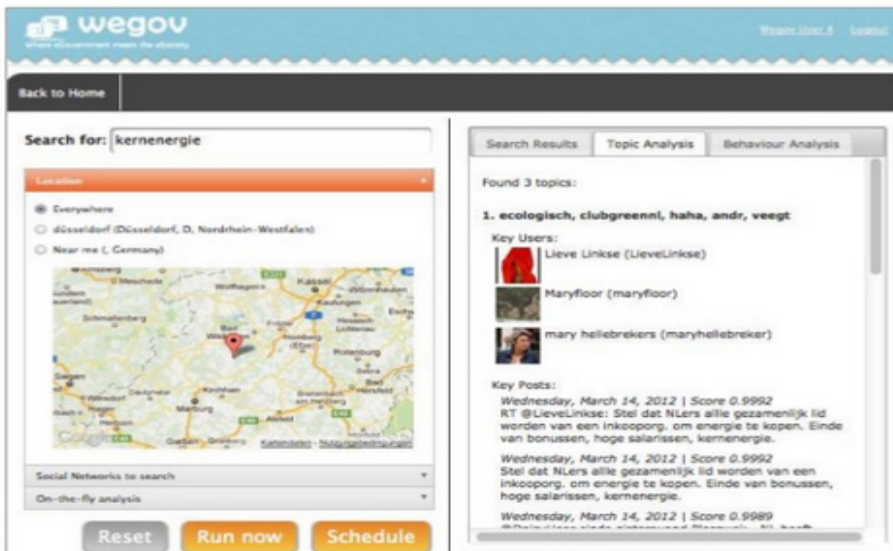


출처: <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet> (2015년 1월 6일 오후 2시 10분 최종 접속)

□ 여론 모니터링을 위한 비정형 빅데이터 분석의 활용

- 인터넷 소셜미디어 기반의 여론이 부상하면서, 이들 신규매체에서 대량으로 생산되는 비정형 빅데이터를 여론 모니터링을 위해 활발히 사용함
- 미국에서는 트위터의 실시간 분석을 통해 감기 트렌트를 모니터링하여 보건정책에 활용하였으며, 유럽에서는 소셜미디어 데이터 분석을 정책의 다양한 분야에서 활용할 수 있도록 WeGov 등과 같은 툴박스를 제작하여 제공함
- 국내에서는 2013년 문화체육관광부에서 빅데이터 분석을 통해 국민인식 및 여론변화 추이를 분석한 바 있으며, 국토연구원과 국립재난안전연구원에서는 빅데이터를 활용한 맞춤형 부동산/도시재난/복지 정책 발굴을 모색하였음
- 그러나, 선행 연구에서 활용한 비정형 빅데이터 분석방법들은 국민여론의 공간적 분포와 시계열 변이를 추적하지 못해, 국민수요 및 지역 맞춤형 정책지원에 한계가 있음

<WeGov 툴박스>



출처: [http://www.slideshare.net/timo\\_wandhoefer/wegov-user-guide-toolbox-25](http://www.slideshare.net/timo_wandhoefer/wegov-user-guide-toolbox-25) (2015년 1월 6일 오후 2시 10분 최종 접속)

### 3. 여론 모니터링을 위한 비정형 빅데이터 시공간분석 모델

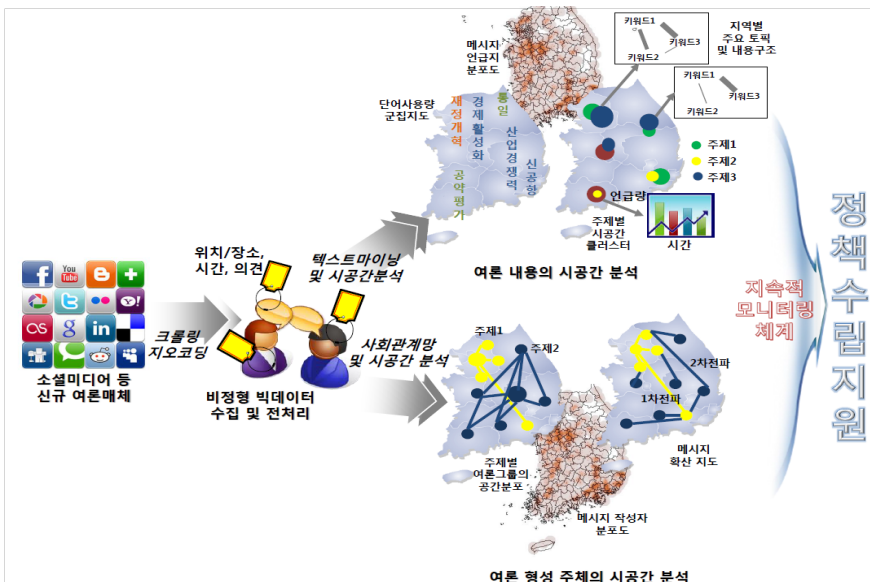
#### 1) 기본방향 및 개념

□ 본 연구에서는 지역 및 수요 맞춤형 정책을 지원하기 위해, 비정형 빅데이터를 위한 새로운 분석방법인 비정형 빅데이터 시공간분석 모델을 제안함

□ 비정형 빅데이터 시공간분석 모델의 기본방향과 개념

- (기본방향) 인터넷 소셜미디어 상의 국민여론이 지역별, 시기별로 변화하는 패턴을 체계적으로 분석하고 지속적으로 모니터링하여, 국민수요 및 지역 맞춤형 정책수요 발굴을 지원하고 여론 기반 정책수립의 시의성을 제고함
- (개념) 인터넷 소셜미디어에서 생성되는 비정형 빅데이터에 크롤링, 텍스트마이닝, 시공간분석 등의 방법을 함께 적용하여, 인터넷 소셜미디어에 반영된 국민여론의 시공간적 변화패턴을 분석하는 자료 중심의 여론 모니터링 방법

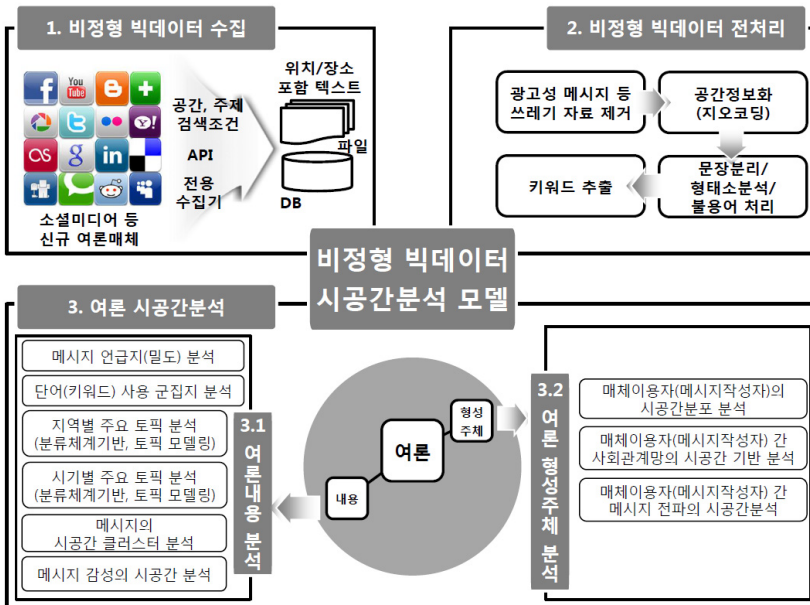
#### <여론 모니터링을 위한 비정형 빅데이터 시공간분석 모델의 개념>



## 2) 전체 구성

- 비정형 빅데이터 시공간분석 모델은 ① 비정형 빅데이터의 수집, ② 비정형 빅데이터의 전처리, ③ 비정형 빅데이터의 시공간분석의 세 단계로 구성
  - 비정형 빅데이터 수집 단계에서는 API나 유·무료 소프트웨어를 이용하여, 트위터, 페이스북, 블로그, 카페, 온라인 신문 등 다양한 인터넷 소셜미디어 매체에서 시민들이 논의한 내용을 지속적으로 수집·저장·관리
  - 비정형 빅데이터 전처리 단계에서는 수집된 여론 데이터에서 광고성 메시지, 욕설, 중복 메시지 등을 제거하고, 메시지를 관련된 위치/장소 정보와 연계하며 (공간정보화 혹은 지오코딩), 메시지를 정제하여 키워드를 추출
  - 비정형 빅데이터 시공간분석 단계에서는 공간분포분석, 시계열 데이터 분석, 텍스트마이닝, 사회관계망분석 등을 융복합 활용하여, 여론 내용과 형성주체의 시공간적 변화 패턴을 탐색

<여론 모니터링을 위한 비정형 빅데이터 시공간분석 모델의 전체 구성>



### 3) 비정형 빅데이터의 수집과 전처리

#### □ 비정형 빅데이터의 수집

- 여론 모니터링을 위한 비정형 빅데이터(여론 빅데이터)는 인터넷 뉴스미디어와 카페, 블로그, 트위터, 페이스북 등의 소셜미디어 서비스에서 수집 가능
- 자료 수집 전 인터넷 소셜미디어의 이용자 프로파일(인터넷, 모바일 서비스 등을 구매할 수 있는 경제력을 갖춘 20~40대의 도시 거주자)이 여론 모니터링의 주요 대상으로 적합한지 사전에 고려해야 함
- 여론 빅데이터의 수집을 위해서는 크롤링, FTP, Open API, RSS, 스트리밍 기술 등을 활용할 수 있음

#### □ 비정형 빅데이터의 전처리

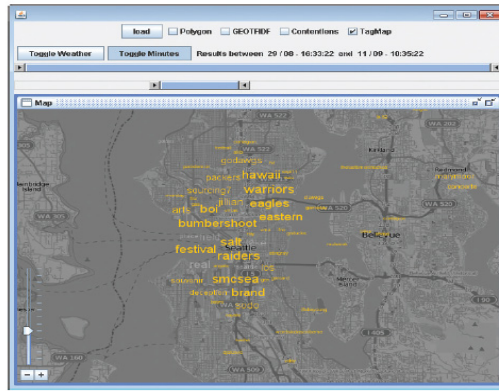
- 여론 빅데이터의 품질개선을 위해 분석에 부적합하거나 중복된 데이터, 특히 광고성 및 악의성 스팸을 사용자, 단어, URL 등의 반복패턴을 분석하여 제거
- 시공간분석을 위해 GPS좌표, 지명/장소 데이터베이스, 사회관계망 기반 추정 등을 이용하여 여론 관련 메시지를 위치/장소와 연계(지오코딩 혹은 공간정보화)
- 공간정보화된 메시지는 문장 단위에서 단어분리, 형태소분석, 불용어처리, 키워드 추출 등의 과정을 통해 주요 키워드 셋트로 정제

### 4) 여론 내용의 시공간분석

#### □ 메시지 작성위치 및 단어사용 (시공간) 군집 분석

- 여론 내용 분석의 첫 단계는 수집된 메시지들의 공간분포를 분석하는 것으로, 메시지 위치도, 메시지 건수의 단계구분도, 밀도 분석 등의 방법 적용 가능
- 어디에서 무슨 이슈가 주로 논의됐는지를 파악하기 위해, 주요 키워드가 빈번히 사용되는 대표지점을 찾아 키워드를 표시하는 단어사용 군집지 분석 이용 가능

### <단어사용 군집지 분석결과와 예시>

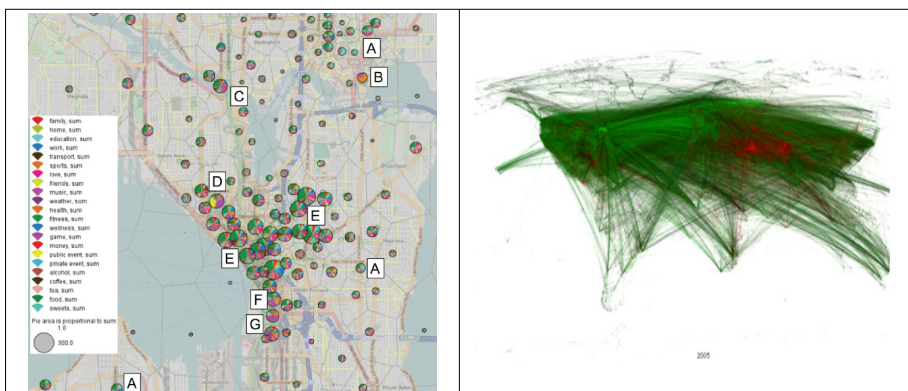


출처: Andrienko 외 2013, Thematic Patterns in Georeferenced Tweets Through Space-Time Visual Analytics, 4 페이지

### □ 여론 트렌드(토픽) 및 감성의 시공간 분석

- 여론 트렌드의 시공간적 경향성을 파악하기 위해, 수집된 메시지들을 지역별로 분류하고 그 의미와 맥락에 따라 특정 주제(토픽)로 나누어 주제별 구성의 시공간적 변화를 살펴볼 수 있음
- 여론 트렌드에 반영된 사람들 감성이 어떠한 시공간 분포를 보이는지 특정 지역과 장소에 관한 사람들의 평판은 어떠한지를 파악하기 위해, 메시지 및 장소별 감성 점수를 계산하여 지도로 표출 가능

### <여론 토픽(좌) 및 감성(우)의 시공간분석 결과 예시>



출처: Andrienko 외 2013, Thematic Patterns in Georeferenced Tweets Through Space-Time Visual Analytics, 6 페이지

출처: Leetaru, K. H. 2012. Culturomics 2.0: Forecasting Large-scale Human Behavior Using Global News Media Tone in Time and Space.

## 5) 여론 형성주체의 시공간분석

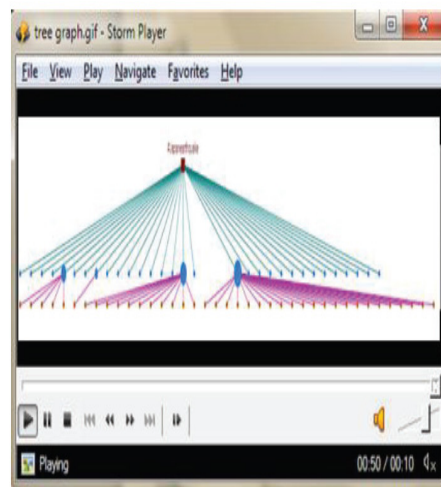
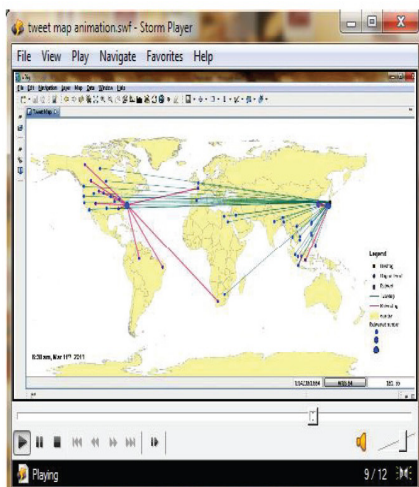
## □ 메시지 작성자의 위치 및 사회관계망의 공간분포 분석

- 인터넷 소셜미디어 상에서 여론 형성을 주도하는 사람들이 실제 현실공간에서 어떻게 분포하는지를 이해하기 위해, 메시지 작성자의 위치, 메시지 작성자별 메시지 작성지점들의 대표 위치나 범위를 분석할 수 있음
- 인터넷 소셜미디어 공간에서 벌어진 사람들과의 상호작용, 즉 사회관계망이 실제 현실공간에서는 어떤 형태를 띠는지를 보기위해 사회관계망에 참여한 사람들의 공간적 분포를 네트워크 형태로 시각화하여 분석할 수 있음

## □ 메시지 전파의 시공간분석

- 여론 공간 상의 사회관계망을 공간정보화하고 이를 상호작용(메시지)의 유형에 따라 분류해보면, 특정 메시지나 이슈가 어떤 사람과 장소를 매개로 어떻게 지리적으로 확산되는지 그 패턴을 파악할 수 있음

### <사회관계망 기반 메시지의 공간적 확산 분석의 예시>



파란색과 분홍색 선은 각각 메시지의 1차, 2차(리트윗) 전파를 나타냄

출처: Ma, D. 2013. Visualization of Social Media Data: Mapping Changing Social Networks. 43 페이지.

#### 4. 결론 및 향후과제

- 본 연구에서 제안한 비정형 빅데이터 시공간분석 모델은 인터넷 소셜미디어에서 형성되는 여론의 지역적, 시간적 변화 동태를 파악할 수 있는 방법을 제공하므로, 지역 및 수요 맞춤형 정책수요 발굴을 위한 정책수단으로 활용할 수 있을 것임
- 본 연구의 결과를 토대로 한 정책제언은 아래와 같음
  - 비정형 빅데이터를 활용한 여론 분석을 통해 국민수요와 지역 맞춤형 정책을 발굴하려는 정부/공공기관이 증가하고 있으므로, 공공부문에서 공동활용할 수 있는 여론 모니터링 체계를 구축하여 정책역량 강화 및 예산 중복투자 예방
  - 비정형 빅데이터 시공간분석 모델을 「정책지도」, 「공간 빅데이터 체계 구축 사업」 등의 국가 및 지방 사업과 연계하여, 정책지도의 다양성을 증대하고 여론 관련 공간 빅데이터의 활용성 제고
  - 비정형 빅데이터 시공간분석 모델의 정책활용을 위해 중앙정부 및 지자체 내 관련 조직(정책기획과, 국민소통과, 주민참여과 등) 간 협력체계를 구축하고, 정책수요 발굴시 비정형 빅데이터 활용을 권고하는 업무지침 마련
  - 비정형 빅데이터 시공간분석 모델의 기본 원리, 적용 방법, 효과 등에 대해 체계적 교육과 홍보를 실시하고, 활용사례를 적극 개발·공유하여 여론 기반 정책수요 발굴을 위한 첨단 분석기법을 확산
- 연구의 한계와 향후 과제
  - (한계) 2개월의 짧은 연구기간으로 인해 비정형 빅데이터 시공간분석 모델의 정책적 적용을 위한 상세 방안 및 실증 사례를 발굴하지 못함
  - (과제) 향후 비정형 빅데이터 시공간분석 모델의 실증과 적용을 위한 구체적인 정책연구를 수행해야 하며, 비정형 빅데이터의 공간정보화 및 공간데이터와의 연계 분석과 같은 방법론적 개선방안도 추가로 연구할 필요가 있음

# 차례

|                                                 |           |
|-------------------------------------------------|-----------|
| 발 간 사 .....                                     | i         |
| 주요 내용 및 정책제안 .....                              | iii       |
| 요 약 .....                                       | iv        |
| <b>I. 서론 .....</b>                              | <b>3</b>  |
| 1. 연구의 배경과 목적 .....                             | 3         |
| 2. 연구의 범위와 방법 .....                             | 5         |
| 3. 선행연구 현황 및 연구의 차별성 .....                      | 7         |
| <b>II. 여론 모니터링과 비정형 빅데이터 분석 .....</b>           | <b>13</b> |
| 1. 여론 모니터링의 개념 .....                            | 13        |
| 1) 여론 모니터링의 개념 .....                            | 13        |
| 2) 여론 모니터링의 등장배경 .....                          | 15        |
| 3) 여론 모니터링의 방법 .....                            | 17        |
| 2. 여론 모니터링과 비정형 빅데이터 .....                      | 20        |
| 1) 인터넷 소셜미디어 기반의 여론 부상 .....                    | 20        |
| 2) 여론 모니터링을 위한 비정형 빅데이터의 활용 .....               | 23        |
| 3) 비정형 빅데이터 기반 여론 모니터링 방법과 활용 사례 .....          | 24        |
| 3. 시사점 .....                                    | 32        |
| <b>III. 여론 모니터링을 위한 비정형 빅데이터 시공간분석 모델 .....</b> | <b>37</b> |
| 1. 기본방향 및 개념 .....                              | 37        |
| 1) 기본방향 .....                                   | 37        |
| 2) 개념 .....                                     | 39        |
| 2. 여론 모니터링을 위한 비정형 빅데이터 시공간분석 모델 .....          | 40        |

|                                  |           |
|----------------------------------|-----------|
| 3. 여론 관련 비정형 빅데이터의 수집과 전처리 ..... | 43        |
| 1) 여론 관련 비정형 빅데이터의 수집 .....      | 43        |
| 2) 여론 관련 비정형 빅데이터의 전처리 .....     | 46        |
| 4. 여론의 시공간분석 .....               | 50        |
| 1) 여론 내용의 시공간분석 .....            | 50        |
| 2) 여론 형성주체에 관한 시공간분석 .....       | 63        |
| <b>IV. 결론 및 향후과제 .....</b>       | <b>71</b> |
| 1. 연구의 결과 및 정책제언 .....           | 71        |
| 1) 연구의 결과와 의의 .....              | 71        |
| 2) 정책 제언 .....                   | 72        |
| 2. 연구의 한계와 향후 과제 .....           | 74        |
| 1) 연구의 한계 .....                  | 74        |
| 2) 향후 과제 .....                   | 74        |
| 참고문헌 .....                       | 77        |
| SUMMARY .....                    | 83        |
| 부록1. 시민참여형 여론모니터링 방법과 사례 .....   | 85        |

|                                      |    |
|--------------------------------------|----|
| 〈표 1-1〉 선행연구와의 차별성 .....             | 9  |
| 〈표 2-1〉 정책 거버넌스의 발전 .....            | 15 |
| 〈표 2-2〉 e-Governance의 발전단계와 특성 ..... | 16 |
| 〈표 2-3〉 인터넷 패러다임의 진화 .....           | 20 |
| 〈표 3-1〉 여론 빅데이터 공급원의 종류 및 예시 .....   | 44 |
| 〈표 3-2〉 여론 빅데이터 수집 기술 .....          | 45 |
| 〈표 3-3〉 비정형 빅데이터 수집 특성 .....         | 45 |
| 〈표 3-4〉 여론 자료 관련 위치/장소 정보의 유형 .....  | 47 |

|                                                     |    |
|-----------------------------------------------------|----|
| 〈그림 1-1〉 연구의 수행 체계 .....                            | 6  |
| 〈그림 2-1〉 여론의 속성 .....                               | 17 |
| 〈그림 2-2〉 여론 모니터링 방법의 종류 .....                       | 18 |
| 〈그림 2-3〉 인터넷 사용자 중 연령대별 소셜미디어 이용률의 변화 .....         | 21 |
| 〈그림 2-4〉 미국 시민참여국 홈페이지 .....                        | 22 |
| 〈그림 2-5〉 폭염 연관어 분석 결과 .....                         | 26 |
| 〈그림 2-6〉 키워드빈도 시계열 분석의 사례 .....                     | 26 |
| 〈그림 2-7〉 전기부족 관련 감성분석 결과 .....                      | 27 |
| 〈그림 2-8〉 WeGov 툴박스 .....                            | 29 |
| 〈그림 2-9〉 복지정책 분야에 대한 국민 관심도 .....                   | 29 |
| 〈그림 2-10〉 ‘대구시’ 검색어를 이용한 서치 네트워크 .....              | 31 |
| 〈그림 2-11〉 ‘농림수산식품부’ 트위터계정의 팔로어(follower) 네트워크 ..... | 31 |
| 〈그림 3-1〉 비정형 빅데이터 시공간분석 모델의 기본방향 .....              | 38 |
| 〈그림 3-2〉 비정형 빅데이터 시공간분석 모델의 개념 .....                | 40 |
| 〈그림 3-3〉 비정형 빅데이터 시공간분석 모델의 전체 구성 .....             | 41 |
| 〈그림 3-4〉 텍스트 데이터의 지오코딩 방법 .....                     | 48 |
| 〈그림 3-5〉 텍스트 데이터의 전처리 과정 .....                      | 49 |
| 〈그림 3-6〉 비정형 빅데이터 전처리 결과 .....                      | 50 |
| 〈그림 3-7〉 서울시 트윗의 위치도 및 위치건수의 단계구분도 예시 .....         | 51 |
| 〈그림 3-8〉 서울지역 트윗의 공간밀도 및 등치선도의 예시 .....             | 52 |
| 〈그림 3-9〉 단어사용 군집지 분석 결과의 예시 .....                   | 53 |
| 〈그림 3-10〉 단어사용 군집지 분석의 원리 .....                     | 54 |
| 〈그림 3-11〉 분류체계 기반 여론 토픽 분석의 원리 .....                | 55 |
| 〈그림 3-12〉 분류체계 기반 여론 토픽 분석 결과의 예시 .....             | 56 |
| 〈그림 3-13〉 LDA 토픽 모델링 알고리즘의 문서-주제 모델 .....           | 57 |
| 〈그림 3-14〉 토픽모델링 기반 여론 내용 분석의 원리 .....               | 57 |
| 〈그림 3-15〉 여론 토픽 내용구조의 지역별 차이 .....                  | 58 |
| 〈그림 3-16〉 토픽모델링을 활용한 Science 논의내용의 경향분석 .....       | 59 |

|                                              |    |
|----------------------------------------------|----|
| 〈그림 3-17〉 여론 메시지의 시공간 클러스터 분석 결과 .....       | 60 |
| 〈그림 3-18〉 여론 감성 공간분포 분석의 원리 .....            | 61 |
| 〈그림 3-19〉 여론 감성 공간분포 분석의 예시 .....            | 62 |
| 〈그림 3-20〉 메시지 작성자의 공간분포 분석을 위한 기초자료 .....    | 63 |
| 〈그림 3-21〉 사용자별 메시지 작성 대표위치 분석의 예시 .....      | 64 |
| 〈그림 3-22〉 사용자 위치자료와 사회관계망 자료의 연계 .....       | 66 |
| 〈그림 3-23〉 사회관계망의 지리적 네트워크 분석 예시 .....        | 66 |
| 〈그림 3-24〉 사회관계망 내 소그룹들의 공간분포 비교·분석의 예시 ..... | 67 |
| 〈그림 3-25〉 사회관계망 기반 메시지 확산 분석의 예시 .....       | 68 |



chapter I

서론



# 서론

국민행복을 추구하는 국정기조에 따라 수요맞춤형 정책의 중요성이 부각되면서, 국민의 정책수요를 시의성 있게 파악할 수 있는 정책수단의 필요성이 증대하고 있다. 이 장은 국민의 정책수요를 이해하기 위해 인터넷, 소셜미디어 등 신규 매체에 반영된 여론을 모니터링할 수 있는 비정형 빅데이터 시공간분석 모델을 개발하는 연구의 수행 배경, 목적, 방법 및 선행연구와의 차별성을 제시하였다.

## 1. 연구의 배경과 목적

2014년 6월 4일은 제6회 전국 동시지방선거가 시행된 날이다. 이번 선거에서 흥미로운 대목은 미시 공간과 관련된 선거공약이 많았다는 점인데, “동네에서 먹고 살자”, “골목까지 행복한 복지국가” 등이 그 예이다. “동네”, “골목”과 같은 주민의 생활공간과 함께 정책관계자들은 주민 생활의 질에도 주목하고 있다. 최근 정부에서 주도하고 있는 “지역행복생활권”, “지역행복지표”, “국토체감지표” 등의 아젠다는 국가정책의 초점이 지역주민 개개인의 행복증대와 생활만족으로 옮겨갔으며, 이는 곧 지역과 그 주민의 특성과 수요에 최적화된 정책의 발굴과 시행이 정책 입안자와 실무자의 주요 현안임을 반영한다.

지역맞춤형 및 수요맞춤형 정책의 부상으로 국민과 지역의 정책수요를 시의성 있게 발굴하고 그 수립을 지원할 수 있는 정책수단에 대한 수요가 증가하고 있다. 과거에는 국가 이념, 전문가 의견, 총량적 통계를 토대로 국가정책 수요가 조사·확정된

후 지방자치단체로 전달되는 성향이 강했다. 그러나 이러한 하향식 방법은 주민수요보다는 정치적 목표에 기반한 정책의 우선 추진(4대강 개발 등), 정책적 의사결정 과정의 불투명성 등 여러 문제점을 낳았다. 최근 쌍방향 소통, 시민참여, 협력·합의를 통한 정책수립 패러다임이 부각되어, 하향식 정책과정을 국민 의견의 직·간접적 수렴을 통해 보완하려는 움직임이 커지고 있다.

국가와 지역의 정책과정에서 정부가 국민 여론을 청취하고 이를 정책 형성에 반영하는 것은 새로운 일이 아니다. 정부는 정책의제를 설정하고 구체적 정책을 결정할 때 설문조사, 여론조사, 정책포럼, 자문위원회, 정책협의회, 세미나, 간담회, 정책토론회 등의 방법을 통해 정책과정의 시민참여를 허용하고 있다. 그러나, 직접적 시민참여 기반의 정부-시민 협의 수단의 경우, 국민 참여가 형식적이기 쉽고, 참여자 수가 제한될 뿐만 아니라 참여자 선정에 유관관계자의 의견과 이해가 반영되어 수집여론의 대표성이 떨어진다는 단점이 있다. 더불어, 실제로 시민참여의 장을 마련하는 준비절차가 단순하지 않고 상당한 비용이 소요되어 정부-시민 간 직접적 협의는 그 시행횟수가 낮은 편이다. 결국, 전통적 정부-시민 협의 방법들은 주요 정책의제에 대한 공동 이해의 토대를 다지는 데에는 적합하나, 변화하는 국민여론 트렌드를 시의성 있게 파악하여 국민과 지역의 수요에 맞춤형된 정책의 발굴에는 그 효용성이 떨어진다고 하겠다.

전통적 정부-시민 협의 방법을 보완하고 정보통신기술의 발달에 따른 국민들의 의사소통방식 변화를 고려하고자, 최근 인터넷, 소셜네트워크서비스(Social Networking Services; SNS) 등 신규매체로 표출된 국민여론을 분석하고 그 동태를 모니터링하는 움직임이 커지고 있다. 예를 들어, 문화체육관광부는 「빅데이터 분석을 통해 본 2013년 국민 인식 변화」(2013) 과제에서 다양한 소셜미디어 빅데이터를 분석하여, 국민생활 트렌드와 정책관심도를 파악하고 국민관심도를 반영한 정책 방향을 제시하고자 하였다. 문화체육관광부와 같이 신규매체에서 수집되는 비정형 빅데이터를 정책에서 활용하려는 시도가 늘어나는 이유는, 이들 데이터가 시민들이 일상사 중 자발적으로 공유한 의견을 대량으로 포함하고 있어, 데이터 분석시 국민여론의 정확한 경향성을 파악할 수 있기 때문이다.

비정형 빅데이터 분석의 정책분야 활용사례는 꾸준히 증가하고 있지만, 현재까지는 대부분 상용서비스를 이용한 전국적 여론 트렌드의 수시 분석에 머물고 있다. 그러나, 이들 상용서비스들은 대체로 데이터 수집, 가공, 분석 과정을 투명하게 공개하지 않고, 키워드 분석, 감성 분석, 연관어 분석 등으로 제한된 범위의 분석기법을 제공하며, 분석결과의 의미 해석도 용이하지 않다는 한계점이 있다. 수요맞춤형, 특히 지역맞춤형 정책 발굴의 경우, 지역별 여론 트렌드와 그 변화가 중요하여 비정형 빅데이터 분석에 대한 시공간적 접근법이 필요하다. 따라서, 기존 상용서비스들은 지역맞춤형 정책 수요발굴에 적용하기 어려우며, 새로운 방법론이 필요하다.

이러한 배경으로 이 연구는 비정형 빅데이터 분석을 활용한 기존의 정책수요 발굴 방법을 보완하고자 한다. 구체적으로, 이 연구의 목적은 인터넷, 소셜미디어 등 신규매체에 반영된 여론의 시공간적 변화 트렌드를 파악할 수 있는 새로운 분석방법을 개념적 수준에서 개발하고 이의 정책적 함의를 밝히는 것이다. 즉, 비정형 빅데이터, 텍스트마이닝, 시공간분석 등을 융합하여, 신규매체에 반영된 여론의 시공간패턴을 지도로 표출하는 분석기법들을 살펴보고, 이들의 정책활용 방향을 제시하고자 한다.

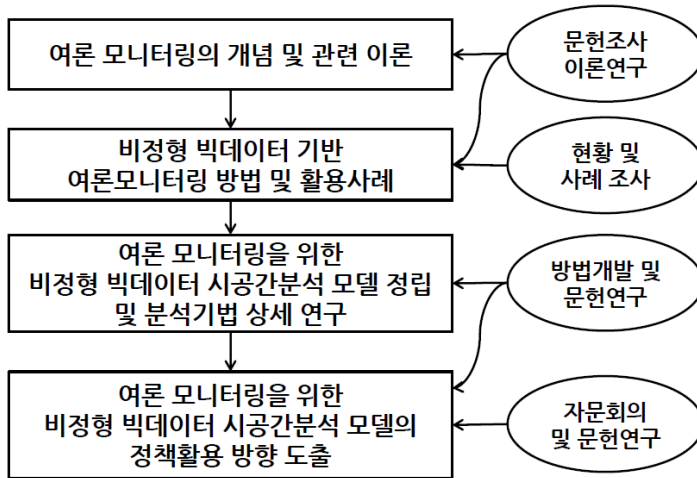
## 2. 연구의 범위와 방법

이 연구의 내용은 크게 세 가지 범위로 나누어진다. 첫째, 여론 모니터링의 개념과 등장배경을 이론적으로 살펴보고, 비정형 빅데이터를 활용한 여론 모니터링 방법과 그 활용 사례들을 검토한다. 둘째, 비정형 빅데이터 기반으로 국민여론을 시공간적으로 분석하는 방법론의 개념과 모델을 정립한다. 이 때, 여론의 시공간 분석방법들을 여론 내용과 형성 주체에 대한 기법들로 분류하여 구체적 적용 방법과 사례들을 제시한다. 셋째, 비정형 빅데이터 기반 시공간분석 방법들의 정책활용 가능성 및 그 방향에 대해 제언하고, 본 연구의 한계와 향후 과제를 논의한다.

〈그림 1-1〉은 이 연구의 수행방법과 수행체계이다. 우선, 문헌조사와 이론연구를 통해 본 연구에 적용되는 여론 모니터링의 개념을 정립하고, 그 등장배경 및 분석 프레임워크를 살펴보았다. 두 번째로, 인터넷, 소셜미디어 등 신규매체의 등장과

그로 인한 국민의사소통 양식의 변화 및 비정형 빅데이터의 출현, 여론 모니터링에 있어 비정형 빅데이터를 활용할 때 기대되는 효과 등을 이론적으로 고찰하였다. 다음으로, 비정형 빅데이터 기반 여론 모니터링 방법들과 그 특징 및 활용사례를 조사한 후, 이들 방법들의 개선을 위한 시사점을 도출하였다.

<그림 1-1> 연구의 수행 체계



다음으로, 비정형 빅데이터를 이용하여 여론 트렌드의 시공간적 변화를 모니터링할 수 있는 새로운 분석방법론을 개발하고 그 구체적 적용 방안 및 사례들을 예시의 형태로 제시하였다. 먼저, 새로운 시공간 분석방법론의 개념 및 기본방향을 정립하고, 이론적 수준에서 분석방법론 개발 모델과 그 구성 체계를 정의하였다. 본 연구에서 제시하는 분석방법론의 실제적 이해를 위해, 분석방법론을 구성하는 상세 기법들을 여론의 내용과 형성주체를 대상으로 하는 두 그룹으로 분류하고 그 실제 적용 방법과 결과의 예시들을 문헌연구를 통해 정리하였다.

마지막으로, 비정형빅데이터 및 공간분석 전문가 등과의 자문회의 및 문헌조사를 통해 본 연구에서 제시한 시공간 분석방법론의 정책활용 가능성을 진단하고, 그 방향에 대해 제언하였다. 또한, 본 연구 및 그 결과의 한계를 설명하고, 향후 추진해야 할 연구과제들을 도출하였다.

### 3. 선행연구 현황 및 연구의 차별성

이 연구와 관련된 선행 연구는 비정형 빅데이터의 정책 활용을 모색한 연구를 살펴보았으며, 크게 ‘비정형 빅데이터의 국토정책 활용방안’에 관한 연구와 ‘비정형 빅데이터 기반 재난관리정책 고도화 방안’에 관한 연구로 구분할 수 있다.

‘비정형 빅데이터의 국토정책 활용방안’에 관한 연구 중에서는 김대중·윤서연(2013)의 「빅데이터를 활용한 국토정책반응 모니터링 및 정책수요 예측방안」, 이영주 외(2013)의 「공간 빅데이터 기반의 복지·공간 융합모델 시범연구」, 김미정 외(2013)의 「과학적 국토정책을 위한 공간빅데이터 활용방안」을 구체적으로 살펴 보았다.

김대중·윤서연(2013)의 연구는 다양한 빅데이터의 국토정책 활용방안을 제시하였는데, 비정형 빅데이터의 이용 측면에서는 소셜미디어, 인터넷 뉴스, 블로그 등의 비정형 빅데이터와 공간빅데이터의 연계 분석을 통해 부동산정책에 대한 국민반응을 모니터링하고 미래 정책수요를 예측하는 방안을 논의하였다.

이영주 외(2013)의 연구는 복지통계, 복지시설 관련 공간자료, 복지 관련 소셜미디어 데이터, 설문조사 결과를 융합하여 공간빅데이터를 구축하고, 이를 활용하여 영유아 복지여건을 수요와 인프라의 두 축에서 진단하는 방안을 제시하였다. 비정형 빅데이터를 장소이름 및 지리좌표 기반으로 공간정보로 변환하고 영유아 복지시설에 대한 주민의견의 내용과 감성을 공간적으로 분석하여, 비정형 빅데이터 공간정보화의 필요성 및 복지수요 분석의 활용 가능성을 제시하였다.

김미정 외(2013)의 연구는 비정형 빅데이터를 포함, 각종 자료를 융합하여 국토정책을 지원할 수 있는 공간빅데이터 체계를 구상하고 그 구축 계획을 수립하였다. 특히, 침수재해 모니터링, 영유아 복지정책 지원, 임대차 수요예측 등의 국토문제 해결에 있어 비정형 빅데이터의 활용방안을 모색하였다.

‘비정형 빅데이터 기반 재난관리정책 고도화 방안’에 관한 연구로는 최선화 외(2013)의 「비정형 빅데이터 수집 및 분석모델 개발」과 최선화·배병걸(2013)의 「소셜 빅데이터 재난관리 운영방안 및 이슈탐지 기법연구」를 검토하였다. 두 연구는

동일 연구과제의 성과로, 국가재난의 전조를 감지하고 신속 대응하기 위해 소셜미디어 빅데이터를 수집, 가공, 분석하는 방법론을 개발하고 이를 재난관리 업무에 적용하는 방안을 제시하였다.

상기 검토된 선행 연구들은 정책수요 발굴, 정책에 대한 국민반응 모니터링, 재난상황 정보의 신속 수집 등 비정형 빅데이터의 다양한 정책활용 방안을 제시하였다. 하지만, 이영주 외(2013)의 연구를 제외한 다른 선행 연구들은 국민이나 주민 의견의 시공간적 분포 패턴 및 그 변화상을 분석할 수 있는 방법론을 제공하지 않아 지역 및 수요 맞춤형 정책 지원에는 제한적이다.

이영주 외(2013)의 연구의 경우, 비정형 빅데이터의 공간정보화와 위치 기반 감성분석 등을 시도하였으나, 여론의 트렌드 파악이나 시공간적 변화상의 지도 표출에 있어서는 기존의 공간분석 틀을 벗어나지 못했다. 본 연구는 국민 여론의 공간적, 시간적 변화를 탐색적으로 분석할 수 있는 지도 기반의 비정형 빅데이터 분석방법론을 체계화하고, 이의 정책활용 가능성 및 방향을 제시한다는 점에서 선행 연구들과 차별된다.

<표 1-1> 선행연구와의 차별성

| 구 분                | 선행연구와의 차별성                                                                                                                                      |                                                                                                                                            |                                                                                                                                                                                                     |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                    | 연구목적                                                                                                                                            | 연구방법                                                                                                                                       | 주요 연구내용                                                                                                                                                                                             |
| 주<br>요<br>선행<br>연구 | 1                                                                                                                                               | <ul style="list-style-type: none"> <li>• 문헌조사</li> <li>• 전문가 자문</li> <li>• 컨퍼런스 참가</li> <li>• 자료 구축 및 분석</li> <li>• 전문기업과 공동 연구</li> </ul> | <ul style="list-style-type: none"> <li>• 빅데이터 개념 정립</li> <li>• 빅데이터 활용사례 및 정책/기술 동향</li> <li>• 국토정책을 위한 빅데이터 활용모델 개발</li> <li>• 빅데이터 관련 기술 조사 및 정리</li> <li>• 빅데이터 기반 부동산정책반응 모니터링 사례 연구</li> </ul> |
|                    | 2                                                                                                                                               | <ul style="list-style-type: none"> <li>• 문헌조사</li> <li>• 전문가 자문</li> <li>• 공간 및 비정형 빅데이터 분석</li> <li>• 설문 조사</li> </ul>                    | <ul style="list-style-type: none"> <li>• 공간 빅데이터의 개념 정립 및 복지정책 활용 필요성</li> <li>• 복지-공간 융합 모델 정립</li> <li>• 영유아 복지여건 진단을 위한 공간 빅데이터 분석(비정형 빅데이터를 활용한 복지수요 분석, 공간 데이터를 활용한 복지시설 공급 분석 등)</li> </ul>   |
|                    | 3                                                                                                                                               | <ul style="list-style-type: none"> <li>• 선행연구 조사</li> <li>• 전문가들과 협동 연구</li> <li>• 빅데이터 분석 관련 전문기업과 협력 연구</li> </ul>                       | <ul style="list-style-type: none"> <li>• 빅데이터 개념과 국토정책과의 관계 정립</li> <li>• 국내외 공간빅데이터 추진 현황 조사</li> <li>• 공간빅데이터 체계 기본구상</li> <li>• 공간빅데이터 체계 구축계획 및 활용 모델 수립</li> </ul>                             |
|                    | 4                                                                                                                                               | <ul style="list-style-type: none"> <li>• 문헌조사</li> <li>• 분류체계, 용어모델 등 지식모델 개발</li> <li>• 분석모델 개발</li> <li>• 프로토타입 개발</li> </ul>            | <ul style="list-style-type: none"> <li>• 빅데이터 개념 정의 및 관련기술 조사</li> <li>• 재난관리 개념정의, 유형분류, 빅데이터 활용 해외사례 조사</li> <li>• 재난관리 분류체계, 용어모델 개발</li> <li>• 재난관리를 위한 비정형 빅데이터 모니터링 모델 개발</li> </ul>          |
|                    | 5                                                                                                                                               | <ul style="list-style-type: none"> <li>• 국내외 문헌 검토</li> <li>• 현황 조사 및 분석</li> <li>• 분석기법 개발 및 검증</li> </ul>                                | <ul style="list-style-type: none"> <li>• 빅데이터 정의 및 동향 분석</li> <li>• 소셜 빅데이터 재난관리 활용현황 조사 및 개선방안 마련</li> <li>• 소셜 빅데이터 기반 재난 전조발생 이슈 탐지기법 개발 및 실험 분석</li> </ul>                                      |
| 본 연구               | <ul style="list-style-type: none"> <li>• 기존 여론 모니터링 도구를 보완하기 위해, 국민 여론의 사공간적 변화 트렌드를 모니터링할 수 있는 비정형 빅데이터 공간 분석방법을 개발하고 이의 정책활용 방향 제시</li> </ul> | <ul style="list-style-type: none"> <li>• 문헌조사 및 이론 연구</li> <li>• 현황 및 사례 조사</li> <li>• 이론 및 방법개발</li> <li>• 자문회의</li> </ul>                | <ul style="list-style-type: none"> <li>• 여론 모니터링 관련 이론 고찰</li> <li>• 비정형 빅데이터 기반 여론 모니터링 방법 및 활용 사례 조사</li> <li>• 여론 모니터링을 위한 비정형 빅데이터 사공간분석 모델의 정립 및 정책 활용 방향에 대한 제언</li> </ul>                    |



chapter II

여론 모니터링과  
비정형 빅데이터 분석



## 여론 모니터링과 비정형 빅데이터 분석

이 장에서는 여론 모니터링의 개념과 필요성에 대한 이론 및 정책적 배경을 검토하고, 비정형 빅데이터를 활용한 여론 모니터링의 사례를 살펴본다. 먼저, 여론 모니터링의 의미, 정책과의 연관성, 여론 분석의 기본 틀을 고찰한다. 다음으로, 인터넷 소셜미디어의 등장으로 인한 여론형성 공간의 변화와 비정형 빅데이터의 대중화에 대해 논의한다. 마지막으로, 여론 모니터링을 위한 비정형 빅데이터 활용의 의의와 사례들을 검토한 후, 방법론적 시사점을 도출한다.

### 1. 여론 모니터링의 개념

#### 1) 여론 모니터링의 개념

일반적으로 여론(Public Opinion)이란 대중(the mass) 혹은 공공 사안에 관여하는 개인들의 집단적 의견을 말한다(김남두, 2010). 따라서, 단순히 개인 의견을 표출하거나 하나의 정보를 전달하는 것은 여론에 해당하지 못하며, 여론형성에는 여러 시민들이 특정 “메시지(이슈)를 공유하고 서로 관심사를 교환하는”(박상호, 2012) 상호 대화과정이 수반된다. 즉, 공통 관심사와 이를 쟁점화할 수 있는 다수의 시민이 여론형성의 필수 조건인 것이다(박상호, 2012).

공통 관심사에 대한 시민들의 집단적 견해라고 할 수 있는 여론의 모니터링이란 비교적 새로운 개념으로, 여론이 어떠한지, 어떻게 변화하고 있는지 그 동향을

지속적으로 조사·분석하여 정보와 지식으로 변환하는 일련의 활동들을 뜻한다. 민간 부문의 경우 대부분 여론 창출과 증대를 위해 여론을 모니터링하고자 한다. 공공 부문의 경우, 여론 모니터링을 통해 국민과의 소통을 활성화하고, 정책 발굴·수립·평가 등 정부활동과 정책을 지원하고자 한다.

정부활동과 정책 지원에 있어 여론 모니터링이 필요한 이유는 두 가지이다. 첫째, 여론과 시민참여를 토대로 한 정책은 민주주의의 기본원칙일 뿐만 아니라 정책과정을 여러 가지 방식으로 개선할 수 있기 때문이다. OECD(2001)에 따르면, 여론과 시민참여는 정책 입법안의 잠정적 효과에 대한 시민 평가 등 새로운 정보를 제공하여 정책의 질을 향상시킨다. 또한, 정책집행이 성공적으로 이루어질 가능성을 높이며, 정책 관련 이해당사자들이 정책과정에 기여할 기회를 제공하여 정책결정과정의 정통성을 강화할 수 있다. 뿐만 아니라, 정책과정의 개방으로 시민-정부 간 파트너십을 구축할 수 있으며, 이는 공공 문제의 해결을 보다 수월하게 만들 수 있다.

정부활동과 정책 지원을 위해 여론 모니터링이 필요한 두 번째 이유는 정책 거버넌스와 정보통신기술이 발전하면서 여론의 조사·수렴과 같은 수동적 방법이 아닌 적극적 여론 이해의 수단이 필요해졌기 때문이다. 본 장의 아래에서 보다 상세히 설명하겠지만, 인터넷 등 전자통신 수단을 통해 국민이 정책과정에 직접 참여할 수 있는 기회가 증가하고 있으며, 이러한 변화를 정책 관계자와 국민 모두 긍정적으로 받아들이고 있다. 한편, 인터넷과 소셜미디어 등 온라인 매체들이 보편화 되면서 여론의 형성, 소통, 정책과정 반영을 위한 정책 도구와 방법에 대한 변화도 촉구되고 있다. 이의 대표적인 예가, 2012년 미국 대통령 선거로 빅데이터 기반 상시적 여론분석을 토대로 선거전략을 구사하여 여론 모니터링에 대한 사람들의 관심을 불러일으켰다.<sup>1)</sup> 2010년대 이후 우리나라에서도 여론에 관한 자료 수집과 분석을 자동화하여 이를 선거, 정책 등 다양한 분야에 활용하고자 하는 시도가 증가하고 있다.<sup>2)</sup>

---

1) <http://www.ciokorea.com/news/22802> (2015년 1월 6일 오후 3시 45분 최종 접속)

2) <http://www.asiatoday.co.kr/view.php?key=349917> (2015년 1월 6일 오후 3시 45분 최종 접속)

## 2) 여론 모니터링의 등장배경

여론 모니터링이 등장하게 된 배경 요인 중 핵심은 정책 거버넌스의 변화이다. 일반적으로 거버넌스란 “공공 및 사적 개인들과 제도들이 공공 목적을 달성하기 위해 자신들의 공통적인 업무를 관리하고, 자원을 통제하고 권력을 행사하는 무수한 방법의 집합이며, 갈등적인 이해나 다양한 이해관계들이 수용되면서 상호협력적인 행동이 취해”(Commission on Global Governance, 1995)지는 과정이다. 민주화, 세계화, 정보통신기술의 발전 등 각종 여건의 변화로 정책 분야의 거버넌스는 전통적 통치 방식에서 e-Governance 패러다임으로 빠르게 진화하고 있다(〈표 2-1〉).

〈표 2-1〉 정책 거버넌스의 발전

| 구분   | 전통적 통치                               | 거버넌스                                                 | e-Governancne                            |
|------|--------------------------------------|------------------------------------------------------|------------------------------------------|
| 의사결정 | 소수의 중앙집권적 의사결정                       | 확산된 분권적 의사결정                                         | 다수의 통합적 의사결정                             |
| 방향성  | 대의적 간접민주주의                           | 다원적 협의 민주주의                                          | 다수적 직접 민주주의                              |
| 가치   | 효율성(efficiency), 대응성(responsiveness) | 효율성(efficiency), 민주성(democracy), 대응성(responsiveness) | 민주성, 효율성, 민주적 관리 (democratic management) |
| 주체   | 정부                                   | 네트워크                                                 | 시민, 시장                                   |

\* 출처: 한국전산원, 2004, “e-Governance 구현을 위한 과제와 전략”, p.21 〈표 5〉 재인용

전통적으로 정책관련 의사결정은 정부와 국민을 대표하는 소수의 정책 입안자들이 전담하는 관료적, 가부장적 구조였다. 정책과정에 국민이 직접 참여할 수 있는 기회는 투표 등으로 제한되었으며, 정책 관련 정보의 국민공유조차 드물었다. 이러한 수직적, 폐쇄적 거버넌스 구조는 정책과정의 효율성과 대응성을 제고하여 빠른 경제성장을 이룩하는데 기여하기도 했지만, 민주주의의 실현과는 거리가 멀다는 단점이 있다.

1990년대 중반 수평적 네트워크를 이루는 다원화된 주체들의 의사결정이 공공 활동과 정책의 성패를 좌우하는 핵심 이슈로 부각되고, 이는 ‘거버넌스’라는 용어로 지칭되었다. 거버넌스 기반의 정책 환경에서는 정부뿐만 아니라, 지방정부, 국민(시민; 주민), 기업, 국제기구, 초국적 기업, 풀뿌리 조직, 비영리단체 등에도 정책 과정에 참여할 기회가 주어지며, 협의와 합의를 통해 의사결정이 내려진다.

정보통신기술이 발전하면서 거버넌스에 참여하는 다양한 주체들이 온라인 네트워크와 사이버 공간을 통해 정책관련 각종 정보를 공유하고, 정책이슈를 논의하며, 관련 의사결정 과정에도 직접 참여할 수 있게 되었다. 이와 같이 “정보통신기술을 이용하여 정책결정과정에 국민의 참여를 확대하며, 행정업무의 신뢰성, 투명성을 보장하는 정부형태”를 e-Governance라 한다(문신용, 2006). 정책 및 관련 행정차원에서 e-Governance는 대의제보다는 개인과 집단이 정책과정에 직접적, 적극적으로 참여한다는 특징이 있다. 이때, 개인과 집단은 선거, 투표, 공청회 등 제도적으로 허용된 장치 이외 다양한 방식으로 정책과정에 참여하고, 일반적으로 정보를 제공받기보다는 쌍방향적 소통과 합의를 원한다(한국대통령학연구소, 2008).

〈표 2-2〉와 같이 e-Governance는 정보 제공(e-Information), 상호작용(e-Interaction), 통합(e-Integration)의 단계로 발전하는데, 정보기술 기반의 정보제공에서 정부-시민간 의사소통 촉진 및 공동 의사결정으로 시민의 정책참여 범위를 확대하는 방식으로 진보한다. e-Government 또는 전자정부는 개념적으로 e-Information과 e-Interaction에 가까우며, 시민이 정보기술 기반으로 의사결정에 직접 참여하는 e-Integration 단계는 여전히 달성해야 할 목표로 남아있다.

〈표 2-2〉 e-Governance의 발전단계와 특성

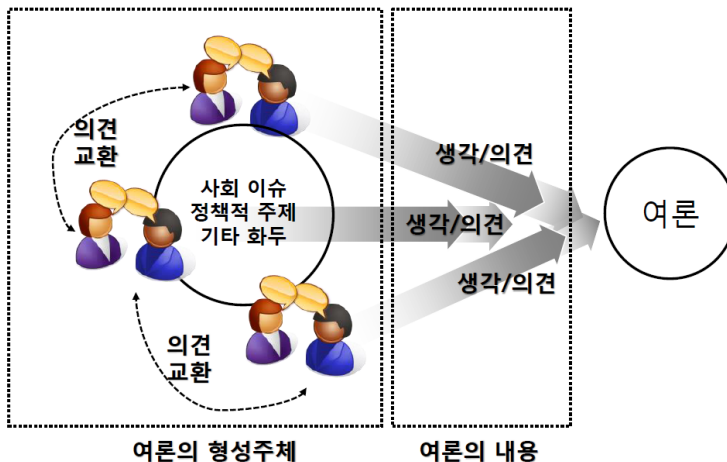
| 발전단계                 | 특성                                                                                                                                     |
|----------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| 1단계<br>e-Information | 시민들에게 참여에 필요한 각종 정보를 온라인을 통하여 제공하는 단계(정부의 각종 정책·시책·법안·사업 및 선거와 관련한 후보자, 공약 등에 관한 자료들)                                                  |
| 2단계<br>e-Interaction | 네트워크 상에 만들어진 게시판과 e-community 등을 통해 의견을 수렴하고, e-polling을 통하여 여론을 수렴하는 단계                                                               |
| 3단계<br>e-Integration | 정부의 각종 정책 및 의사결정에 시민들이 직접 참여하여 의견을 표출하고 시민과 정부가 공동으로 의사결정을 이루어 나가는 단계로서, 정책에 대한 결정은 물론, 선출직을 뽑는 선거 등 모든 결정이 온라인을 통해 이루어지는 단계(e-voting) |

\* 출처: 한국전산원, 2004, “e-Governance 구현을 위한 과제와 전략”, p.25 〈표 7〉 재인용

### 3) 여론 모니터링의 방법

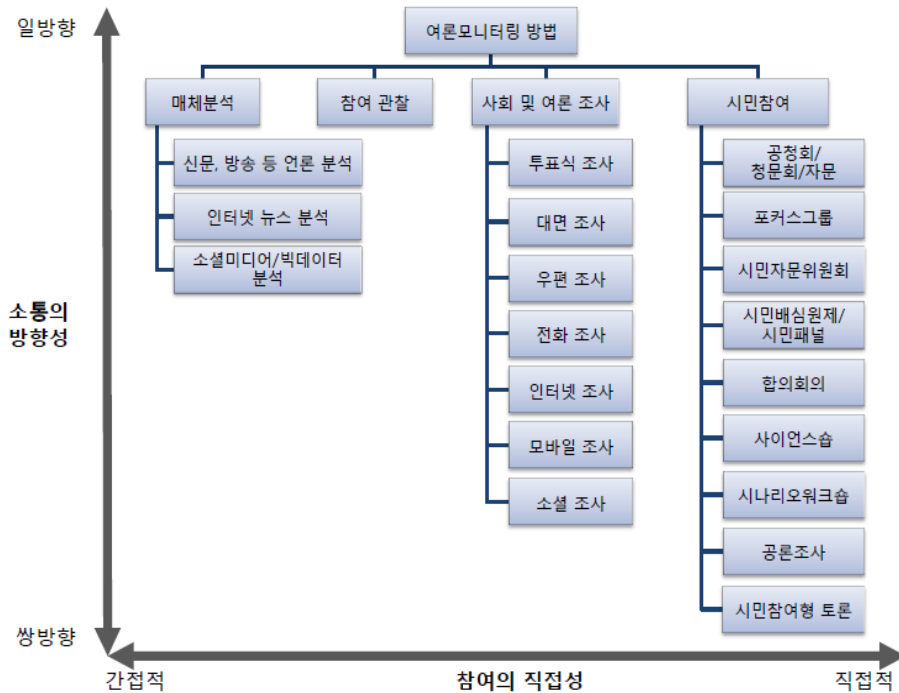
공통 관심사에 대한 시민들의 집단적 견해를 여론이라고 할 때, 여론 모니터링의 대상이 되는 여론의 속성은 크게 내용과 형성주체 두 가지이다(〈그림 2-1〉). 여론의 내용은 집단적 논의의 대상이 되는 사안, 즉 사회의 주요 이슈, 화두, 혹은 정책적 주제에 대한 사람들의 공통된 생각이나 견해를 말한다. 한편, 여론의 형성주체는 이런 생각이나 의견을 개진한 사람들에 해당한다. 여론 모니터링에 있어, 내용 분석은 다양한 언어로 표현된 말과 글 속에서 일관되게 드러나는 공통된 아이디어와 주제, 사람들의 논의 태도와 관련된 감정을 간파하는데 초점을 둔다. 여론의 형성 과정에서 대화와 의견교환 등 사람들 간의 상호작용이 반드시 일어나므로, 상호 관계를 맺은 사람들과 그 관계의 특성을 분석하는 것이 여론 형성주체에 대한 분석의 주가 된다.

〈그림 2-1〉 여론의 속성



〈그림 2-2〉는 여론 모니터링을 위해 활용될 수 있는 구체적 방법들을 보여준다. 이들 방법들은 여론 수집과정에서 일어나는 소통의 방향성 관점에서 일방향적인 소통과 쌍방향적 소통을 기반으로 방법들로 분류될 수 있다. 또한, 여론을 형성하는 주체인 시민참여의 직접성을 기준으로 직접적 방식과 간접적 방식으로 나뉜다.

<그림 2-2> 여론 모니터링 방법의 종류



\* Donsbach & Traugott(2008), 모효정·권복규(2009) 등의 논의를 토대로 도식화

<그림 2-2>에서 가장 간접적인 방법은 매체분석과 참여관찰로, 매체분석은 각종 미디어에 반영된 논의내용의 분석을 통해 여론을 이해하고자 하며 참여관찰은 연구자가 시민들의 일상사에 직접 관찰하면서 주요 화두와 그 특성을 파악하는 접근법이다. 매체가 다양화되면서, 매체분석은 신문, 방송, 인터넷 뉴스 등 언론분석에서, 최근에는 소셜미디어/빅데이터 분석으로 그 영역을 확장하고 있다.

매체나 관찰의 기제에 아니라, 시민의 직접적 참여가 시작되는 여론 모니터링 유형은 사회 및 여론조사이다. 사회조사는 사회적 현상에 관한 정성적, 정량적 질문들을 시민들에게 던져 사회인식과 여론의 동태를 분석하는 접근법인 반면, 여론조사는 시민들의 의견을 묻는데 주력한다. 사회조사는 그 구성에 따라 응답자의 생각과 의견을 파악하도록 설계되기도 하며, 이 경우 사회조사와 여론조사는 그 경계가 흐려진다. 사회/여론 조사는 조사 도구에 따라 그 종류가 다양한데, 한두 가지의

간단한 질문에 대해 다중선택 답안에서 일부를 고르는 투표형 조사(Public Opinion Polling), 직접 시민을 만나서 조사하는 대면조사, 전화, 인터넷, 핸드폰, 소셜미디어 등의 소통도구를 기반으로 하는 조사들이 있다.

시민참여 방법들은 시민 중 일부 대표자들이 특정 안전에 대한 논의에 직접 참여할 수 있는 기제를 제공한 후 그 속에서 표출되는 사람들의 생각과 의견들을 수렴하고, 검토·분석하는 방식으로 여론을 파악한다. 공청회/청문회/자문은 참석자들 간 의사소통이 다소 일방향이적이고 시민이 비교적 간접적으로 의사를 개진하는 방식으로, 시민들은 주로 추척기관의 발표나 사전에 선정된 의제에 대해 듣고 질문을 하거나 추가 의견을 제공한다. 보다 시민참여가 활발해지는 방법들이 포커스그룹, 시민자문위원회, 시민배심원/시민패널, 합의회의이다. 소규모 그룹의 시민들이 자유롭게 토론하고 질의응답하며, 필요시 자문을 요청한다. 사이언스숍과 시나리오워크숍 등은 연구개발 활동이나 지역개발 정책 수립시 시민들의 참여를 유도하고 그 과정에서 특정 이슈에 대한 시민들의 생각과 의견을 읽을 수 있는 여론 모니터링 방법이다. 마지막으로 공론조사나 시민참여형 토론회는 특정 사회적 이슈나 정책현안에 대해 시민들이 숙고하고 토론할 기회를 제공하여, 언론보도나 매체에 의해 편향된 생각과 의견이 아닌 양질의 여론을 형성될 수 있도록 한다.

본 연구에서는 최근 두각을 나타내고 있는 비정형 빅데이터 기반의 여론 모니터링 방법에 초점을 둔다. 정책 거버넌스의 발전으로 사회/여론 조사, 시민참여 등 직접적 시민참여 기반의 방법들도 새로이 관심을 끌고 있으나, 이들 방법에 대한 논의는 본 연구의 범위를 벗어나므로 본문에서는 다루지 않는다. 다만, <부록 1>에서 <그림 3-1>에서 소개된 방법들 중 일부 대표적인 방법들에 대해서만 간략히 설명한다. <그림 3-1>에 표시된 방법들 중 본 연구에서 논의하지 않은 방법들에 대한 상세 정보는 「외교통상 정책결정에서의 의견수렴과정에 관한 연구」(한국대통령학연구소, 2008), 「울산 북구 주민참여정책 추진성과 분석 및 제도화」(울산광역시 북구, 2013) 등의 문헌을 참조할 수 있다.

## 2. 여론 모니터링과 비정형 빅데이터

### 1) 인터넷 소셜미디어 기반의 여론 부상

1980~90년대의 PC시대에서 인터넷 시대와 모바일 시대를 거쳐 지금은 스마트 시대로 나아가고 있다(최진명, 2012). 이러한 정보통신기술의 진화로 인터넷 패러다임 또한 정보제공 중심의 웹1.0에서 ‘참여·공유·개방’의 웹2.0, 그리고 개별 사용자의 특성과 요구에 맞게 개인화된 웹의 형태인 웹3.0으로 발전하고 있다(<표 2-3>). 인터넷은 그 출현 직후부터 사람의 의사소통을 위한 미디어로 활용되어 왔고, 따라서 인터넷 패러다임의 진화는 사람들의 의사소통 양식을 변화시키고 있으며 이는 다시 e-Governance의 발달과 맞물려 정책과정에도 영향을 미치고 있다.

<표 2-3> 인터넷 패러다임의 진화

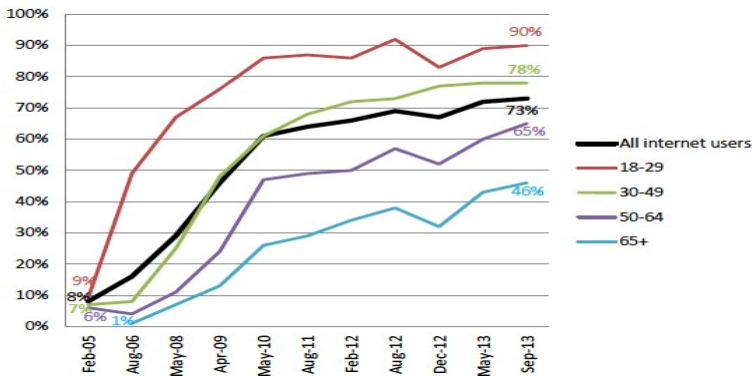
| 구분   | 웹 1.0                                                                     | 웹 2.0                                                                                     | 웹 3.0                                                                           |
|------|---------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| 특징   | <ul style="list-style-type: none"> <li>정부 중심</li> <li>일방향 정보제공</li> </ul> | <ul style="list-style-type: none"> <li>시민 중심</li> <li>양방향 정보제공</li> </ul>                 | <ul style="list-style-type: none"> <li>개인 중심</li> <li>개인별 정보제공</li> </ul>       |
| 기반기술 | <ul style="list-style-type: none"> <li>브라우저/웹 저장</li> </ul>               | <ul style="list-style-type: none"> <li>브로드밴드</li> <li>Rich Link/Content Models</li> </ul> | <ul style="list-style-type: none"> <li>시맨틱 기술</li> <li>센서 네트워크</li> </ul>       |
| 주류형태 | <ul style="list-style-type: none"> <li>전문가가 제작한 특제 주제 중심의 웹사이트</li> </ul> | <ul style="list-style-type: none"> <li>개인이 템플릿을 이용해 만든 블로그/홈페이지, 소셜미디어</li> </ul>         | <ul style="list-style-type: none"> <li>데이터를 자동으로 처리해 제공하는 맞춤형 정보 서비스</li> </ul> |
| 중심관계 | <ul style="list-style-type: none"> <li>전문가-방문자</li> </ul>                 | <ul style="list-style-type: none"> <li>이용자-이용자</li> </ul>                                 | <ul style="list-style-type: none"> <li>데이터-이용자</li> </ul>                       |

\* 출처: 이해정(2007), 변미리(2012)의 논의 내용 정리

인터넷 패러다임의 변화가 스마트 폰 환경으로 전이되면서, 최근 많은 주목을 끌고 있는 이슈는 소셜미디어의 활용 증대이다. 소셜미디어(Social Media)는 “웹 2.0을 바탕으로 하는 인터넷 기반 애플리케이션의 총체이며 사용자가 콘텐츠를 만들고 교환할 수 있도록 하거나 혹은 소비자가 스스로 만든 미디어”(Kaplan & Haenlein, 2010)이다. 미디어 형태 측면에서 소셜미디어는 블로그, SNS, 위키, 마이크로블로깅, 팟캐스트 등 다양한 미디어를 포괄하며, 쌍방향의 즉각적 의사소통 지원, 개방형 서비스, 매쉬업 및 정보 융합의 용이성 등의 특징을 지닌다(고영삼, 2011).

Pew Research Center에 따르면, 2014년 현재 약 18억의 인터넷 사용자가 소셜미디어를 이용하고 있으며<sup>3)</sup>, 2013년에는 성인 인터넷 사용자 중 약 73%가, 18-29세 인터넷 사용자 중 약 90%가, 30-49세 인터넷 사용자 중 약 78%가 소셜미디어를 사용했다고 한다(<그림 2-3>)<sup>4)</sup>. 국내의 경우, 한국인터넷진흥원의 인터넷이용실태 조사에 따르면, 20대의 소셜미디어 사용률이 87.5%, 30대가 74%, 만 6-19세가 52%로, 국내에서도 젊은 층의 이용률이 높게 나타났다(이윤희, 2014).

<그림 2-3> 인터넷 사용자 중 연령대별 소셜미디어 이용률의 변화



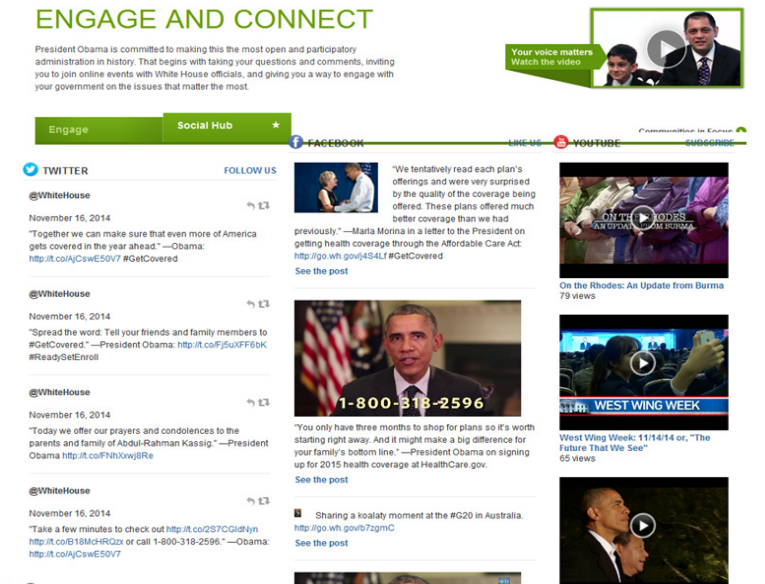
출처: <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet> (2015년 1월 6일 오후 2시 10분 최종 접속)

정보통신기술 기반의 정보공유와 온라인 소통 활성화에 더하여, 최근 소셜미디어의 확산은 참여행동과 직접소통을 증진하고 정책 공론장을 활성화하여 정책과정 전반에 다양한 변화를 야기하고 있다. 많은 사람들이 SNS를 통해 실시간으로 정보를 주고받으며 자신들의 의사를 국가 및 지역의 정책에 반영하고 싶어 한다. 또한, 블로그, 트위터와 같은 소셜미디어로 표출된 개개인의 의견이 집합되어 여론을 형성하면서 정책과정에 보다 강한 영향력을 발휘하게 되었고, 심지어 풀뿌리 여론의 형성 공간을 소셜미디어가 대체할 수 있다(박상호, 2012)는 평가까지 받고 있다.

3) <http://www.statista.com/topics/1164/social-networks> (2005년 1월 6일 오후 2시 10분 최종 접속)

4) 2013년 7월 18일부터 9월 30일까지, 18세 이상의 인터넷 사용자 5,112명을 대상으로 조사되었다(<http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet>; 2005년 1월 6일 오후 2시 10분 최종 접속).

<그림 2-4> 미국 시민참여국 홈페이지



출처: <http://www.whitehouse.gov/engage/social-hub> (2015년 1월 6일 오후 2시 10분 최종 접속)

상기한 바와 같이 소셜미디어를 통한 시민의 정책참여 수요가 증대하면서, 국내외 정부기관들은 정책수단으로서 소셜미디어의 활용 방안을 여러모로 모색 중에 있다(서원석·조세현·송희준, 2012; Kavanaugh 외, 2012). 미국의 오바마 정부는 2009년 시민참여국을 창설하고 정부-시민 간 의사소통을 위해 정보기술을 적극 활용 중인데, 최근에는 ‘We the People’, ‘e-타운홀 미팅’ 등의 프로그램을 통해 다양한 SNS를 통해 각종 정책관련 행사에 시민의 자발적 참여를 유도하고 중요 사안에 대해서는 시민들이 직접 참여할 수 있는 기회를 제공하고 있다(<그림 2-4>).

국내에서도, 대통령실, 안전행정부, 문화체육관광부, 국토교통부 등 대다수의 중앙 정부 기관이 SNS 계정을 도입 운영하여 양방향 소통을 시도하고 있으며, 국민신문고의 국민제안 코너나 서울시의 천만상상 오아시스 등의 정책제안 서비스가 이루어지고 있다(이원태·차재관·신호철, 2012). 국가 및 지역 정책 이해당사자들도 이러한 변화 흐름에 맞춰 소셜미디어를 이용하여 사회 문제점이나 정책수요에 빠르게 대응할 필요가 있다.

## 2) 여론 모니터링을 위한 비정형 빅데이터의 활용

인터넷 소셜미디어 기반 여론의 부상은 여론 모니터링의 방법론적 변화를 촉구하고 있다. 과거에는 신문, 방송, 토론회 등 전통적 매체들을 활용하여 여론의 동향과 변화를 살펴보았다. 여론이 형성·표출되는 공간이 오프라인, 지면매체, 방송 등에서 온라인의 인터넷 소셜미디어로 확대되면서, 여론 모니터링의 대상이 되는 자료의 공급원 역시 인터넷 뉴스, 카페, 블로그, 트위터 등의 신규매체로 포함하게 됐다.

인터넷 소셜미디어 등 신규매체들은 시민들의 자발적 의견공유와 쌍방향 소통을 지원하는 플랫폼 역할을 하는데, 누구나 차별 없이 언제 어디서든 생각, 의견, 감정을 표현할 수 있도록 하고 그 결과물을 디지털 데이터로 수집·관리한다. 이러한 플랫폼 환경에서 수집되는 여론 데이터는 그 형태와 종류가 다양하며, 빠른 속도로 끊임없이 생성된다. 결과적으로, 신규매체에서 생산하는 여론 데이터는 그 크기가 매우 방대한 빅데이터의 특성을 띤다. 또한, 텍스트, 사진 등의 형태로 표현되어 데이터 내의 구조가 정해져 있지 않는 비정형 데이터의 특성을 보인다. 따라서, 인터넷 소셜미디어 등 신규매체를 통해 수집되는 여론 데이터를 분석·모니터링하기 위해서는, 비정형 빅데이터를 다루어야 하며 이들 데이터에서 가치있는 정보를 추출할 수 있는 정교한 분석기법들이 필요하게 되었다.

신규매체 중심의 비정형 빅데이터 기반 여론 모니터링은 데이터 수집, 처리, 분석을 위한 새로운 방법론을 요구하지만, 전통적 매체 기반의 접근법의 한계를 보완할 수 있다. 먼저, 신문, 방송, 토론회 등 전통적 매체에서는 포착하기 어려운 시민들의 일상적 여론 활동과 정책 수요에 대한 이해를 높일 수 있다. 비록 신규매체의 이용자들이 20~40대의 젊은 계층으로 한정되긴 하지만, 이들 젊은이들은 출근, 식사, 업무, 친목 등의 일상 활동을 수행하면서도 자유롭게 의견을 개진하고 교환할 수 있다. 지면, 스튜디오, 회의장 등 공식적 공간에서 논의하기 어려운 주제들도 타인을 의식할 필요가 없는 신규매체에서는 진솔하게 논의될 수 있다. 그러므로 신규매체에서 수집된 비정형 빅데이터를 여론 모니터링에 활용할 경우, 보다 신선하고 다양한 여론을 포착하고 국민들의 실질적인 정책수요를 발굴할 가능성이 높아진다.

다음으로, 비정형 빅데이터 기반 여론 모니터링은 여론 트렌드 변화를 보다 신속히 포착할 수 있게 한다. 기존의 여론 모니터링 방법에서는 소수의 전문가들이 발품을 팔아 자료를 수집하고 매체의 내용을 일일이 읽고 분석하는 과정을 거쳐야 하므로, 여론 트렌드의 파악 속도가 느리다. 반면, 트위터, 페이스북 등 인터넷 소셜미디어는 실시간으로 여론 관련 메시지를 생산하며 이들 데이터는 크롤링, API 등의 방법으로 자동으로 수집할 수 있다. 따라서, 크롤링, 텍스트마이닝, 사회관계망 분석 등 다양한 정보기술과 분석기법을 융합 활용하여, 자동/반자동화된 여론 트렌드 추출이 가능해진다. 결과적으로, 여론의 경향성과 변화 패턴을 파악하는 속도가 빨라지게 된다.

### 3) 비정형 빅데이터 기반 여론 모니터링 방법과 활용 사례

컴퓨터의 대중화, 인터넷 사용량의 폭발적 증가, 모바일 애플리케이션과 소셜미디어의 확산 등으로 오늘날의 세계는 바야흐로 빅데이터 시대에 진입하였다. 21세기의 원유라고도 불리는 빅데이터의 활용이 기업 이윤, 정부 정책, 행정의 효율성을 좌우할 것으로 예측된다(Goes, 2014). 우리나라의 경우 특히 빅데이터에 대한 관심이 높는데, 국가 및 지역 정책을 위한 빅데이터 활용방안을 다양한 각도에서 모색 중이다(김대중·윤서연, 2013; 김미정 외, 2013; 이석주 외, 2013; 최선화·배병걸, 2013). 본 소절에서는 인터넷 뉴스, 소셜미디어 등 신규 매체에서 생산하는 비정형 빅데이터를 이용하여 국민 여론의 현황과 동태를 분석하는 방법론과 그 활용사례들을 살펴본다.<sup>5)</sup>

#### (1) 텍스트마이닝 기반 매체분석

##### ① 개요

인터넷 뉴스, 소셜미디어 등 신규매체는 다양한 규모의 언론집단과 시민 개개인이 국가와 지역의 정책에 관한 정보, 생각, 의견, 감정을 서로 교환하고 소통할 수 있는

---

5) 소셜미디어의 국가/지역 정책 활용방안에 대해서도 많은 논의가 있으며, 정책홍보 및 국민소통을 위한 소셜미디어의 활용전략을 관해서는 서원석·조세현·송희준(2012)과 광운대학교 산학협력단(2014)의 연구를 참조하길 바란다.

여론형성의 새로운 창구가 되고 있다. 이들 매체는 접근이 쉽고, 사용자가 원할 때 자유로이, 대부분 자발적으로 자료를 제공할 수 있으며, 공유된 자료가 인적, 기술적 네트워크를 타고 빠르게 전파될 수 있을 뿐만 아니라, 한 번 수집된 자료는 상당 기간 지속적으로 축적한다는 장점이 있다. 신규매체를 통해 수집되는 자료는 텍스트, 이미지, 동영상 등 그 형태가 다양한데, 그 중 텍스트 자료가 주를 이루고 그 양이 방대하다. 따라서, 신규매체에 반영된 여론과 정서적 반응이 어떠한지, 이들 패턴은 어떻게 변하는지를 보기 위해서는 방대한 양의 텍스트 자료를 자연어처리하고 계량분석하는 텍스트마이닝 방법들이 활용된다. 일반적으로 텍스트마이닝은 그 의미를 찾기 힘든 텍스트 기반의 자료에서 새로운 정보를 발견해내는 일련의 방법과 기술들을 총칭하며(솔트룩스, 2008), 여론 모니터링의 맥락에서 자주 이용되는 텍스트마이닝 기법으로는 동시출현단어분석, 키워드빈도 분석, 감성분석, 이슈/토픽 추출 등이 있다.

동시출현단어 분석이란 수집된 텍스트 문서에서 키워드들이 동시 출현하는 패턴을 파악해내는 분석기법이다. 예를 들어, 하나의 트윗에서 동시 출현한 키워드의 쌍을 추출하고 전체 문서집합에서 키워드 쌍별 발생 빈도와 연결 관계를 분석하면, 트위터 상의 주요 관심 이슈와 그 연계성의 변화를 추적할 수 있게 된다(진설아 외, 2013). 키워드 중 특정 주제를 선택하고 이와 동시출현한 다른 키워드 간의 관계를 살펴보는 분석을 특히 연관어 분석이라고 하며, 키워드들을 주제별로 분류한 후 연관어분석을 실시하면 주요 이슈간의 관련성을 보다 구조적으로 파악할 수 있는 장점이 있다. 연관어분석의 사례는 <그림 2-5>와 같다(최선화·배병걸, 2013).

키워드빈도 분석은 특정 문서 집단 내에서 자주 언급되는 키워드들을 추출하고 이들이 언급되는 빈도에 따라 순위를 부여하는 분석방법이다. 이 방법에서 키워드의 순위는 단순 빈도로 결정될 수 있으나, 동시출현단어 분석으로 도출된 키워드 간의 네트워크 그래프를 기반으로 그 중심성에 따라 순위가 계산(페이지랭크 등)되기도 한다(진설아 외, 2013). 키워드빈도 분석을 실시간이나 일정 시간간격으로 수행하여 사람들의 관심사가 어떻게 변하는지 그 추이를 탐색할 수 있다. <그림 2-6>은 키워드 빈도를 시계열 분석한 사례이다.

<그림 2-5> 폭염 연관어 분석 결과



\*출처: 최선화·배병걸(2013) 그림 3.9(p.55) 발췌

<그림 2-6> 키워드빈도 시계열 분석의 사례

|    | 3월      |    | 4월      |        | 5월    |       | 6월      |       | 7월      |     | 8월      |     | 9월      |     |
|----|---------|----|---------|--------|-------|-------|---------|-------|---------|-----|---------|-----|---------|-----|
| 순위 | 연관어     | 건수 | 연관어     | 건수     | 연관어   | 건수    | 연관어     | 건수    | 연관어     | 건수  | 연관어     | 건수  | 연관어     | 건수  |
| 1  | 양도세     | 5  | 양도세     | 10,644 | 양도세   | 4,473 | 양도세     | 1,911 | 양도세     | 968 | 양도세     | 781 | 양도세     | 738 |
| 2  | 양도세면제   | 1  | 양도세면제   | 200    | 양도세면제 | 113   | 양도세면제   | 44    | 양도세면제   | 25  | 양도세면제   | 27  | 양도세면제   | 9   |
| 3  | 양도세종과폐지 | 1  | 양도세면제   | 77     | 양도세면제 | 56    | 양도세면제   | 29    | 다주택자양도세 | 7   | 양도세면제   | 12  | 양도세종과폐지 | 5   |
| 4  | 양도세면제   | 1  | 양도세면제   | 30     | 양도세면제 | 13    | 양도세종과폐지 | 11    | 양도세면제   | 6   | 양도세면제   | 10  | 다주택자양도세 | 4   |
| 5  |         |    | 양도세종과폐지 | 21     | 양도세면제 | 13    | 양도세종과폐지 | 11    | 양도세면제   | 5   | 다주택자양도세 | 5   | 양도세종과폐지 | 4   |
| 6  |         |    | 양도세면제   | 17     | 양도세면제 | 11    | 양도세종과폐지 | 10    | 양도세면제   | 4   | 다주택자양도세 | 3   | 양도세면제   | 4   |
| 7  |         |    | 양도세종과폐지 | 16     | 양도세면제 | 8     | 양도세면제   | 6     | 양도세면제   | 4   | 양도세면제   | 3   | 양도세종과폐지 | 3   |
| 8  |         |    | 양도세면제   | 14     | 양도세면제 | 7     | 양도세면제   | 5     | 양도세면제   | 4   | 양도세면제   | 6   | 양도세면제   | 2   |
| 9  |         |    | 양도세면제   | 14     | 양도세면제 | 5     | 다주택자양도세 | 4     | 양도세면제   | 4   | 양도세면제   | 3   | 양도세면제   | 2   |
| 10 |         |    | 양도세면제   | 12     | 양도세면제 | 4     | 양도세면제   | 3     | 양도세면제   | 4   | 양도세면제   | 2   | 양도세면제   | 2   |
| 11 |         |    | 양도세면제   | 11     | 양도세면제 | 4     | 양도세면제   | 3     | 양도세면제   | 3   | 양도세면제   | 1   | 다주택자양도세 | 1   |
| 12 |         |    | 양도세면제   | 11     | 양도세면제 | 4     | 양도세면제   | 2     | 양도세면제   | 2   | 양도세면제   | 1   | 양도세면제   | 1   |
| 13 |         |    | 양도세면제   | 7      | 양도세면제 | 4     | 양도세면제   | 1     | 양도세면제   | 2   | 양도세면제   | 1   |         |     |
| 14 |         |    | 양도세면제   | 6      | 양도세면제 | 2     | 양도세면제   | 1     | 양도세면제   | 1   |         |     |         |     |
| 15 |         |    | 양도세면제   | 6      | 양도세면제 | 2     | 양도세면제   | 1     | 다주택자양도세 | 1   |         |     |         |     |
| 16 |         |    | 양도세면제   | 5      | 양도세면제 | 2     | 양도세면제   | 1     | 양도세면제   | 1   |         |     |         |     |
| 17 |         |    | 양도세면제   | 4      | 양도세면제 | 2     | 양도세면제   | 1     | 양도세면제   | 1   |         |     |         |     |
| 18 |         |    | 양도세면제   | 3      | 양도세면제 | 2     | 양도세면제   | 1     | 양도세면제   | 1   |         |     |         |     |
| 19 |         |    | 양도세면제   | 3      | 양도세면제 | 1     |         |       | 양도세면제   | 1   |         |     |         |     |
| 20 |         |    | 양도세면제   | 2      | 양도세면제 | 1     |         |       | 양도세면제   | 1   |         |     |         |     |
| 21 |         |    | 다주택자양도세 | 2      | 양도세면제 | 1     |         |       | 양도세면제   | 1   |         |     |         |     |
| 22 |         |    | 양도세면제   | 2      | 양도세면제 | 1     |         |       |         |     |         |     |         |     |

\*출처: 김대중·윤서연(2013) <표 4-13>(p.108) 발췌

감성분석이란 “텍스트에 나타난 사람들의 태도, 의견, 성향과 같은 주관적인 데이터를 분석하는 자연어 처리 기술”(신수정, 2014)이다. 2012년 미국 대선에서 이 기법이 이용되면서 그 인기가 상승하였으며, 최근에는 민간이나 공공에서 마케팅 도구로서 도입하고 있다. 감성분석은 표본으로 선택된 사람들에게 긍정/부정 등을 묻지 않고 대량의 텍스트 자료에서 사람들의 정서 상태를 추정함으로써 대중의 반응을 (반)자동으로 분석한다는 장점이 있다. 자료수집, 감성과 관련된 문장요소만을 추출하는 주관성탐지, 추출된 키워드들이 긍정인지 부정인지를 계량화하는 극성탐지의 세 단계를 거쳐 감성분석이 이루어지며(신수정, 2014), 그 결과는 <그림 2-7>과 같이 키워드나 문서를 몇 가지의 감정 카테고리로 분류하는 형태를 띤다.

<그림 2-7> 전기부족 관련 감성분석 결과



\*출처: 최선화-배병걸(2013) 그림 3.10(p.56) 발췌

마지막으로 이슈/토픽 추출은 방대한 텍스트나 멀티미디어 자료에서 특정 주제나 이슈, 주제그룹들을 자동으로 파악, 추출하는 분석 기법이다(Sobkowicz 외, 2012). 이슈/토픽 추출에 대한 접근법은 다양한데, 일반적으로 감독분류(Supervised Classification)와 무감독분류(Unsupervised Classification)로 구분할 수 있다. 감독분류에서는 분석가가 특정 이슈나 토픽에 관한 것이라 판단한 일련의 훈련(학습)용 문서를 선택하고 이 문서들의 단어구성, 문장구조 등을 통해 이슈/토픽을 찾아내는

분류 모델을 먼저 개발한다. 이후, 그 모델을 신규로 확보된 문서들에 적용하여 미리 선정된 이슈/토픽 중 하나로 할당한다. 반면, 무감독분류에서는 특정 문서군 내에 포함된 단어들을 그 맥락과 의미를 기반으로 유사도에 따라 몇 개의 그룹으로 분류한 후 포괄적인 주제어를 부여하는 방식을 취한다. 두 방법 모두 이슈/토픽 추출시 유용하나, 여론분석에서 있어 주요 관심이슈가 미리 파악된 경우에는 감독분류가 보다 효율적이며, 여론상태에 대한 선입견 없이 인터넷뉴스, 소셜미디어 등에서 회자되는 공통이슈를 파악하고자 한다면 무감독분류가 더 적합하다 하겠다.

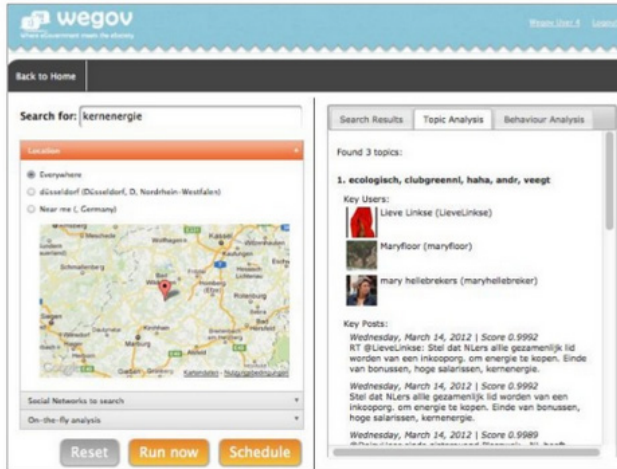
## ② 활용 사례

텍스트마이닝 기반 매체분석의 대표적 사례는 2009년 H1N1 독감 유행시 미국에서 트위터 기반으로 실시간 감기트렌드 분석을 한 것이다(AAPOR, 2014). Chew-Eysenbach(2009)는 “H1N1”, “swine flu”라는 키워드를 포함한 트윗을 수집·분석하여, 이들 메시지가 주로 정보전달을 목적으로 생성되었으나 독감과 같은 질병에 대한 사람들의 태도나 경험을 보여주는 좋은 자료라고 보고한 바 있다. 유사하게, Paul-Drezde(2009)는 감기에 관한 트윗 메시지의 양과 미국 질병통제센터에서 발표한 공식 감기통계가 높은 연관관계에 있음을 보여주고, 소셜미디어 자료가 장시간에 걸친 건강관련 모니터링에 효과적으로 주장하여 미국의 보건정책에 시사점을 제시한 바 있다. 유럽에서는 유럽연합위원회(European Commission) 차원에서 정책 관련 의사결정자들이 다양한 소셜미디어와 이들 매체에서 수집된 자료에서 주요 이슈와 토픽을 추출할 수 있도록 지원하는 소프트웨어 프로그램을 개발하는 WeGov 프로젝트를 진행하였고 그 산출물로 WeGov 툴박스를 제작한 바 있다(Addis 외, 2010). <그림 2-8>은 WeGov 툴박스의 메시지 검색 및 토픽 분석 결과 화면을 보여준다.

국내에서는 2013년 문화체육관광부에서 「빅데이터 분석을 통해 본 2013 국민 인식 변화」 과제를 수행하면서 국민인식 및 여론변화 추이 파악을 위해 특정 주제어의 언급량 분석, 감성 연관어 분석 등을 실시한 바 있다(<그림 2-9>). 또한, 안정행정부 국립재난안전연구원에서 소셜미디어 빅데이터를 활용한 맞춤형 재난관리 운영방안을 연구한 바 있다(최선화·배병걸, 2013). 특히, 이 연구에서는 재난관련 감성분석,

재난이슈 탐지기법 등을 개발하고 이를 활용하여 소셜미디어를 실시간 모니터링함으로써 재난관리 단계별 맞춤형된 정보서비스를 개발할 것을 제안하였다.

<그림 2-8> WeGov 툴박스



\*출처: [http://www.slideshare.net/timo\\_wandhoefer/wegov-user-guide-toolbox-25](http://www.slideshare.net/timo_wandhoefer/wegov-user-guide-toolbox-25) (2015년 1월 6일 오후 2시 10분 최종 접속)

<그림 2-9> 복지정책 분야에 대한 국민 관심도

▪ '복지'의 연관어 변화

| NO | 2012.06~ | 2012.09~ | 2012.12~ | 2013.3~ |
|----|----------|----------|----------|---------|
| 1  | 사회       | 사회       | 사회       | 사회      |
| 2  | 경제       | 경제       | 박근혜      | 사회복지    |
| 3  | 경제       | 경제       | 경제       | 경제      |
| 4  | 국민       | 박근혜      | 경제       | 교육      |
| 5  | 사회복지     | 후보       | 한국       | 경제      |
| 6  | 교육       | 문화인      | 국민       | 한국      |
| 7  | 한국       | 박근혜      | 문화인      | 국민      |
| 8  | 사회       | 교육       | 사회복지     | 지음      |
| 9  | 지음       | 공약       | 공약       | 지역      |
| 10 | 정부       | 안철수      | 후보       | 문화      |
| 11 | 박근혜      | 사회복지     | 노인       | 사람      |
| 12 | 서울       | 경제민주화    | 예산       | 정부      |
| 13 | 생활       | 정부       | 정부       | 대선      |
| 14 | 지역       | 대선       | 교육       | 서비스     |
| 15 | 사람       | 한국       | 재균       | 사회      |
| 16 | 대선       | 예산       | 대통령      | 노인      |
| 17 | 경제민주화    | 문화       | 지음       | 박근혜     |
| 18 | 대선       | 불자리      | 문화       | 생활      |
| 19 | 문화       | 지음       | 존        | 업무      |
| 20 | 재균       | 서비스      | 사회       | 문화      |
| 21 | 예산       | 지역       | 아름       | 서울      |
| 22 | 문화       | 정치       | 서울       | 공무원     |
| 23 | 노인       | 사람       | 생활       | 시장      |
| 24 | 서비스      | 재균       | 재단       | 영남      |
| 25 | 생활       | 노인       | 지역       | 예산      |

따루가 말한다. 핀란드엔 학원이 없다. 공교육에 대한 믿을 때문이다. 2차 대전 이후 핀란드가 빈곤할 때 복지를 시작했다.

한국에만 있는 산후조리원. 출산 후 산후조리원 가는 게 당연한 서비스 순서인 듯 되버렸는데 우리나라 복지정책 아직 멀었다.

서울 사는 65세 이상 노인들이 갈만한 노인정과 여가 보낼 수 있는 장소 만들어주는 복지 대책은 어떻게?

복지도 경제민주화든 예전 참리화, 과학화하듯 하다. 동이 제대로 쓰이도록!

\*출처: 문화체육관광부(2013) 슬라이드 #16 발췌

부동산정책의 효과 및 관련 수요 예측에 있어서도 텍스트마이닝 기반의 소셜미디어 매체분석이 유용함을 보여주는 연구도 있다. 김대중·윤서연(2013)은 부동산 임대차 관련 정책에 대한 국민반응을 모니터링하는데 감성분석, 키워드 빈도분석이 유용하며, 주요 키워드빈도의 추이 분석을 통해 떠오르는 부동산 수요도 파악 가능성을 제시한 바 있다.

## (2) 여론형성에 관한 사회관계망 분석

### ① 개요

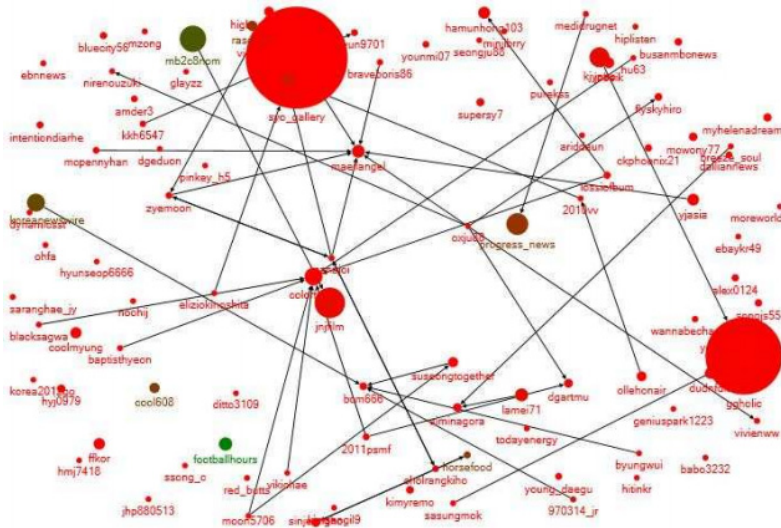
텍스트마이닝 방법의 초점은 여론 내용의 이해에 있다. 여론 모니터링에 있어 여론의 내용파악과 함께 중요한 이슈는 누가 여론형성을 주도하고 여론이 어떻게 파급되는가이다. 소셜미디어의 경우, 메시지의 작성자, 응답자, 청취자, 재전파자 등의 정보가 기록되기 때문에 여론형성과 전파의 경위나 패턴을 분석할 수 있다. 사회관계망 분석이 특히 이러한 패턴분석에 유용하며, 그 결과는 국가와 지역의 정책과정에 필요한 여론을 수집할 때 소셜미디어 상의 여론형성을 어떻게 주도하면 되는지, 또는 여론내용 분석시 어떠한 점에 주의하면 되는지에 대한 정보를 제공한다.

사회관계망 분석은 “특정 사회 시스템에서 어떤 구성요소들이 어떤 구도의 관계망을 형성하고 있는지를 나타내는데 유용한 연구 방법이다”(조성은 외, 2011). 이 정의에 따라, 소셜미디어 이용자들의 세계를 온라인 상의 사회 시스템으로 간주할 수 있다. 이 경우 구성요소는 소셜미디어 이용자 개인, 트위터에서는 트위터 계정을 보유하고 트윗 활동을 하는 이용자가 되며, 누가 메시지를 생성하고 누가 이 메시지를 팔로잉(following)하는지를 네트워크 형태의 자료로 구성하고 이를 시각화하면 트위터 상에서 이슈가 어떻게 형성되고 전파되는지를 보다 쉽게 이해할 수 있다.

서치 네트워크(Search Network)와 에고 네트워크(Ego Network)는 트위터 이용자 간의 관계망 분석시 자주 활용되는 시각화 기법이다(전채남, 2011). 서치 네트워크는 동일 키워드를 포함한 트윗 메시지를 작성한 이용자들 간의 관계를 시각적으로 보여준다. <그림 2-10>이 그 예인데, 네트워크 상의 노드가 녹색일수록 더 많은

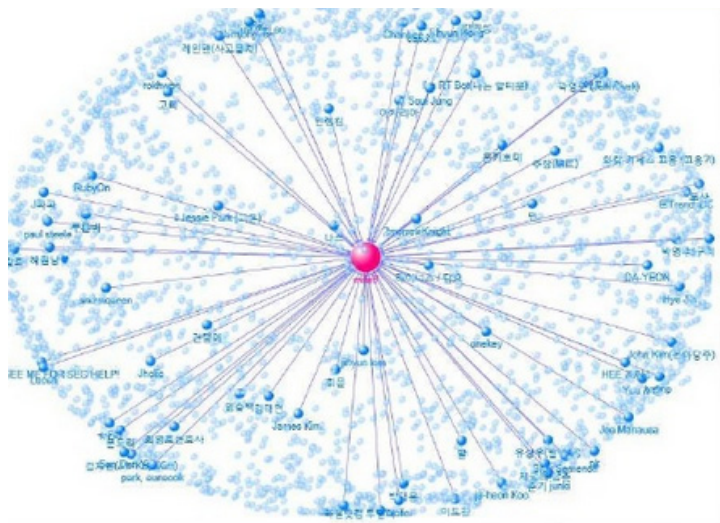
메시지를 작성하고 있고 노드의 크기가 클수록 많은 팔로어(follower)가 있음을 뜻한다.

<그림 2-10> '대구시' 검색어를 이용한 서치 네트워크



\*출처: 전채남(2011) 슬라이드 #21 발췌

<그림 2-11> '농림수산식품부' 트위터계정의 팔로어(follower) 네트워크



\*출처: 조성은 외(2011) 슬라이드 #19 발췌

반면, 에고 네트워크는 특정 트위터 이용자를 대상으로 이 이용자와 다른 이용자들이 어떤 관계망을 형성하고 있는지를 보여준다. 에고 네트워크에서는 중앙노드는 선택된 특정 이용자를, 다른 노드들은 선택된 이용자를 팔로잉하거나 팔로잉된 이용자를, 선의 굵기는 두 이용자 간에 상호 응답, 즉 상호소통이 잦았음을 뜻한다. <그림 2-11>는 에고 네트워크의 예시이다.

## ② 활용 사례

사회관계망 분석의 국내 활용사례로는 농림수산물식품부와 대구시의 트위터 운영실태를 분석한 연구와 경상북도의 도민 맞춤형 계획 수립의 경우가 있다. 조성은 외(2011)는 2011년 농림수산물식품부의 트위터 계정 운영 실태를 사회관계망 기법으로 분석하여, 정부-시민 간 온라인 소통을 개선하고 수집되는 여론자료의 질을 향상시킬 수 있는 방향을 제시하였다. 전채남(2011)은 대구시 트위터와 페이스북의 운용결과 분석에 서치 네트워크, 에고 네트워크 등 시각화 중심의 사회관계망 분석을 적용한 후, 대구시의 시민소통 효과를 제고하고 SNS 상 여론형성 과정을 개선할 수 있는 방안을 제안하였다. 마지막으로, 더아이엠씨(2014)는 경상북도의 요청으로 도청 홈페이지와 네이버, 다음 등 포털사이트에서 수집된 빅데이터를 사회관계망으로 분석하여 문화관광, 창의경제 등의 분야에 대한 정책수요를 제시하였다.

## 3. 시사점

본 장에서 논의한 여론 모니터링의 개념, 방법, 비정형 빅데이터 활용 사례 등을 종합해보면, 민주주의와 함께 시작된 여론 모니터링이 시민참여의 확산, 신규매체(소셜미디어 등)와 방법론(텍스트마이닝 등)의 발달로 그 중요도와 활용도가 함께 높아지고 있음을 알 수 있다. 정책지원의 관점에서 보았을 때, 여론 모니터링 방법론의 최근 변화는 네 가지 시사점을 던져준다.

첫째, 중앙 및 지방 정부는 시민을 정책과정의 주요 주체로 전제하고 이들의 목소리를 직간접적으로 들을 수 있는 다양한 여론창구를 마련하고 지속적으로 운영해야

한다. 민주주의의 개념상 시민이 정책과정의 중심이 되는 것은 당연함에도 불구하고, 그간 시민들의 목소리, 즉 국민여론을 수렴하는 것은 실효성이 낮은 형식적 절차로 여겨지는 경우가 많았다. 사회/여론 조사의 확산, 공론조사와 시민참여형 토론의 이용 증대, 소셜미디어 등 시민참여의 대안 출현 등은 정부와 시민 모두 국민참여의 중요성에 대한 인식을 바꾸고 있고, 여론을 형성할 수 있는 다채로운 환경과 방법이 실제로 존재함을 말한다. 따라서, 정책 관계자들도 이러한 여론 환경의 변화에 대응할 수 있는 여론 모니터링 체계를 구축하여 다양한 방식으로 표출되는 정책수요를 빠짐없이 포착할 수 있는 역량을 갖추어야 한다.

둘째, 소셜미디어의 활용 영역을 정책홍보뿐만 아니라 국민소통과 여론 모니터링의 부문으로까지 확대하고, 이를 제도적으로 뒷받침할 수 있는 조치가 필요하다. 소셜미디어의 이용이 보편화되면서 이들 신규매체를 활용하여 정책홍보 효과를 제고하려는 정부의 다양한 노력이 있었다. 여론 모니터링에 있어서도 소셜미디어는 막대한 영향을 끼치고 있는데, 소셜미디어로 수집된 여론의 빅데이터 분석 사례가 그 증거라 하겠다. 이처럼 소셜미디어의 용도가 다양화되고 있음에도 불구하고, 중앙 및 지방 정부의 대표 포털이나 소셜미디어는 여전히 정부 주도의 일방적 정책홍보의 창구 수준에만 머물고 있다. 국민이 쉽게 참여할 수 있고 정부-국민 간 소통이 활발히 이루어지며, 이런 참여와 소통의 발자취들을 빅데이터 분석 등의 수단으로 지속적으로 모니터링되어 참여·소통이 정책과정에 지속적으로 반영될 수 있는 제도적 기반 마련이 필요하다.

셋째, 맞춤형 지역정책 마련을 위해 지역 단위에서 여론의 공간적 분포와 시계열 변이를 추적할 수 있는 여론 모니터링 방법의 개발이 필요하다. 사회/여론 조사의 경우 과거 국가 단위에서, 특정 이벤트 시기에만 실시되는 것이, 이제는 광역시와 시군구 단위의 자체 정기조사로 그 형식이 변형되고 있다. 빅데이터 기반의 여론 모니터링도 문화관광부 사례처럼 전국 단위에서 실시되기도 하지만, 최근에는 지방자치단체 내부에서 지역수요의 시의성 있는 이해를 위해 빅데이터 분석기법을 활용하는 사례가 조금씩 등장하고 있다(경상북도 사례). 이와 같이 점진적 증가를 보이는 지방자치단체의 정책수단 수요에 부응하기 위해, 지역여론의 변화와 분포 패턴을 추적 분석할 수 있는 새로운 여론 모니터링 방법들을 개발해야 한다.

넷째, 여론 모니터링을 정책과정에서 쉽게 활용할 수 있도록 공동 정책자산 개념의 여론 모니터링 가이드와 툴킷(Toolkit)을 개발해야 한다. 시민의 직접참여를 통해서든, 빅데이터 분석을 통해서든 다양한 정부기관과 조직에서 여론 모니터링을 활용하는 사례가 늘고 있다. 여론 모니터링의 구체적 초점과 내용은 기관과 사안에 따라 달라지겠지만, 적용방식이나 필요한 분석기법들은 유사하다. 따라서, 기관별로 여론조사나 데이터분석을 위한 외부용역사업을 진행하기보다는, 유럽연합위원회의 WeGov 사례처럼 여론 모니터링의 다양한 수요자들이 함께 사용할 수 있는 소프트웨어 툴킷이나 상세 방법론 가이드를 개발하여 범부처적 차원의 플랫폼을 통해 공유하는 것이 효율적 예산 사용을 위해 바람직하다 하겠다.

# chapter III

## 여론 모니터링을 위한 비정형 빅데이터 시공간분석 모델



# 여론 모니터링을 위한 비정형 빅데이터 시공간분석 모델

이 장은 인터넷, 소셜미디어 등 신규매체에 반영된 여론의 시공간적 변화상을 탐색할 수 있는 비정형 빅데이터 시공간분석 모델의 기본방향과 개념을 정립하고, 상세분석기법과 그 적용사례를 정리하였다. 본 장에서 제시한 분석모델의 핵심은 비정형 빅데이터에 반영된 여론의 주요 트렌드를 파악하고, 이의 시공간적 변화 경향을 추적하는 것이다. 여론 모니터링에 있어 여론 내용과 형성 주체가 주요 관심사가 되므로, 두 측면에서 여론 시공간분석 기법의 상세 적용 방법을 살펴보았다.

## 1. 기본방향 및 개념

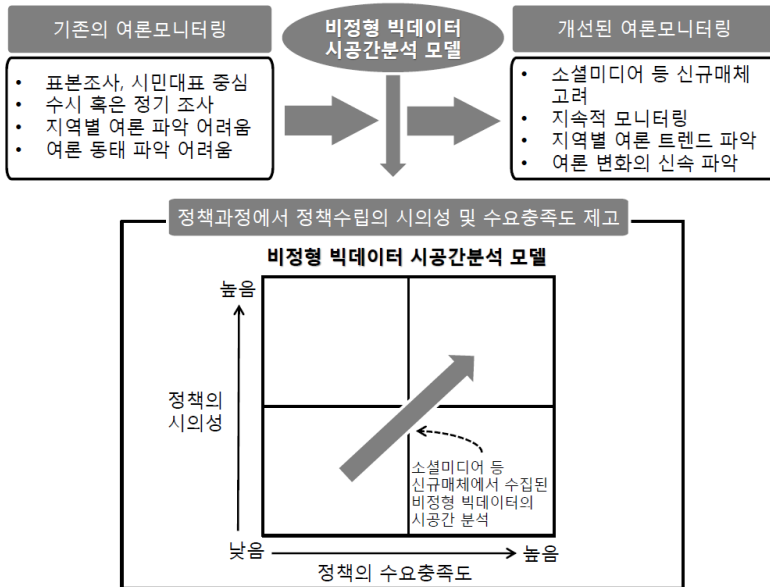
### 1) 기본방향

지금까지 여론 모니터링의 개념, 분석방법, 활용사례를 살펴보았다. 상기의 검토 결과에서 도출된 시사점 중 하나는 소셜미디어, 인터넷뉴스 등 새로운 매체를 통해 형성되는 국민 여론을 보다 적극적으로 모니터링하고 이를 위한 분석방법들도 체계적으로 활용할 필요가 있다는 것이다. 특히, 지역 및 수요 맞춤형 정책 지원을 위해서, 실시간으로 생성되는 여론 데이터에서 트렌드를 파악하고 이의 시공간적 변화를 추적할 수 있는 새로운 분석방법을 개발하여 이를 중앙 및 지방정부를 위한 공동 정책수단으로 발전시켜나갈 필요가 있다고 강조하였다. 이 시사점에 대한 대응으로,

본 장에서는 신규매체에서 시시각각 변화하는 국민 의견에서 여론의 경향성을 찾아내고 이를 시공간적으로 분석할 수 있는 방법론을 제안하고 그 개발모델을 이론적으로 정립한다.

〈그림 3-1〉은 본 연구에서 제시하는 비정형 빅데이터 시공간분석 모델에서 추구하는 기본방향과 목표를 보여준다. 정책과정에서 이용됐던 기존의 여론 모니터링은 대부분 리서치 회사에서 선정한 표본인구나, 간담회, 공청회, 시민위원회 등에 초대된 시민대표들의 생각과 의견을 조사하는 방식으로 이루어졌다. 그 결과, 기존 방법들은 최근에 등장한 인터넷뉴스나 소셜미디어 등의 신규매체에서 형성되는 여론을 충분히 반영하지 못했다. 또한, 시민조사나 시민 직접참여형 여론 모니터링 방법들은 특정 행사 시기에만 실시되거나, 그 시행 간격이 너무 길어 빠르게 변화하는 여론의 신속한 포착이 어려웠다. 뿐만 아니라 종래의 여론 모니터링 방법은 지역별 여론의 차별화 양상이나, 여론 표출의 시공간적 변이에 대한 체계적 접근을 시도하지 않아, 수요 및 지역 맞춤형 정책수요 발굴을 지원하는데 한계가 있었다.

〈그림 3-1〉 비정형 빅데이터 시공간분석 모델의 기본방향



비정형 빅데이터 시공간분석 모델은 기존 여론 모니터링 방법의 취약점을 개선하여, 정책 발굴 및 수립의 시의성과 정책의 수요충족도를 제고하고자 한다. 이를 위해 인터넷뉴스, 소셜미디어 등 새로운 매체 공간에서 형성되고 있는 여론의 체계적 분석을 지향하며, 그 수단으로서 텍스트마이닝, 사회관계망분석, 시공간분석 등 다양한 분석기법의 융합 활용을 시도한다.<sup>6)</sup> 무엇보다도 여론이 공간과 시간의 축 상에서 끊임없이 변화한다는 사실에 중점을 두고, 시기와 지역별 여론 트렌드의 파악과 이의 시공간적 변화를 지속적으로 모니터링하는데 초점을 둔다.

## 2) 개념

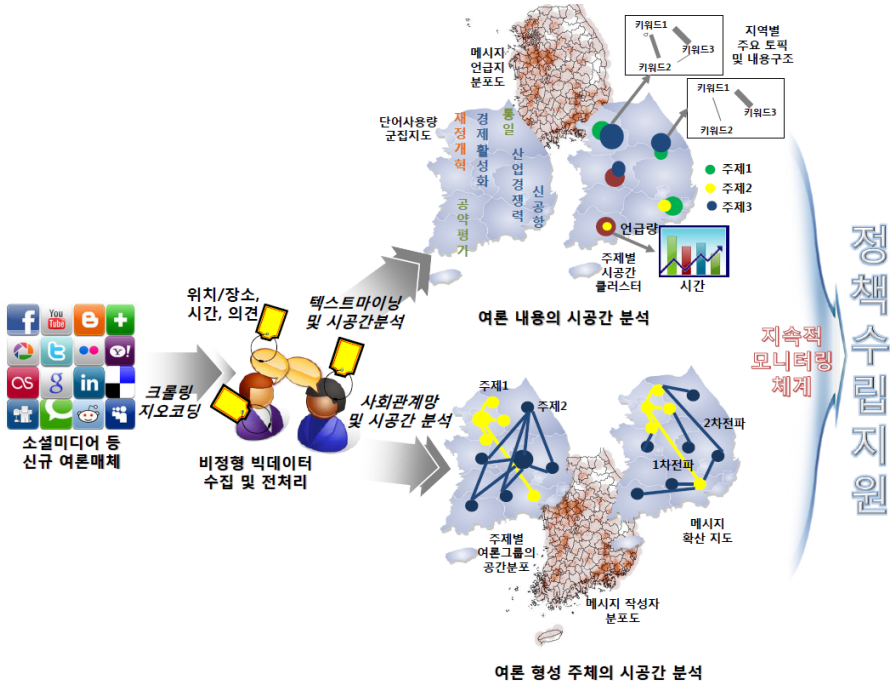
비정형 빅데이터 시공간분석 모델은 인터넷뉴스, 소셜미디어 등 신규 여론 매체에서 생성되는 비정형 빅데이터에 크롤링, 지오코딩, 텍스트마이닝, 사회관계망분석, 시공간분석 등의 방법론을 함께 적용하여, 신규매체에 반영된 국민여론의 시공간적 변화 패턴을 분석하는 자료 중심의 여론 모니터링 방법이다. <그림 3-2>는 비정형 빅데이터 시공간분석 모델의 핵심 개념을 도식으로 보여주는데, 신규매체를 통해 드러나는 국민 여론의 내용과 그 형성 과정의 주요 트렌드는 무엇인지, 그리고 이는 시공간적으로 어떻게 변화하는지를 살펴보는 것이 주요 특징이다.

<그림 3-2>에 표현된 모델의 또 다른 특징은, 텍스트마이닝, 사회관계망분석, 공간분석 등의 과정을 실시간으로 수집되는 여론 자료에 계속적으로 적용할 수 있는 체계를 수립하고자 한다는 점이다. 이런 체계 구축을 통해 단발성의 여론분석과 정책수요 발굴이 아닌, 여론과 정책수요의 지속적 모니터링을 지원할 수 있다.

---

6) 본 연구에서 제안하는 비정형 빅데이터 시공간분석 모델은 선행연구들에서 사용된 비정형 빅데이터 분석기법들을 토대로 한다. 그러나, 본 연구의 모델은 기존 분석기법들을 활용하여 비정형 빅데이터의 내용 및 관련 속성의 시공간적 측면을 분석하는데 초점을 두며, 여러 연구에서 산발적으로 활용된 분석기법들을 여론 모니터링이라는 목적 하에 포괄적 분석모델로 종합한다.

<그림 3-2> 비정형 빅데이터 시공간분석 모델의 개념

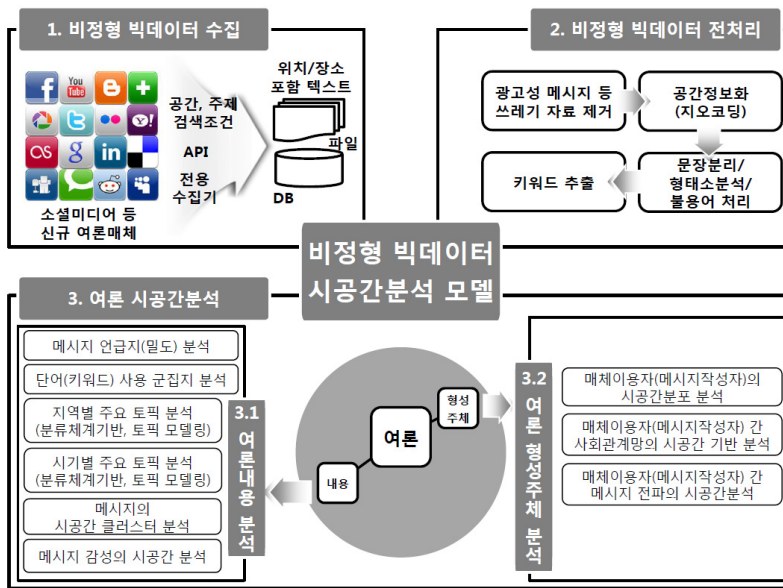


## 2. 여론 모니터링을 위한 비정형 빅데이터 시공간분석 모델

여론 모니터링을 위한 비정형 빅데이터 시공간분석 모델(<그림 3-3>)은 크게 세 가지 요소로 이루어진다. 첫 번째 요소는 여론 빅데이터 수집을 관장한다. 트위터, 페이스북, 블로그, 카페, 온라인 신문 등 다양한 여론매체에서 시민들이 논의한 내용을 API나 유·무료 소프트웨어를 이용하여 지속적으로 수집한다. 여론의 시공간 패턴을 살펴보는 것이 본 연구에서 제안한 분석모델의 주요 목표이므로, 자료 수집시 위치나 장소에 관한 검색조건이 적용되어 자료 작성자의 프로필이나 자료 내용에 위치나 장소를 포함한 문서와 메시지들만 수집하게 된다. 또한, 정책적 관심사에 따라 특정 주제어들을 검색조건으로 사용하여, 분석할 자료의 내용적 범위를 축소할 수도 있다. 수집된 자료들은 텍스트 값으로 구성되며, 파일이나 데이터베이스 형태로 저장·관리된다.

비정형 빅데이터 시공간분석 모델의 두 번째 요소는 자료 전처리 지원을 목적으로 한다. 인터넷과 소셜미디어 공간은 누구에게나 개방되어 있기 때문에, 광고성 메시지, 욕설, 중복 메시지 등 분석할 필요가 없는 자료도 포함하고 있다. 이들 쓰레기 자료의 제거가 비정형 빅데이터 전처리의 첫 단계이다. 다음으로, 여론 자료의 지역별 분류와 여론의 공간분포 패턴분석을 지원할 수 있도록 소셜미디어 메시지나 인터넷뉴스 등의 문건을 특정 위치나 장소에 할당하는 공간정보화(지오코딩) 단계를 거친다. 지리적으로 참조된 여론자료가 만들어지면, 여론 내용의 경향성을 보다 잘 파악할 수 있도록 텍스트 메시지 내의 문장들을 단어로 분리하고, 형태소 분석과 불용어 처리로 메시지마다 주된 키워드만을 남겨둔다.

<그림 3-3> 비정형 빅데이터 시공간분석 모델의 전체 구성



<그림 3-2>에서 제안된 모델의 세 번째 요소는 수집된 메시지와 문서들에 시공간분석을 적용하는 것이다. 본 연구의 2장에서 논의한 것처럼, 여론 모니터링 시 분석의 대상이 되는 여론의 속성은 내용과 형성 주체로 분류할 수 있으며, 시공간 분석은 이 두 가지 속성에 차례로 적용할 수 있다.

여론의 내용 분석은 수집된 메시지나 문서와 관련된 위치들의 공간적 밀도를 살펴보는 것에서 시작하여, 키워드별 빈발(집중) 위치를 지도로 표출하는 단계로 발전한다. 다음으로, 행정구역이나 관심사 유사도 등 여러 기준으로 지역을 분류하고, 지역별 시민들의 논의 내용을 대표 주제로 요약하여 여론 트렌드를 파악한다. 이 때, 자료 전처리 단계에서 공간화된 메시지와 문서들은 그 참조 위치에 따라 사전에 정의된 지역으로 할당된다. 지역별 여론의 대표 주제는 키워드와 상위 주제들의 관계를 정의한 분류체계나 토픽 모델링<sup>7)</sup>과 같은 텍스트마이닝 기법을 이용하여 자동 혹은 반자동으로 추출한다. 지역 여론의 주요 토픽에 대하여 관련 논의 내용의 구조를 상세히 파악해야 하는 경우, 동시출현단어 분석과 그래프 시각화 기법들을 추가 적용하여 토픽별 주제 네트워크를 제작할 수도 있다(강범일 외, 2013). 여론 내용의 트렌드가 파악되면, 여론에 투영된 사람들의 감성적 태도, 즉 긍부정 정도의 공간 패턴을 분석한다.

여론 자료에 시간 속성이 포함된 경우, 위에서 논의된 공간분석 방법들을 월, 분기 등 일정 시간 간격으로 적용하여, 그 변화 동향을 검토할 수 있다. 이 때, 시계열 지도, 애니메이션 등을 활용하여 여론 내용의 시공간적 변화를 손쉽게 살펴볼 수 있어야 한다. 또한, 메시지나 문서들 간의 시공간 거리를 계산하여, 일정 공간거리 내에서 특정 시간대에 최소 몇 건 이상의 문서와 메시지가 생성된 시공간 클러스터들을 탐지하고, 클러스터별 주요 토픽구성을 분석할 수 있다.

비정형 빅데이터에 시공간분석을 적용할 때, 두 번째 주요 분석대상은 여론형성의 주체이다. 즉, 신규매체라는 사이버 공간 상에서 여론형성에 참여하는 사람들은 누구인지, 사이버 공간 상에서 활발히 상호작용하는 사람들이 지리적으로도 인접한 지, 사이버 공간 상에서의 여론 전파가 지리적 공간에서는 어떤 모습을 띠는지 등이 분석의 초점이 된다. 여론 내용 분석의 경우와 같이, 여론형성 주체에 대한 분석도 매체 이용자나 메시지/문서 작성자의 공간적 분포와 이의 시간적 변화를 살펴보는 것으로 시작한다. 다음으로, 댓글<sup>8)</sup>, 팔로잉<sup>9)</sup>, 리트윗<sup>10)</sup> 등의 형태로 의견을 교환함으

7) 토픽 모델링은 대량의 문서 묶음에서 사용된 키워드들의 동시 사용 패턴을 토대로 이 문서 묶음을 대표하는 주요 토픽들을 추출해내는 텍스트마이닝 기법으로, 본 장의 4절에서 상세히 설명한다.

로써 상호작용하는 사람들의 사회관계망을 추출하고 이를 지리적 공간으로 투영하여, 사회관계망의 공간적 범위와 그 변화상을 분석한다. 마지막으로, 리트윗, 링크, 공유 등의 방식으로 특정 생각과 의견이 확산된 경우, 생각과 의견의 원천 생산자와 2·3차 전달자의 사회관계망을 구분하고 이들을 지도 상으로 표출하여 사이버 공간 상의 정보 확산이 실제 지리 공간에서는 어떤 구조를 띠며, 그 구조가 시간이 흐르면서 어떻게 변화하는지를 살펴본다.

본 장의 아래 절들에서는 비정형 빅데이터 시공간분석 모델의 개별 구성요소를 이루는 자료 수집·처리·분석 기법들의 상세 내용을 설명한다.

### 3. 여론 관련 비정형 빅데이터의 수집과 전처리

#### 1) 여론 관련 비정형 빅데이터의 수집

본 연구에서 대상으로 하는 여론 관련 비정형 빅데이터(이하 ‘여론 빅데이터’)의 수집에 있어 가장 먼저 논의해야 할 이슈는 자료의 대표성이다. 본 연구의 2장에서 논의한 것처럼, 인터넷뉴스, 소셜미디어 등의 신규매체는 대다수의 국민 보다는 인터넷, 모바일 서비스 등을 이용할 수 있는 경제력을 갖춘 20~40대의 도시 거주자들이 주로 이용하며, 남성보다는 여성의 이용 비율이 다소 높다(이윤희, 2014). 따라서, 여론 빅데이터를 수집하고자 할 때 먼저 여론 모니터링의 대상이 신규매체 이용자들의 프로파일에 해당하는지 고려해야만 한다.

여론 빅데이터의 공급원은 크게 인터넷뉴스 미디어와 소셜미디어 서비스로 구분할 수 있다. <표 3-1>은 여론 빅데이터의 주요 공급원과 그 예시를 보여주는 데, 이들 자료는 크롤링(Crawling), FTP(File Transfer Protocol), Open API(Application Programming Interface), RSS(Really Simple Syndication), 스트리밍(Streaming) 서비스 등의 기술을 활용하여 수집할 수 있다(<표 3-2>).

8) “인터넷 게시물 밑에 남길 수 있는 짧은 글”(http://ko.wikipedia.org/wiki/맷글, 2015년 1월 4일 오후 8시 15분 최종접속)

9) 트위터 등의 SNS에서 다른 사람이 쓴 글을 구독하는 행동

10) 트위터 서비스 이용자가 다른 사람이 쓴 글을 그대로 자신의 게시물로 등록하는 행동

<표 3-1> 여론 빅데이터 공급원의 종류 및 예시

| 데이터 공급원          |                | 정의/예시                                                                                                                                                                                                                                       |
|------------------|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 대분류              | 소분류            |                                                                                                                                                                                                                                             |
| 인터넷<br>뉴스<br>미디어 | 인터넷<br>뉴스서비스   | <ul style="list-style-type: none"> <li>「신문 등의 진흥에 관한 법률」에 따른 신문, 인터넷신문, 「뉴스통신진흥에 관한 법률」에 따른 뉴스통신, 「방송법에 따른 방송 및 잡지 등 장기간행물의 진흥에 관한 법률」에 따른 잡지 등의 기사를 인터넷을 통하여 계속적으로 제공하거나 매개하는 전자간행물</li> <li>예: 네이버 뉴스, 법률뉴스, 부동산114 인터넷뉴스서비스 등</li> </ul> |
|                  | 인터넷신문          | <ul style="list-style-type: none"> <li>컴퓨터 등 정보처리능력을 가진 장치와 통신망을 이용하여 정치·경제·사회·문화 등에 관한 보도·논평 및 여론·정보 등을 전파하기 위하여 간행하는 전자간행물</li> </ul>                                                                                                     |
|                  | 인터넷<br>멀티미디어방송 | <ul style="list-style-type: none"> <li>광대역통합정보통신망 등을 이용하여 양방향성을 가진 인터넷 프로토콜 방식으로 일정한 서비스 품질이 보장되는 가운데 텔레비전 수상기 등을 통하여 이용자에게 실시간 방송프로그램을 포함하여 데이터·영상·음성·음향 및 전자상거래 등의 콘텐츠를 복합적으로 제공하는 방송</li> </ul>                                          |
| 소셜<br>미디어<br>서비스 | 카페·클럽·<br>동호회  | <ul style="list-style-type: none"> <li>카페/클럽의 예: 다음카페, 네이버 클럽 등</li> <li>동호회의 예: 사진 동호회 '포토아지트', 쇼핑물 동호회 '내가게' 등</li> </ul>                                                                                                                 |
|                  | 블로그·<br>미니홈피   | <ul style="list-style-type: none"> <li>블로그의 예: 다음/네이버 블로그, 이글루스, 티스토리 등</li> <li>미니홈피: 싸이월드 미니홈피 등</li> </ul>                                                                                                                               |
|                  | 마이크로 블로깅       | <ul style="list-style-type: none"> <li>수평적 서비스(다양한 정보 제공): 트위터, 페이스북 등</li> <li>수직적 서비스(특정 관심분야 정보만 제공): 핀터레스트, 허핑턴포스트, 팟잇컴, 인스타그램, 패스, 포스퀘어 등</li> </ul>                                                                                   |
|                  | 인스턴트 메신저       | <ul style="list-style-type: none"> <li>카카오톡, 구글플러스, 네이트온, 버디버디 등</li> </ul>                                                                                                                                                                 |
|                  | 인맥관리 서비스       | <ul style="list-style-type: none"> <li>페이스북, 마이스페이스, 링크나우, 토씨 등</li> </ul>                                                                                                                                                                  |
|                  | 가상현실 서비스       | <ul style="list-style-type: none"> <li>세컨드라이프, 싸이월드 미니라이프 등</li> </ul>                                                                                                                                                                      |

\*출처: 한국정보화진흥원·빅데이터전략연구센터(2012), 김위근·김춘식(2010)

여론 빅데이터 공급원의 유형에 따라 적용 가능한 자료수집 기술이 다른데, <표 3-3>은 매체 유형별 자료수집 방법을 보여준다. 크롤링은 뉴스, 블로그, 인터넷 카페, 트위터, 페이스북, 포스퀘어 등 다양한 매체에서 생성되는 자료를 수집하는데 적용 가능하다. 한편, Open API는 유료 혹은 무료 API를 제공하는 서비스들에만 적용할 수 있으며, 서비스에 따라 제공하는 API 종류가 다르다. 예를 들어, 트위터 서비스의 경우 무료로 이용 가능한 Search API와 Public Streaming API를 제공하며, 유료로 이용할 수 있는 Firehose Streaming API도 함께 제공한다. 유료로 서비스하는 Firehose Streaming API는 사용자가 공개를 동의한 모든 트윗 메시지를 제공한다. 그러나, 무료 서비스인 Search API와 Public Streaming API는 수집된 트윗 메시지 중 일부만 샘플링을 통해 제공한다.

Public Streaming API는 검색 시점부터 생성된 메시지 중 검색조건에 맞는 것들을 1% 무작위 샘플링하여 지속적으로 제공한다.<sup>11)</sup> 한편, Search API는 검색어가 반드시 필요하며 검색조건에 부합하는 메시지 중 검색 시점부터 6~9일 전에 생성된 것들 위주로 자료를 제공하나, 그 결과조차도 트위터에서 수집한 전수 데이터는 아니다.<sup>12)</sup>

<표 3-2> 여론 빅데이터 수집 기술

| 구분       | 특징                                                                                                | 비고         |
|----------|---------------------------------------------------------------------------------------------------|------------|
| 크롤링      | • SNS, 뉴스, 웹 정보 등 인터넷 상에서 제공되는 웹문서 정보 수집                                                          | 웹문서 수집     |
| FTP      | • TCP/IP 프로토콜을 활용하는 인터넷 서버로부터 각종 파일을 수신                                                           | 파일 수집      |
| Open API | • 서비스, 정보, 데이터 등을 어디서나 쉽게 이용할 수 있도록 개방된 API로 데이터 수집방식 제공<br>• 다양한 어플리케이션을 개발할 수 있도록 개발자와 사용자에게 공개 | 실시간 데이터 수집 |
| RSS      | • 웹 기반 최신정보를 공유하기 위한 XML 기반 콘텐츠 배급 프로토콜                                                           | 콘텐츠 수집     |
| 스트리밍     | • 인터넷에서 텍스트, 음성, 오디오, 비디오 데이터를 실시간으로 지속적으로 수집할 수 있는 기술                                            | 실시간 데이터 수집 |

\*출처: 미래창조과학부·한국정보화진흥원빅데이터전략센터(2014)의 <주요 데이터 수집 기술> 표(p.14) 재구성

<표 3-3> 비정형 빅데이터 수집 특성

| 수집 방법     | 휴대 전화 | 이메일 | 논문/특허 | 뉴스  | 블로그 | 인터넷 카페 | 트위터 | 페이스북 | 포스퀘어 |
|-----------|-------|-----|-------|-----|-----|--------|-----|------|------|
| 데이터 베이스   | ○     | ○   | ○     | △   | ×   | ×      | ×   | ×    | ×    |
| 크롤링       | ×     | ×   | △     | ○   | ○   | △      | △   | △    | △    |
| RSS       | ×     | ×   | ×     | ○   | ○   | △      | ×   | ×    | ×    |
| Streaming | ×     | ×   | ×     | △   | ×   | ×      | △   | ×    | ×    |
| Open API  | ×     | ×   | ○     | △   | △   | ×      | △   | △    | △    |
| Agent     | △     | △   | ×     | ×   | ×   | ×      | △   | △    | ×    |
| 수집 간격     | 1분    | 1시간 | 1개월   | 1시간 | 6시간 | 6시간    | 1분  | 20분  | 1일   |
| 최소 수명 주기  | 3개월   | 6개월 | 5년    | 5년  | 3년  | 2년     | 1년  | 1년   | 6개월  |

\*출처: 최광선(2012)

11) 트위터 Streaming API에 관한 상세정보는 <https://dev.twitter.com/streaming/public>(2014년 12월 10일 오후 11시 55분 최종 접속)를 참조할 있다.

12) 트위터 Search API에 관한 상세정보는 <https://dev.twitter.com/rest/public/search>(2014년 12월 10일 오후 11시 55분 최종 접속)를 참조할 있다. Search API의 경우 1회 추출 가능한 메시지 건수, 시간당 API호출 건수, 트위터 계정별 최대 메시지 반환 건수 등 여러 제약이 있기 때문에, 이들 제약의 조합으로 인해 검색조건에 부합하는 트윗 메시지 전수를 추출하는 것은 가능하지 않다.

트위터 API의 예시에서 설명한 것처럼, 특정 매체에서 여론 빅데이터를 수집할 때 가능한 모든 데이터를 수집할 수도 있지만 분석의 관심사나 데이터 처리능력에 따라 각종 검색조건을 부여하는 경우가 많다. 본 연구의 분석모델은 여론 트렌드의 시공간적 변이에 초점을 두므로, 검색조건에 공간적 범위가 사용되는 경우가 많다. 예를 들어, 트위터 Search API를 이용할 때 “near:국토연구원 within:15km”와 같은 연산자를 이용하여 특정 위치나 장소로부터 일정 거리 이내에서 작성된 메시지만을 추출할 수 있다. 그러나 트위터와 달리 문서와 특정 위치/장소를 연결할 수 없는 경우도 있다. 이 때, 분석목적에 부합하기만 하다면, 먼저 가능한 모든 문서를 수집하고 각 문서 내에서 위치/장소 관련 단어들을 추출하여 잠정적 위치정보를 부여한 후(다음 소절 참조) 특정 공간범위에 해당하는 문서들만을 뽑아낼 수 있다. 공간적 범위 외에도 시간범위나 특정 주제어를 검색조건으로 이용할 수 있다.

다양한 기술로 수집된 여론 빅데이터는 관리를 위해 그 원본을 파일이나 데이터베이스로 저장해둔다. 데이터베이스가 이용되는 경우, 데이터의 용량과 수집 주기, 여론 트렌드 분석과 시공간분석의 실시간 지원 여부에 따라 관계형부터 NoSQL 데이터베이스 등 여러 기술 중 서비스 목적에 부합하는 것으로 선택할 수 있다.

## 2) 여론 관련 비정형 빅데이터의 전처리

〈그림 3-3〉에서 볼 수 있듯이, 수집된 여론 빅데이터는 여러 단계의 전처리 과정을 거친다. 이 중 첫 단계는 분석에 부적합거나 중복된 데이터 항목을 제거하여 수집된 여론 데이터의 품질을 개선하는 것이다. 텍스트 기반의 온라인 문서나 메시지(이하 ‘메시지’)가 주를 이루는 여론 빅데이터에서 데이터 품질에 영향을 미치는 가장 큰 요소는 광고성 혹은 악의성 스팸이다. 스팸의 여과를 위해 다양한 기법이 적용될 수 있는데, 동일 사용자가 동일 URL이나 단어를 포함한 메시지를 빈번히 작성하는 경우, 특정 매체에서 생성한 메시지를 소수의 사용자가 독점하는 경우, 특정 사용자가 유사 내용의 메시지를 지속적으로 작성하는 경우를 집중 관리한다(하수옥 외, 2012).

여론 빅데이터 전처리의 두 번째 단계는 본 연구에서 제시한 시공간분석 모델에서 매우 중요한데, 메시지를 위치/장소와 연계, 즉 지오코딩하는 공간정보화이다. <표 3-4>는 신규매체에서 수집된 메시지와 관련된 위치/공간 정보의 유형과 그 의미를 설명하고 있다.

<표 3-4> 여론 자료 관련 위치/장소 정보의 유형

| 구분                 | 위치/장소 유형       |                                                                     | 설명/의미                                                                                                                                                                                                                                                                             |
|--------------------|----------------|---------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 문서/<br>메시지         | 작성<br>위치/장소    | 좌표<br>(예: 126.9, 37.8),<br>위치/장소 이름<br>(예: 안양시,<br>**아파트),<br>이동 경로 | <ul style="list-style-type: none"> <li>특정 위치/장소에서 논의된 생각과 의견으로, 반드시 해당 위치/장소에 관한 생각과 의견은 아니나 그 위치/장소와 관련되어 있을 가능성이 높음</li> <li>한 사용자가 작성한 문서/메시지들의 작성 위치/장소 정보가 일정 기간 이상 누적될 경우, 이 사용자가 활동하고 있는 공간의 범위와 그 일관성을 파악할 수 있으며 이를 통해 작성 위치/장소가 그 사용자의 거주지나 직장일 가능성을 추정할 수 있음</li> </ul> |
|                    | 언급<br>위치/장소    | 좌표,<br>위치/장소 이름                                                     | <ul style="list-style-type: none"> <li>특정 위치/장소에 관한 생각과 의견</li> </ul>                                                                                                                                                                                                             |
| 문서/<br>메시지의<br>작성자 | 등록<br>위치/장소    | 좌표,<br>위치/장소 이름                                                     | <ul style="list-style-type: none"> <li>특정 위치/장소에서 활동한다고 밝힌 사용자의 생각과 의견이나, 그 사용자가 반드시 관련 위치/장소에서 거주하거나 일한다고는 볼 수는 없음</li> </ul>                                                                                                                                                    |
|                    | IP 등록<br>위치/장소 | 좌표,<br>위치/장소 이름                                                     | <ul style="list-style-type: none"> <li>문서/메시지가 시스템을 통해 등록되는 경우, 작성자의 등록 위치/장소가 아닌 시스템 운영 IP가 등록된 위치/장소가 주로 기록됨</li> </ul>                                                                                                                                                         |

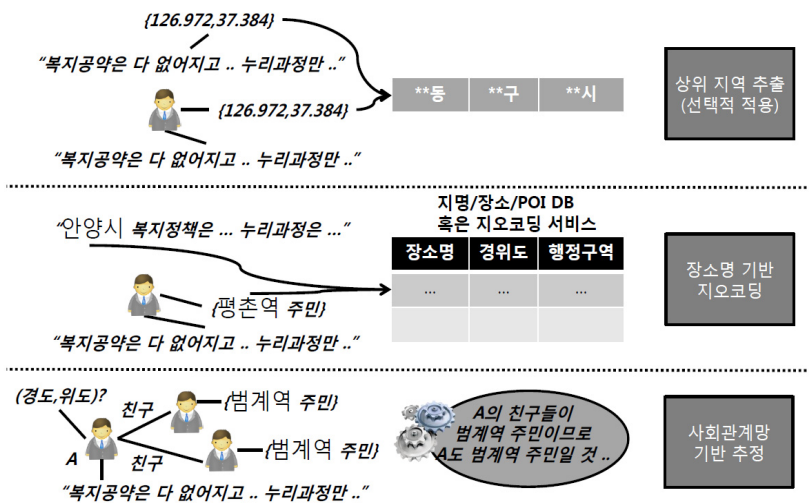
여론을 포함한 메시지는 작성, 언급, 작성자 위치/장소 정보를 포함하는데, 각각 의미하는 바가 다르다. 메시지가 특정 위치와 장소를 언급하고 있는 경우, 그 위치/장소 및 해당 지역에 관한 여론으로 간주할 수 있다. 반면, 메시지에 작성 혹은 작성자의 위치/장소가 메타데이터의 형태로 추가되어 있는 경우, 이들 자료에 기재된 내용은 그 위치/장소에 머물렀던 사람들의 여론으로 반드시 그 위치/장소에 관한 논의라고는 할 수는 없다. 다만, 작성 혹은 작성자의 위치/장소를 포함하므로 그 위치/장소와 관련성은 높을 것으로 간주할 수 있다. 또한, 특정 사용자의 메시지를 작성한 위치들을 일정 기간(3개월<sup>13)</sup> 등) 이상 수집할 수 있고, 이 기간

13) “한국에서는 상주의 개념을 3개월로 한정하여 ① 3개월간 계속 살았거나, ② 3개월을 살지 않았어도 앞으로 그 곳에서 살 기간과 이미 살아온 기간을 합쳐 3개월 이상이 되거나, ③ 역시 3개월을 계속 거주한 일이 없지만 앞으로 3개월 이상을 계속 살 것으로 예상되는 장소를 상주지로 규정

동안 사용자가 일관되게 특정 위치와 시점에서 메시지를 작성했다면 그 위치를 거주지(이른 아침이나 저녁 시간)나 직장(낮 시간)으로 간주할 수 있다. 이 경우, 수집된 메시지들은 지역 주민이나 주요 활동인구의 여론으로 볼 수 있다(Andrienko 외, 2013; Ghosh·Guha, 2013).

<그림 3-4>는 텍스트 데이터의 지오프코딩 사용되는 주요 기법들을 보여준다. 트위터 등 신규매체에서 수집된 메시지가 GPS 기반의 작성 위치나 작성자의 위치를 명시한 경우, 추가적인 공간정보화가 반드시 필요하지는 않다. 그러나, 분석 목적에 따라 메시지 위치를 포함하는 상위 행정 구역이나 집계 지역을 파악하는 지오프코딩을 수행할 수 있다.

<그림 3-4> 텍스트 데이터의 지오프코딩 방법



\*출처: 하수옥 외(2012)의 논의내용 재구성

신규매체에서 수집되는 메시지 중 상당수는 정확한 작성/작성자 위치를 포함하고 있지 못하며, 메시지 내에서 장소 이름을 언급하거나 메시지 작성자의 등록 장소를 텍스트 값으로 명기한 경우가 많다. 이 경우, 장소 이름에 해당하는 단어들을 추출하고

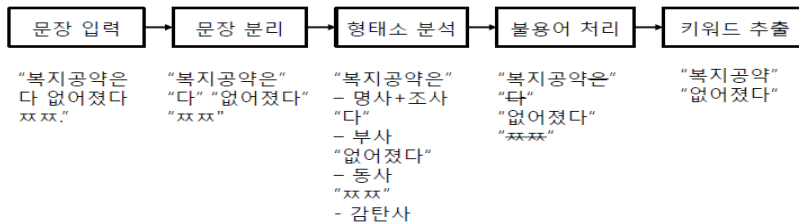
하였다.” (<http://terms.naver.com/entry.nhn?docId=1110027&cid=40942&categoryId=31612>; 2015년 1월 4일 오후 10시 44분 최종 접속)

지명/장소명/POI<sup>14)</sup> 데이터베이스나 지오코딩 서비스를 통해 해당 장소에 관한 지리 좌표나 행정구역 등 위치참조정보를 계산해야 한다. 이 때, 메시지 내에 여러 장소이름이 언급되거나 관련 시공간 맥락 정보가 부재하여 장소 이름이 나타내는 정확한 위치가 불분명할 수 있다. 이 경우, 분석 목적에 따라 장소와 관련된 위치를 결정하는 규칙을 설정해야 하며, 관련 예는 이영주 외(2013)의 연구를 참조할 수 있다.

여론 모니터링시 수집된 메시지와 그 작성자에 관한 위치정보가 전혀 없지만, 관련 위치정보를 추출해야 하는 경우도 있다. 이 경우, 메시지 작성자의 사회관계망 데이터를 추출하여 작성자 친구나 지인들의 위치정보를 기반으로 작성자가 활동하는 주요 공간범위를 추정하는 방법이 있다. 그러나, 그 추정 결과의 신뢰도가 낮으므로 여론 빅데이터의 시공간분석시 유의해야 한다.

여론 빅데이터 전처리의 나머지 단계에서는 메시지나 문서에 포함된 문장들에서 실제 분석에 쓰일 주요 키워드들만을 정제하여 뽑아낸다(<그림 3-5>). 이를 위해 먼저 메시지를 구성하는 문장들을 구문 단위로 세분하고(문장 분리), 각 구문을 구성하는 단어들의 형태소를 분석하여(형태소 분석) 명사, 동사, 형용사, 부사 등과 같은 주요 품사를 포함하는 단어들만 남긴다. 남겨진 단어들에서 조사, 어미 등 불용어를 모두 처리하여(불용어 처리) 주요 키워드들만을 추출한다(키워드 추출). 분석필요에 따라, 키워드들은 서로 통합되기도 하고 유사의미의 다른 키워드로 대체되기도 한다.

<그림 3-5> 텍스트 데이터의 전처리 과정



14) Points of Interest의 약자로 주요 랜드마크, 건물, 상점, 시설물 등 사람들의 관심을 많이 끄는 장소들의 위치 정보를 포함한 데이터베이스

여론 빅데이터 전처리의 결과 <그림 3-6>과 같은 구조의 분석용 초기 자료가 만들어진다. 본 연구에서는 여론의 지속적 시공간분석을 위한 체계 구축을 지향하며, 이 경우 <그림 3-6>에 제시된 초기 자료가 효율적 분석을 위해 재구조화되기도 한다. 추가적 데이터 변형은 분석기법에 따라 달라지며, 이에 대해서는 세부 분석기법을 다루는 다음 소절에서 논의한다.

<그림 3-6> 비정형 빅데이터 전처리 결과

| 메시지 ID | 메시지 작성자 ID | 메시지 작성 시점 | 메시지 원본 | 메시지 키워드 | 위치/장소 유형 | X (경도) | Y (위도) | 장소명 | 음면등 | 시군구 | 시도 | 기타 지역 | ... |
|--------|------------|-----------|--------|---------|----------|--------|--------|-----|-----|-----|----|-------|-----|
| ...    |            |           |        |         |          |        |        | ... | ... |     |    |       |     |
|        |            |           |        |         |          |        |        |     |     |     |    |       |     |

#### 4. 여론의 시공간분석

이 소절에서는 여론의 내용과 형성주체의 시공간 패턴을 분석할 수 있는 다양한 방법들을 설명한다. 이영주 외(2013), Andrienko 외(2013), Ghosh·Guha (2013), Leetaru(2011), Ma(2012), Neuhaus(2013)의 연구 등 국내외 최신 연구 결과를 토대로 이들 분석방법의 기본원리와 적용 사례를 제시한다.

##### 1) 여론 내용의 시공간분석

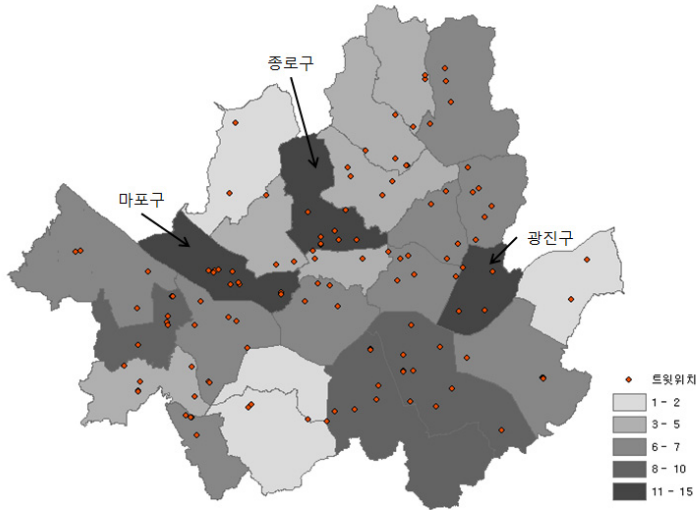
###### (1) 메시지 작성위치의 공간분포 분석

인터넷 카페, 트위터, 페이스북 등의 신규매체를 이용하는 국민들은 어디에서 무슨 이슈에 대해 주로 논의하고, 이 패턴은 어떻게 변할까? 여론 모니터링시 누구나 쉽게 제기할 수 있는 질문이다. 수집한 여론 데이터를 이용하여 이 질문을 답하고자 할 때, 먼저 수집된 메시지가 어디에서 작성됐는지 그 공간분포를 살펴볼 수 있다.

메시지 작성위치의 공간분포를 분석하기 위한 기법은 다양하며, <그림 3-6>과 같은 구조의 기초 분석 자료에 이들 기법을 그대로 적용할 수 있다. 가장 단순한 분석방법은 개별 메시지의 언급 위치를 지도 상에 점으로 뿌려주거나,

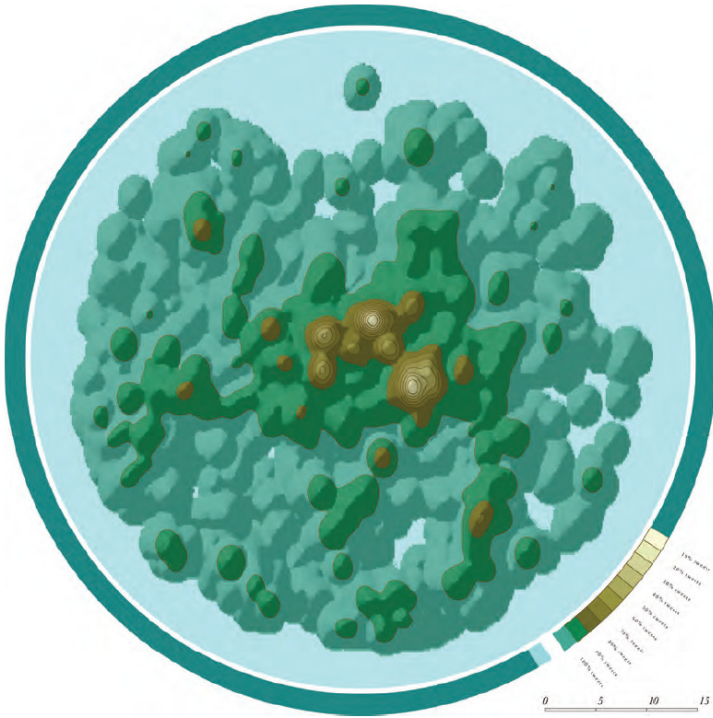
격자, 읍면동, 시군구 등의 구역을 기반으로 메시지 작성건수를 단계구분도의 형태로 표현하는 것이다. <그림 3-7>은 서울시를 대상으로 한 가상의 사례로, 159건의 트윗의 위치와 그 건수를 단계구분도로 나타낸 예시이다.

<그림 3-7> 서울지역 트윗의 위치도 및 트윗건수의 단계구분도 예시



메시지 위치 지도나 단계구분도 외에, 메시지 작성이 주변 지역에 비해 특별히 활발한 핫스팟 지역이 어디인지 분석할 필요가 있다. 이 경우, 분석의 초점은 면적, 인구 등 메시지 모집단 관련 변수 대비 실제 메시지 작성건수가 기대 이상으로 높은 지역을 찾는 것이다. 점 밀도(Point Density), 커널(Kernel) 밀도 분석 등이 핫스팟 분석에 자주 활용된다. 커널 밀도 분석의 경우 일정 간격의 격자지도를 생성하고, 각 격자의 중심점에서 특정 반경 내에 있는 점들의 커널 밀도를 계산한 후 그 합계를 표준화한 결과를 연속 면(Surface)의 형태로 표현한다. 각종 밀도 분석의 결과는 그 밀도 값을 일정 간격으로 구분하여 등치선도(Isolines)의 형태로도 나타낼 수 있다. <그림 3-8>은 서울 지역에서 생성된 트윗 메시지의 공간밀도를 연속 면과 등치선도로 나타낸 예시이다.

<그림 3-8> 서울지역 트윗의 공간밀도 및 등치선도의 예시



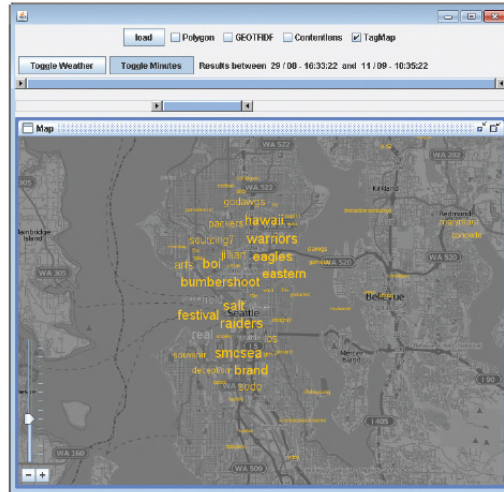
\*출처: Neuhaus(2013)의 Figure 9.73 (p.335)

## (2) 단어사용 군집지 분석

앞서 논의했듯이, 어디에서 무슨 이슈를 논의하는가는 여론분석시 자주 제기되는 질문이며, 이에 대한 답은 여론의 지역차를 이해하는 데 핵심적 역할을 할 수 있다.

지역별 여론 이슈를 직관적으로 파악할 수 있는 분석방법 중 하나는 Andrienko 외(2013)가 소개한 단어사용 군집지 분석(Term-usage Cluster Analysis)이다. 이 방법에서는 메시지에 포함된 주요 단어들이 기대 이상으로 빈번히 사용되는 지점들을 찾고, 지명 대신 이 지점들과 관련된 단어들을 지도 상에 표현한다. <그림 3-9>는 단어사용 군집지 분석의 결과를 예시적으로 보여주는데, 미국 시애틀 지역에서 작성된 위치 기반 트윗 메시지들을 이용하였다.

<그림 3-9> 단어사용 군집지 분석 결과의 예시



\*출처: Andrienko 외(2013)의 Figure 2(a) (p.4)

단어사용 군집지 분석의 원리는 <그림 3-10>에서 설명하고 있다. <그림 3-6>의 기초 분석 자료에서 모든 메시들에 포함된 키워드들 중 고유한 것들을 추출한다. 이 때, 각 고유 키워드를 포함한 메시지들의 위치와 시점, 즉 시공간 좌표를 함께 추출하여 벡터 행렬 형태로 저장한다. 다음으로, 고유 키워드 관련 시공간 좌표 행렬에 k-means 군집 분석을 적용하여, 고유 키워드를 포함하고 시공간 거리가 가까운 메시지들로 구성된 소수의 시공간 핫스팟들을 찾아낸다. 이 군집분석 과정을 고유 키워드 각각에 적용한 후, 키워드별로 추출된 핫스팟들의 중심점과 메시지 밀도를 계산하여, 고유 키워드의 가중치로 할당한다. 마지막으로, 각 키워드와 연관된 가중치와 시공간 핫스팟들의 중심점 좌표를 이용하여, 키워드를 지도 상에 표시하여 <그림 3-9>와 같은 단어사용 군집지도를 만들어낸다.

지도 상에 표시된 모든 키워드는 시간 좌표를 가지므로, 슬라이드바(Slide Bar)와 같은 탐색적 분석도구를 이용하여 단어사용 공간패턴의 시간적 변화를 동적으로 살펴볼 수 있다. 또한, 분석가가 특정 키워드를 선택하면, 이와 연관된 모든 메시지들의 위치를 지도 상에 표시하여 특정 이슈가 논의되는 공간적 범위도 세밀하게 분석할 수 있다.

<그림 3-10> 단어사용 군집지 분석의 원리



\*출처: Andrienko 외(2013)의 논의 내용을 토대로 도식화

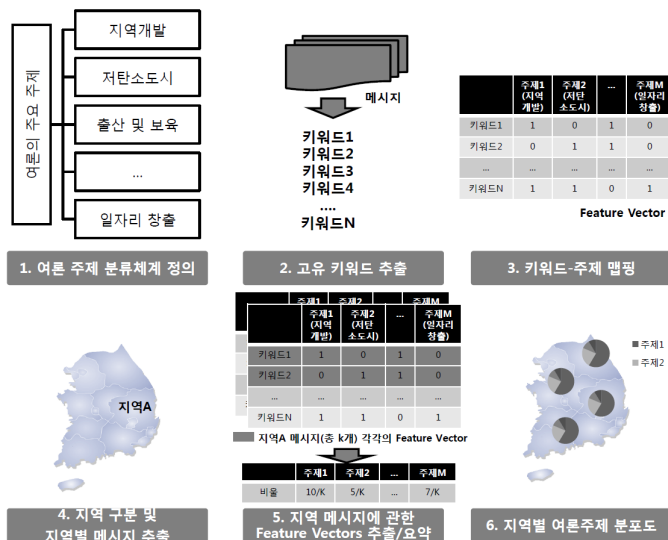
### (3) 여론 트렌드 분석(토픽 분석)

여론 관련 메시지의 작성 위치 및 주요 단어가 중심으로 사용된 군집지역 탐색 등은 여론 데이터의 전반적 시공간분포 패턴 파악에 유용하다. 그러나, 이들 결과는 방대한 여론 빅데이터 내에 숨어있는 주요 토픽 및 트렌드를 보여주는 데에는 한계가 있다. 메시지 작성위치 지도나 밀도의 경우 대부분 인구밀도와 유사한 공간패턴을 보여 많은 정보를 제공하지는 못한다(Andrienko 외, 2013). 한편, 단어사용 군집지 분석의 경우 지도 시각화를 위해 k-means 군집분석을 통해 추출된 시공간 핫스팟들의 중심지를 이용하는데, 이들 핫스팟의 위치는 그 의미의 해석이 쉽지 않다(Andrienko 외, 2013). 키워드 이상의 유의미한 정보를 끌어내기 위해서는, 메시지들을 그 의미와 맥락에 따라 특정 주제 그룹으로 분류하고 이 주제들의 시공간패턴을 분석할 필요가 있다. 이를 위한 두 가지 방법이 분류체계와 토픽 모델링 기반의 여론 토픽 분석이다.

## ① 분류체계 기반 방법

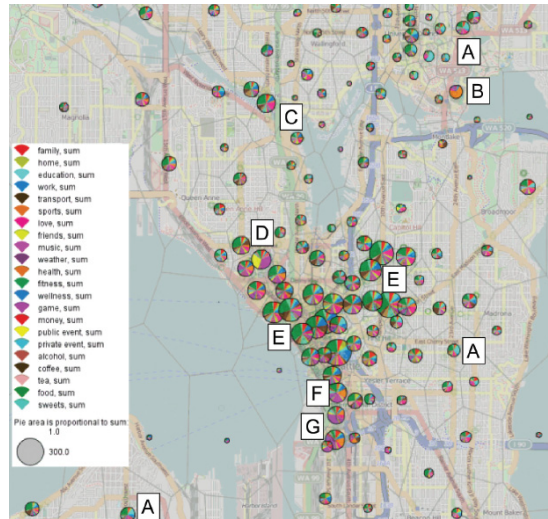
〈그림 3-11〉은 분류체계 기반의 방법의 분석 원리를 도식화한 것이다. 이 분석방법의 첫 단계는 여론 주제들을 정의한 온톨로지나 주제 트리 등의 분류체계를 정의하는 것이다. 다음으로, 수집한 메시지를 구성하는 고유 키워드들을 추출한다. 세 번째 단계로, 각각 고유 키워드가 어떤 주제에 해당하는지를 결정하여, 행은 키워드이고 열은 주제이며 셀의 값은 1이나 0을 포함하는 이진 행렬인 Feature Vector를 생성한다. 네 번째로, 사전에 정의되었거나(행정구역 등), 메시지 작성위치의 시공간적 거리 분석으로 도출된(Voronoi Tessellation 등) 지역구분체계를 이용하여, 각 소지역에서 작성된 메시지들을 추출한다. 다섯 번째로, 각 지역 메시지별로 Feature Vector를 추출하여 합산한다. 마지막으로, Feature Vector 합산 결과를 토대로, 지역 메시지에서 자주 논의되는 여론 주제들의 분포를 차트지도로 시각화하여 지역별 주요 여론 토픽을 직관적으로 표출한다. 〈그림 3-12〉는 미국 시애틀에서 생성된 트윗 메시지에 분류체계 기반의 방법을 적용한 사례로, Voronoi Tessellation을 통해 소지역 경계를 결정하였다.

〈그림 3-11〉 분류체계 기반 여론 토픽 분석의 원리



\*출처: Andrienko 외(2013)의 논의 내용을 토대로 도식화

<그림 3-12> 분류체계 기반 여론 토픽 분석 결과의 예시



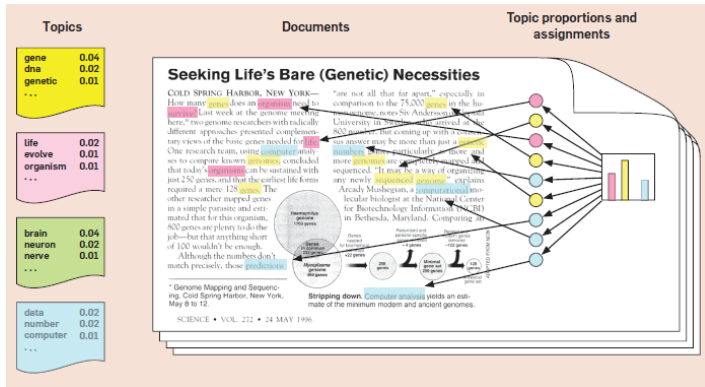
\*출처: Andrienko 외(2013)의 Figure 3 (p.6)

## ② 토픽 모델링 기반 방법

여론 토픽 분석을 위한 두 번째 방법은 토픽 모델링(Topic Modeling)이다. 토픽 모델링은 텍스트 자료 내 단어들의 사용 빈도를 통계적으로 분석하여 전체 입력 자료를 관통하는 잠재 주제들을 추출해내고 이 주제들 간의 연관관계나 시간적 변화를 살펴보는 텍스트마이닝 알고리즘이다(Blei, 2012).

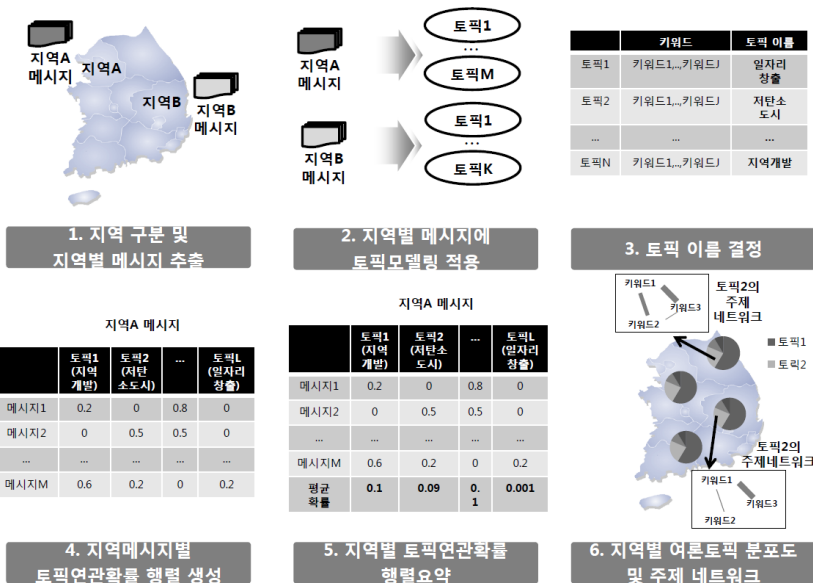
토픽 모델링에 대한 접근법은 다양한데, Latent Dirichlet Allocation(LDA) 방법이 가장 단순하면서도 널리 활용되고 있다(Blei, 2012). <그림 3-13>에서 볼 수 있듯이, LDA 방법은 텍스트 자료들은 알려지지 않은 여러 주제로 구성되어 있고, 이 주제들은 다양한 단어들로 표현된다고 전제한다. 또한, 텍스트자료 내 문서마다 이 주제들의 구성 비율이 다르며, 이 비율을 결정하는 확률분포가 존재한다고 전제한다. 이러한 전제들 하에, LDA 기법은 문서 내 주제구성 비율을 결정짓는 확률분포를 추정하며, 그 과정에서 전체 텍스트 자료를 주도하는 주제들과 그 주제들을 구성하는 주요 단어 집합들을 찾아낸다. LDA 기법의 상세 작동원리와 통계 및 컴퓨팅 측면에 대해서는 Blei(2012)의 논문을 참조할 수 있다.

<그림 3-13> LDA 토픽 모델링 알고리즘의 문서-주제 모델



\*출처: Blei(2012)의 <Figure 1>(p.78) 발췌

<그림 3-14> 토픽모델링 기반 여론 내용 분석의 원리

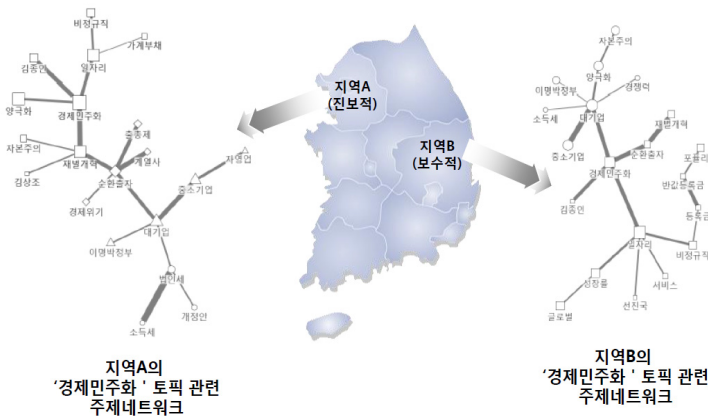


토픽 모델링 기반의 여론 트렌드 시공간분석의 원리는 <그림 3-14>와 같다. 사전에 정의되었거나(행정구역 등), 특정 분석으로 도출된(Voronoi Tessellation 등) 지역구분체계를 이용하여, 각 지역에서 작성된 메시지들을 추출한다. 지역별 메시지에 LDA 토픽모델링을 적용하여, 주요 토픽과 이를 구성하는 키워드들을

도출한다. LDA 토픽모델링의 경우 빈번하게 동시 언급되는 키워드 그룹을 추출하여 여론 토픽의 존재는 보여주지만, 각 토픽의 이름은 분석가가 키워드를 기반으로 할당해야 한다. 이 때 키워드 구성이 조금 상이하더라도 동일 토픽 이름을 할당할 수 있는데, 그 이유는 동일 토픽이라도 다른 키워드들을 통해 표현될 수 있기 때문이다.

토픽 이름이 결정되면, 지역메시지 각각이 특정 토픽에 연관될 확률을 행렬 형태로 계산하고, 전체 지역메시지가 각 토픽에 연관될 평균 확률을 계산한다. 이 결과 값을 이용하여, 지역별 여론의 토픽분포를 <그림 3-12>와 유사한 차트지도의 형태로 표현할 수 있다. 또한, 동일 토픽에 대하여 지역별 키워드 구성 네트워크를 생성하여, 같은 토픽에 대한 여론이라도 그 상세 내용은 지역마다 어떤 차이를 보이는지 분석할 수 있다. 진보적 성향의 지역 A와 보수적 성향의 지역 B가 있고, 두 지역 주민들이 모두 ‘경제민주화’라는 토픽에 대해 관심을 보였다고 가정해보자. <그림 3-15>는 ‘경제민주화’라는 공통 토픽에 대한 상세 논의내용을 주제 네트워크로 표현한 것이다. 비록 여론의 주요 토픽이 동일할지라도, 진보적인 지역 A에서는 경제민주화의 효과(재벌개혁 등)에 대한 관심이 높은 반면, 보수적인 지역 B에서는 경제민주화에 대한 우려(대기업, 경쟁력 등)를 보다 많이 표출하였다.

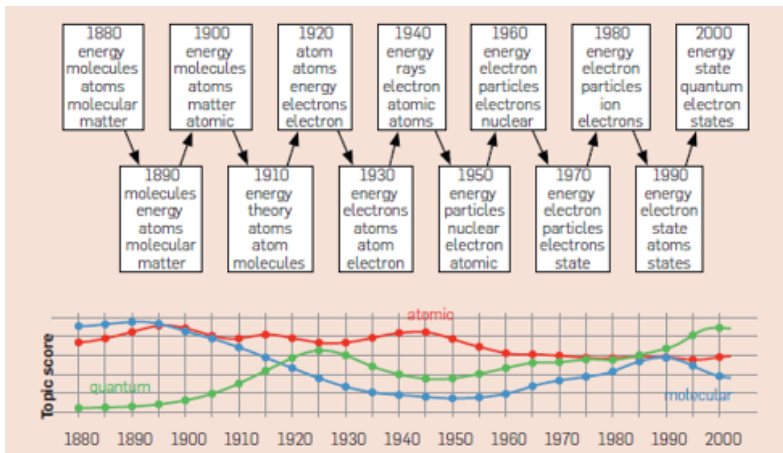
<그림 3-15> 여론 토픽 내용구조의 지역별 차이



\*출처: 강범일 외(2013)의 <그림 1>(p.324) 재구성

토픽 모델링의 주된 장점은 방대한 여론자료를 대표 토픽들로 요약할 수 있다는 것이다. 이에 더해, 주요 토픽의 시간적 변화상을 지속적으로 분석할 수 있다는 점에서 토픽 모델링은 여론 모니터링에 유용하다. <그림 3-16>은 1880년부터 2002년까지 수집된 Science 잡지 기사에 토픽 모델링을 적용한 결과이다. 이 그림에서 볼 수 있듯이, 토픽 모델링을 활용하면 여론을 구성하는 주제들과 이의 세부 키워드 구성이 시간의 흐름에 따라 어떻게 변화하는지, 그 추이를 보다 효과적으로 모니터링 할 수 있다.

<그림 3-16> 토픽모델링을 활용한 Science 논의내용의 경향분석



\*출처: Blei(2012)의 <Figure 5>(p.81) 중 일부 발췌

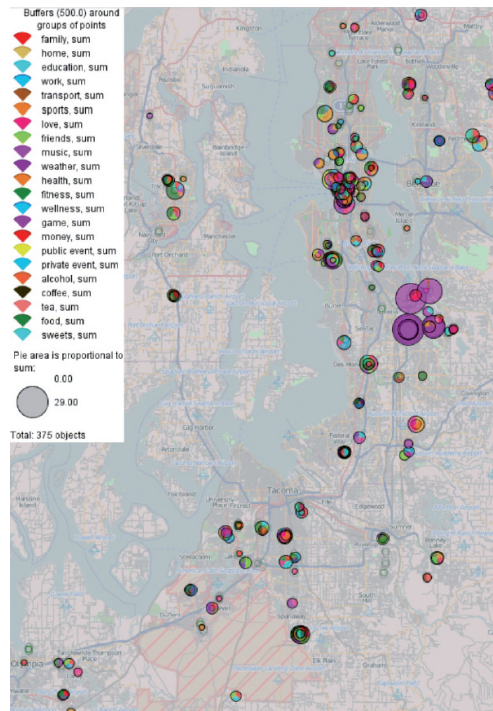
#### (4) 메시지의 시공간 클러스터 분석

어디에서 무슨 주제들의 여론이 주로 형성되는가라는 기본 의문이 만족되면, 여론 형성 활동이 특히 활발하거나 저조한 지역은 어디이고, 이러한 예외적 현상이 벌어진 시기는 언제인지 등의 추가 질문을 던질 수 있다. 예를 들어, 어린이 보호구역 관련 정책수요가 높은 지역을 찾고 싶을 때, 최근 어린이 보호구역에 대한 논의가 예외적으로 활발했던 시공간 클러스터를 찾아 클러스터별 상세 토픽을 분석할 수 있겠다.

일정 기준 이상의 형성 활동이 특정 시기에만 집중되어 분석가가 특별히 관심을 쏠아야 하는 예외상황을 포착하기 위해서, 시공간 클러스터 분석을 여론 관련

메시지에 적용할 수 있다. <그림 3-6>의 기초 분석 자료를 보면, 모든 메시지는 시공간 좌표를 가진다. 이들 3차원 좌표를 이용하여 개별 메시지 간의 시공간적 거리를 계산할 수 있다. 다음으로, 최소 메시지 건수, 메시지 간 최대 공간거리, 메시지 간 최대 시간거리 등의 기준을 적용하여, 메시지들의 시공간 클러스터를 찾아낼 수 있다. <그림 3-17>은 미국 시애틀에서 생성된 트윗 메시지들의 시공간 클러스터를 보여주는데, 이 때 적용된 클러스터 기준은 최소 10건의 메시지, 최대 500m의 메시지 간 공간거리, 최대 15분의 메시지 간 시간거리였다. <그림 3-17>에서 시공간 클러스터들은 그 공간 범위에 따라 크기가 달라지며, 공간 범위가 유사하나 시간 범위가 다를 경우 중첩된 원들로 표현되었다. 또한, 각 클러스터 내에 포함된 메시지들에 여론 토픽 분석을 적용하여, 클러스터별 여론 토픽 구성을 차트의 형태로 표현하였다.

<그림 3-17> 여론 메시지의 시공간 클러스터 분석 결과



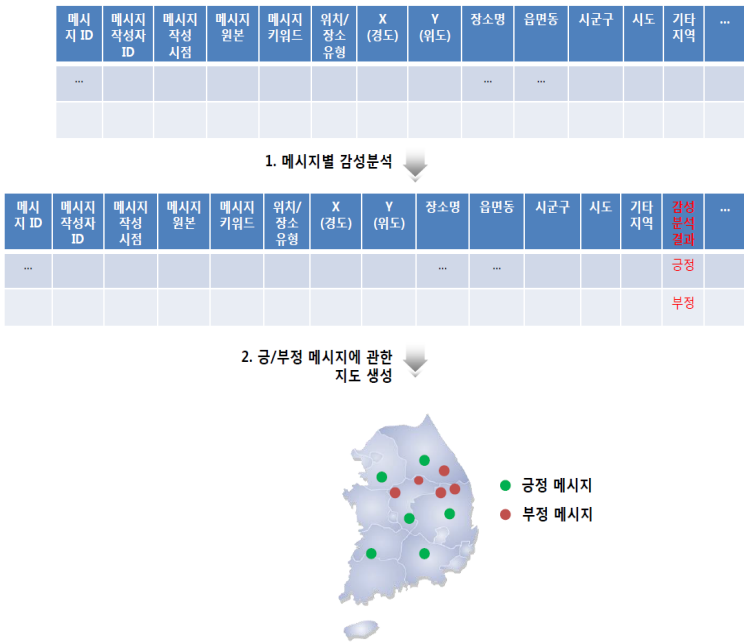
\*출처: Andrienko 외(2013)의 <Figure 6>(p.8) 발췌

(5) 메시지 감성의 시공간 분석

본 연구의 제2장에서 논의한 것처럼, 특정 이슈에 대한 사람들의 태도나 성향을 파악하고자 할 때 감성분석을 주로 이용한다. 대체로 감성분석은 여론 전반에 반영된 사람들의 정서(긍정/부정)를 살펴볼 때 활용된다. 하지만, 시공간적 맥락에서 여론에 반영된 사람들의 감성은 어떤 시공간 분포를 보이는지, 특정 지역/장소에 관한 사람들의 평판은 어떠한지도 분석할 수 있다.

여론에 반영된 감성의 공간 분포를 분석하는 가장 단순한 접근법은 위치참조된 메시지들을 지도 상에 점의 형태로 뿌리고, 이들의 색상을 메시지의 긍/부정 점수에 따라 다르게 표현하는 것이다(<그림 3-18>).

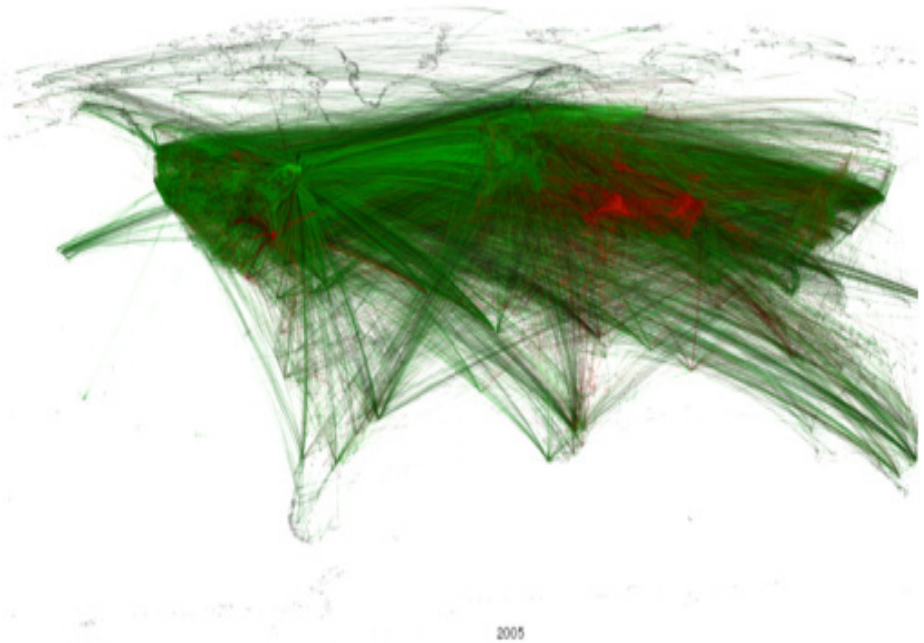
<그림 3-18> 여론 감성 공간분포 분석의 원리



여론 메시지의 긍/부정 점수를 활용하여, 메시지에 언급된 도시나 장소들에 관한 사람들의 언급 태도나 성향도 분석할 수 있다. 먼저, 수집된 여론 데이터에 포함된

모든 장소/도시를 추출한다. 다음으로, 각 장소/도시를 포함한 모든 메시지들의 평균 감성 점수를 계산하여 이들 장소/도시에 할당한다. 세 번째 단계로, 두 개 이상의 장소/도시를 동시 언급하고 있는 메시지들을 추출하고, 함께 언급된 장소/도시를 선으로 연결한다. 네 번째로, 각 선의 끝점에 해당하는 두 도시/장소들을 동시 포함한 메시지들을 추출하고, 이들의 평균 감성 점수를 선에 할당한다. 마지막으로, 주요 장소/도시 및 이들을 잇는 선을 네트워크 지도 형태로 표출한다. <그림 3-19>는 2005년 뉴욕타임즈에서 언급된 도시와 랜드마크에 관한 감성분석로, 초록색 지역은 그 지역 관련 긍정적 논의가 많으며 빨간색 지역은 부정적 논의가 많음을 뜻한다. 여론 데이터에 포함된 장소들에 대해 감성분석을 적용할 경우, 유사한 분석결과를 산출할 수 있다.

<그림 3-19> 여론 감성 공간분포 분석의 예시



\*출처: Leetaru(2012)의 <Figure 12> 발췌

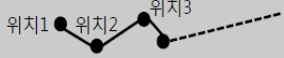
## 2) 여론 형성주체에 관한 시공간분석

### (1) 메시지 작성자의 공간분포 분석

인터넷 카페, 트위터, 페이스북 등의 온라인 매체를 이용하여 여론 형성을 주도하는 사람들이 실제 현실공간에서는 어디에 위치할까? 온라인이라는 가상의 공간에서 활동하므로 여론 형성주체들의 실제 지리적 분포에는 아무런 경향성이 없을까? 아니면, 실제 지리적으로 가까이 사는 사람들이 온라인 매체에서도 활발히 의견을 주고받는 것일까? 여론 형성주체에 관한 이러한 질문들도 여론 모니터링에 있어 중요한 역할을 하는데, 이는 온라인 매체 상의 여론 형성과정에 대한 이해를 높일 수 있을 뿐만 아니라, 온라인 매체와 실제 지리공간 상에서의 여론 형성이 어떻게 연관되어 있는지를 파악할 수 있도록 도와주기 때문이다.

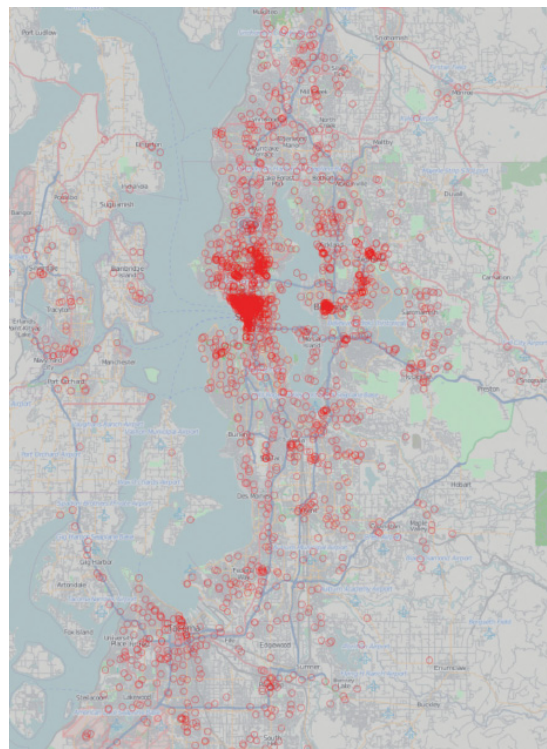
여론 형성주체들의 공간분포를 파악하기 위해서는, 수집된 여론 데이터를 작성한 사람들의 활동 위치나 장소와 관련된 기초 분석 자료를 먼저 구축해야 한다. 트위터, 페이스북 등의 SNS 서비스에서는 사용자가 서비스를 가입할 때 사용자 주소나 거주/활동지를 기입하도록 하므로, 이들 정보를 활용하여 메시지 작성자의 위치/장소를 파악할 수 있다. 또한, 사용자가 작성한 메시지들의 작성 위치도 사용자가 활동하는 공간적 범위를 보여줄 수 있는데, 이 경우 하나의 위치나 장소가 아닌 이동경로와 같은 일련의 위치 정보들을 수집해야 한다. 여론 형성주체인 메시지 작성자들의 공간분포 분석을 위한 기초 자료는 <그림 3-20>과 같은 구조를 가진다.

<그림 3-20> 메시지 작성자의 공간분포 분석을 위한 기초자료

| 메시지 작성자 ID | 기본 주소<br>(서비스 등록시 기재) | 서비스<br>등록 시점 | 메시지 작성 위치                                                                                                     |
|------------|-----------------------|--------------|---------------------------------------------------------------------------------------------------------------|
| 홍길동        | 경기도 안양시 동안구           | 2014/12/31   | { 위치1, 위치2, ..., 위치N }<br> |
| ...        |                       |              |                                                                                                               |

메시자 작성자의 위치 자료가 구축되면, 이들 위치의 공간분포를 분석할 수 있다. 메시지 작성위치에 대한 공간 분석의 경우와 마찬가지로, 위치도(<그림 3-7>, 단계구분도(<그림 3-7>), 밀도분석(<그림 3-8>)과 같은 기법을 이용하여 메시지 작성자들의 공간분포 패턴을 탐색할 수 있다. 또한, 사용자별 메시지 작성위치들의 대표점을 추출하여 그 분포의 공간적 변이를 살펴볼 수 있으며, 메시지 작성위치들을 연결하는 이동경로들의 밀도(Line Density)를 분석하여 여론 형성 활동이 어디에서 보다 활발한지 살펴볼 수 있다. <그림 3-21>은 사용자별 대표적 메시지 작성위치를 보여준 사례로, 미국 시애틀에서 트위터 사용자들이 트윗을 작성한 위치를 기반으로 사용자별 이동경로를 추출하고, 이 이동경로의 대표점(Medoid)을 원으로 표시하였다(Andrinko 외, 2013).

<그림 3-21> 사용자별 메시지 작성 대표위치 분석의 예시



\*출처: Andrienko 외(2013)의 <Figure 9>(p.10) 발췌

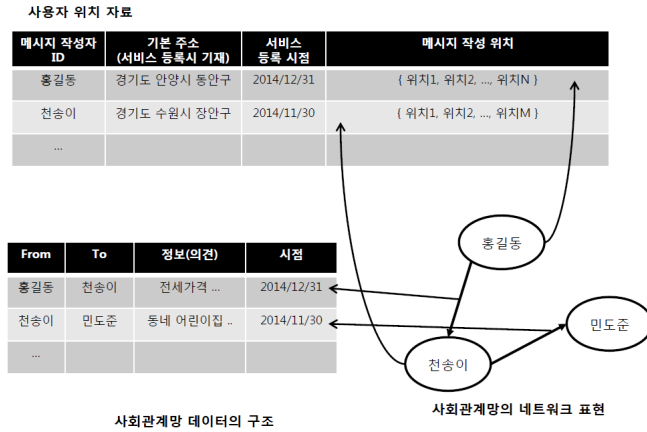
## (2) 사회관계망의 공간분포 분석

인터넷 카페, 트위터, 페이스북 등 신규매체에서 여론은 매체 이용자들 간의 상호작용을 통해 형성된다. 면대면 소통을 통해 의견이 조성되는 오프라인의 여론 공간과 달리, 온라인 공간에서는 친구맺기, 댓글, 팔로잉, 리트윗 등의 특수한 행위를 통해 사람들 간의 소통이 이루어진다. 온라인 매체에서 어떤 사람의 친구는 누구인지, 이 사람의 메시지에 누가 댓글을 달았는지, 이 사람의 메시지들을 리트윗하는 사람들은 누구인지 등의 정보를 추출하여 매체 이용자들 간의 사회관계망을 추출·분석할 수 있다.

2장에서 설명한 것처럼, 여론 데이터에 숨어있는 사회관계망을 분석하는 기존의 방법들은 대부분 사회관계망의 공간적 측면을 고려하지 못했다. 그러나, 신규매체에서 특정 이슈에 관한 논의를 주도하는 사람들이 실제 공간에서 어디에 위치하고 있고, 이들의 사회적 관계가 어떤 공간적 범위에서 형성되고 있는지를 살펴볼 수 있다면 여론의 성격(지역적/전국적), 잠정적 여론확산 경로 등의 파악에 유용할 수 있다. 온라인 매체 이용자의 위치/장소 정보를 추출하고, 의견교환이 활발한 이용자 그룹들의 상호작용을 네트워크 형태로 지도상에 표출하여 이의 변화를 모니터링할 수 있다면, 온라인 사회관계망의 시공간적 특성을 보다 쉽게 이해할 수 있다.

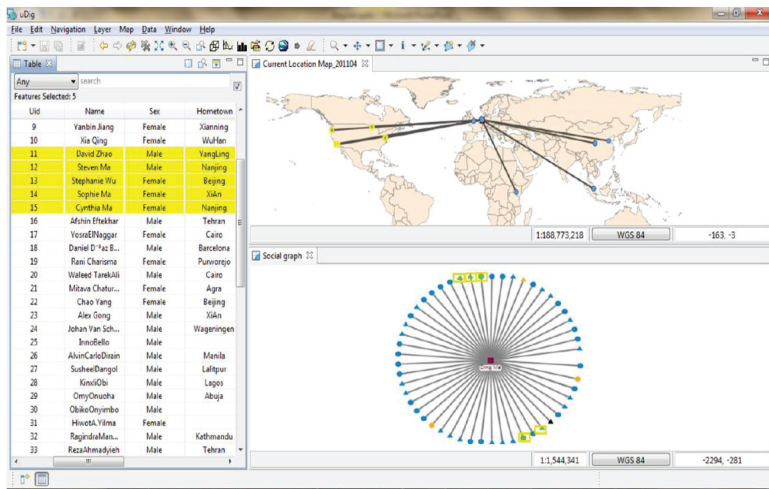
사회관계망의 시공간분석을 위해서는 여론 관련 메시지 작성자들의 위치 자료(<그림 3-20>)와 함께, 어떤 사람이 누구와 어떤 의견교환을 언제 했는지에 대한 정보가 필요하다. 개념적으로 이는 노드(Node)와 링크(Link)로 구성된 네트워크 혹은 그래프로 표현될 수 있다. 데이터 구조 측면에서는 노드(사람)의 쌍들이 교환한 정보(의견)와 시점을 나열한 표로 표현될 수 있다. 사회관계망의 공간적 분포를 분석하기 위해서는, 노드-링크의 사회관계망 자료가 사용자 위치 자료와 연계되어야 하며, 이는 <그림 3-22>에서 도식으로 표현하고 있다.

<그림 3-22> 사용자 위치자료와 사회관계망 자료의 연계



<그림 3-22>와 같은 공간정보화된 사회관계망 자료가 구축되면, 사회관계망의 지리적 위치를 보여주는 다양한 분석을 실시할 수 있다. 우선, 상호작용을 한 사용자들의 위치를 연결한 지리적 네트워크 형태로 분석할 수 있으며, 이때 사용자의 기준 위치는 기본 주소, 메시지 작성 대표지점 등 다양하게 변경할 수 있다. <그림 3-23>은 사회관계망의 지리적 네트워크를 분석하는 예시로, Ding Ma라는 이용자의 트위터 친구들의 공간분포를 사회관계망과 함께 보여준다.

<그림 3-23> 사회관계망의 지리적 네트워크 분석 예시



\* 출처: Ma(2013)의 <Figure 5-15>(p.44) 발췌

여론 데이터 기반의 사회관계망 분석에 있어, 논의 내용을 토대로 관심 이슈가 유사한 소그룹들을 추출하고 이들의 공간적 범위가 어떻게 다른지 살펴볼 있다. 예를 들어, 구제역 확산과 롯데월드 싱크홀 관련 이슈를 논의한 메시지들을 수집하고, 이들 메시지의 작성자들이 어디에 어떻게 분포하는지를 비교할 수 있겠다. <그림 3-24>는 사용자 그룹별 공간분포를 비교·분석하는 가상의 예이다. 구제역 확산과 같은 문제에 대해서는 메시지 작성자의 분포가 <그림 3-24>의 Community 4와 같이 광범위하게, 롯데월드 싱크홀과 같은 문제에 대해서는 메시지 작성자의 분포가 <그림 3-24>의 Community 124와 같이 국지적으로 나타날 수도 있다.

<그림 3-24> 사회관계망 내 소그룹들의 공간분포 비교·분석의 예시



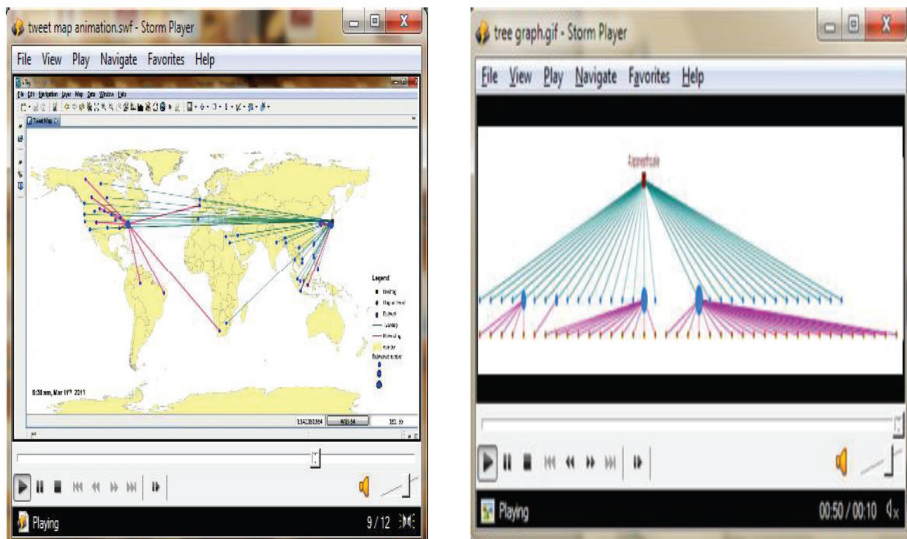
\*출처: Comber 외(2012)의 <Figure 4>(p.295) 발췌

### (3) 메시지 전파의 시공간 분석

여론 데이터 내에 숨어있는 사회관계망은 사람들 간의 소통과 상호작용이 어떻게 이루어지는지를 보여주기 때문에, 특정 이슈가 사람들 사이에서 어떻게 전파되는지를 알고자 할 때 유용하다. 공간정보화된 사회관계망의 경우 사람들 간의 상호작용을 현실 공간으로 투영시킬 수 있는데, 이를 분석하면 특정 메시지나 이슈가 사람을 매개로 어떻게 지리적으로 확산되는지 알 수 있다. 구체적으로,

트위터라는 서비스는 리트윗이라는 과정을 통해 1차 생성된 메시지를 다른 서비스 이용자에서 2차 전파할 수 있는 기능을 제공한다. 따라서, A라는 사람이 작성한 ‘구제역 발생’이라는 트윗 메시지는 1차로 A의 팔로어인 B에게 전달된 후, 리트윗을 통해 B의 팔로어인 C에게까지 전달될 수 있다. A-B, B-C 간의 메시지 전달을 1차, 2차 확산으로 구분하고 이를 지도 상에 표출하면, ‘구제역 발생’이라는 이슈가 A, B, C 세 사람을 통해 A가 사는 경기도 안성에서 B가 사는 경기도 안양을 통해 C가 사는 서울로 확산되는 것을 알 수 있다. <그림 3-25>는 일본지진에 관한 트윗 메시지가 사회관계망을 통해 지리적으로 확산되는 과정을 일관 그래프, 네트워크 지도, 애니메이션을 이용하여 보여주는 분석기법의 예시이다.

<그림 3-25> 사회관계망 기반 메시지 확산 분석의 예시



\* 파란색과 분홍색 선은 각각 1차, 2차(리트윗) 전파를 나타냄.

\* 출처: Ma(2013)의 <Figure 5-13>(p.43) 발췌

chapter IV

결론 및 향후과제



## 결론 및 향후 과제

이 장에서는 지금까지 수행한 연구결과를 요약하고 정책제언을 제시하였다. 본 연구의 한계와 함께, 비정형 빅데이터 시공간분석 모델의 실증과 적용을 위한 향후 연구과제도 제안하였다. 인터넷 소셜미디어에서 생산되는 비정형 빅데이터에서 여론의 시공간적 변화 트렌드를 탐지하는 분석모델은 여론의 지역적, 시간적 변화 동태를 모니터링할 수 있는 방법론을 제공하여 국민의 정책수요에 시의성있게 대응하기 위한 정책수단으로 활용될 수 있을 것으로 기대된다.

### 1. 연구의 결과 및 정책제언

#### 1) 연구의 결과와 의의

본 연구는 국민들의 새로운 의사소통 창구로 떠오른 인터넷과 소셜미디어에서 생성되는 비정형 빅데이터를 시공간적으로 분석하여 여론의 지역적, 시간적 변화 동태를 파악할 수 있는 방법론을 제안하였다. 이 방법론은 기존의 여론 모니터링 방법에서 찾아내기 어려웠던 여론 내용의 지역적, 시기적 차이, 여론 형성주체들의 시공간적 상호작용을 지도와 차트 등으로 시각화할 수 있는 새로운 분석기법을 제공하였다는 데 그 학술적 의의가 있다. 또한, 여론 관련 비정형 빅데이터의 수집, 공간정보화, 정제, 시공간분석 등 분석 체계를 정립하여, 후속 연구와 사업들을 위한 방법론적 가이드를 제공하였다.

정책적 측면에서, 본 연구는 국민여론 기반의 맞춤형 정책수립에 적용가능한 비정형 빅데이터 시공간 분석방법을 개발하여 중앙 및 지방정부가 여론 모니터링시 활용할 수 있는 정책수단을 제공했다는 데 의의가 있다. 국가의 정책기조가 국민 행복과 소통을 지향하면서, 지역과 국민의 수요에 근거한 맞춤형 정책 발굴이 그 어느 때보다도 중요하다. 인터넷 소셜미디어에서 형성된 여론이 국민의 정책수요를 보여줄 수 있는 유용한 정보원으로 인식되고 있으나, 이들 매체에서 쏟아지는 방대한 양의 데이터에서 지역과 국민의 정책수요를 찾아내는 일은 쉽지 않다. 본 연구에서 제안한 비정형 빅데이터 시공간분석 모델은 인터넷 소셜미디어 상의 국민 여론이 지역과 시기에 따라 변화하는 양상을 포착할 수 있는 시각적 분석방법론을 제공하여, 비정형 빅데이터의 정책활용 가능성을 제고할 수 있다.

## 2) 정책 제언

본 연구의 결과를 토대로 국민수요와 지역 맞춤형 정책 구현을 위한 세 가지 정책제언을 할 수 있다.

먼저, 본 연구에서 제안한 비정형 빅데이터 시공간분석 모델을 활용하여 정책인프라 성격의 여론 모니터링 체계를 구축할 필요가 있다. 근래에 비정형 빅데이터 기반 여론 분석을 통해 국민수요 및 지역특성에 맞춤형 정책 발굴하려는 정부/공공기관이 증가하고 있다. 기관마다 초점을 두고 있는 정책은 다르지만, 비정형 빅데이터를 수집하고 분석하는 과정과 방법론은 크게 다르지 않다. 그럼에도 불구하고, 현재는 기관마다 별도로 비정형 데이터의 수집과 분석을 위한 서비스를 구매하거나 구축하고 있다. 이러한 기관별 산발적 접근법은 비정형 빅데이터 분석을 위한 유사 기능들을 구축하는데 중복투자를 유발할 수 있어 예산의 낭비가 우려된다. 또한, 비정형 빅데이터 분석 서비스의 구매나 구축에 적지 않은 비용이 필요하므로, 예산을 충분히 확보하지 못한 기관들의 경우 비정형 빅데이터의 활용을 고려조차 할 수 없다. 따라서, 여러 공공/정부기관들이 함께 사용할 수 있는 정책인프라 성격의 여론 모니터링 체계를 구축·공유하여, 데이터와 여론 기반의 정책수요 발굴을 위한 공공부문의 정책역량을 강화할 필요가 있다.

두 번째로, 비정형 빅데이터 시공간분석 모델을 「정책지도」, 「공간 빅데이터 체계 구축」 등의 정책사업과 연계할 필요가 있다. 비정형 빅데이터 시공간분석 모델은 인터넷 소셜미디어 상에서 형성되는 국민여론의 시공간적 변화패턴을 시각적으로 분석할 수 있는 방법을 제공한다. 최근 광주광역시 광산구, 서울시 등을 중심으로 추진된 정책지도 사업은 정책관련 사회현상을 지도로 표출함으로써, 정책과정의 여러 단계를 지원하고자 한다. 비정형 빅데이터 시공간분석 모델은 그간 정책지도에서 많이 다루지 않았던 국민여론의 지도 표출을 지원하므로, 정책지도의 한 유형으로 발전할 수 있는 가능성이 크다.

또한, 국토교통부에서 현재 중인 「공간 빅데이터 체계 구축 사업」에서는 비정형 빅데이터를 공간화한 다종의 융합데이터를 생산할 예정이다. 따라서 본 연구에서 제안한 비정형 빅데이터 시공간분석 모델과 연계하면, 국토교통부에서 신규로 생산할 융합데이터의 활용도를 제고하는데 기여할 수 있다.

마지막으로, 비정형 빅데이터 시공간분석 모델의 정책활용을 활성화하기 위해서 제도 기반 마련 및 교육·홍보의 강화가 필요하다. 현재 여론 분석 및 모니터링은 중앙부처와 지방자치단체 내에서도 정책기획과, 국민소통과, 주민참여과 등 여러 부서에서 산발적으로 추진되고 있다. 이러한 산발적 접근법은 여론 모니터링에 대한 예산 중복투자를 가져올 수 있을 뿐만 아니라, 투자 대비 효용 가치가 낮은 모니터링 성과품을 산출하는 결과를 초래할 수 있다. 따라서, 관련 조직간 협력적 거버넌스 체계와 공동의 정책수단을 구축하고, 정책관련 여론 분석을 위한 비정형 빅데이터 활용에 관한 업무지침을 수립하여 비정형 빅데이터 시공간분석 모델의 정책활용을 위한 제도적 기반을 마련해야 한다.

비정형 빅데이터 시공간분석 모델의 정책활용 활성화를 위해서는 이 모델의 기본 원리, 적용 방법, 효과 등에 대해 체계적 교육과 홍보도 필요하다. 특히, 맞춤형 정책지원 관련 비정형 빅데이터 시공간분석 모델의 활용사례를 적극 개발·공유하여 여론 기반 정책수요 발굴을 위한 첨단 분석기법을 정책수립자, 연구자, 일반국민 등 다양한 수요자층에 확산할 필요가 있다.

## 2. 연구의 한계와 향후 과제

### 1) 연구의 한계

본 연구에서는 여론 모니터링을 보완할 수 있는 분석방법론의 하나로 비정형 빅데이터 시공간분석 모델을 제안하였다. 2개월의 짧은 연구기간으로 인해 본 연구는 연구 내용과 범위의 측면에서 몇 가지 한계를 가진다.

먼저, 연구의 내용 측면에서는 비정형 빅데이터의 정책활용시 고려해야 할 데이터의 정책적 유의성을 보다 심도있게 다루지 못하였다. 비정형 빅데이터가 인터넷 소셜미디어에서 주로 생산되고 이들 신규매체의 이용이 젊은 계층을 중심으로 널리 확산되고 있으므로, 여론 모니터링 및 여론기반 정책수요 발굴시 비정형 빅데이터를 활용하는 것은 타당한 접근법이다. 그러나, 신규매체의 이용계층이 제한적이고 이들을 통해 수집된 여론 관련 메시지들에 포함된 위치정보의 의미 또한 여러 가지로 해석될 수 있기 때문에, 비정형 빅데이터가 보다 유용하게 활용될 수 있는 정책과 그렇지 못한 정책들을 구분하여 데이터의 분석 및 해석 방안에 대해 보다 명확히 논의할 필요가 있다.

다음으로, 연구 범위의 측면에서 본 연구는 비정형 빅데이터 시공간분석 모델을 개념적으로 정립하였지만, 실제 데이터를 활용한 분석모델의 실증 및 구체적 적용사례까지는 발굴하지 못하였다. 비정형 빅데이터 시공간분석 모델의 정책현장 활용을 위해, 이후 계속해서 자료수집, 분석방법론, 정책적용 방안 등에 대한 검토와 후속 연구가 필요하다.

### 2) 향후 과제

본 연구에서 제안한 비정형 빅데이터 시공간분석 모델의 지속적인 발전을 위해 향후 추진해야 할 과제는 다음과 같다. 먼저, 분석모델의 실증과 적용을 위한 구체적인 정책연구 사례를 발굴해야 한다. 이를 위해 앞서 언급한 「정책지도」, 「공간 빅데이터 체계 구축」 등의 국가사업과 연계한 융합연구를 추진할 필요가 있다.

다음으로 기술 및 방법론적 측면에서 비정형 빅데이터 시공간분석 모델을 보다 정교화할 필요가 있다. 비정형 빅데이터의 시공간분석을 통해 유용한 정보를 추출하기 위해서는 메시지, 문서 등의 비정형 데이터를 공간정보화하는 과정이 매우 중요하다. 비정형 빅데이터가 그 크기가 방대함에도 불구하고, 정확한 지리좌표를 포함한 데이터는 그 비율이 높지 않다. 따라서, 비정형 데이터를 공간정보로 변환할 수 있는 다양한 기술들을 개발하고 이들의 유용성을 검증하는 과정이 필요하다. 이에 더하여, 비정형 빅데이터의 유형과 수집원에 따라 적용가능한 공간정보화 방법이 무엇이며 그 실효성은 얼마나 되는지 평가하여, 여론 모니터링을 위한 비정형 빅데이터 활용의 실제적 가이드라인을 제시할 필요가 있다.

방법론적으로는 여론의 내용과 형성주체의 분석에 중점을 둔 비정형 빅데이터 시공간분석 모델을 국토공간 관련 정형 데이터와 연계·활용할 수 있는 방안을 마련해야 한다. 정책관계자들이 여론 모니터링을 통해 잠재적 정책수요를 탐지한 후에는, 여론 상의 정책수요가 실제 국토공간 현황과 어떻게 맞물려 있는가라는 의문을 제기할 것이다. 이런 의문에 체계적으로 접근할 수 있는 분석모델을 개발하여, 비정형 빅데이터 시공간분석 모델의 실효성을 보다 강화할 필요가 있다.



## 참고문헌

- 강범일·송민·조화순. 2013. 토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구. 한국문헌정보학회지 제47권 제4호 315~334.
- 고영삼. 2011. 미디어 패러다임 변화에 따른 정부의 소셜미디어 커뮤니케이션 방향. IT & Future Society 제13호. 한국정보화진흥원.
- 광운대학교 산학협력단. 2014. 국민 중심의 소통 활성화 방안 연구. 광운대학교 산학협력단.
- 김규원·신형진·권기욱. 2013. 대구사회조사를 활용한 지역정책 개발방안 연구. 대북 경북연구원.
- 김남두. 2010. 미디어 융합환경에서의 여론형성 재구조화에 관한 연구(Ⅱ): 6.2 지방선거기간 인터넷 이용자의 뉴스 이용행태에 대한 사례 분석. 정보통신정책연구원.
- 김대중·윤서연. 2013. 빅데이터를 활용한 국토정책반응 모니터링 및 정책수요 예측방안. 국토연구원.
- 김미정·김대중·이영주. 2013. 과학적 국토정책을 위한 공간 빅데이터 활용방안. 국토연구원.
- 김원근·김춘식. 2010. 한국의 인터넷뉴스서비스. 한국언론진흥재단.
- 김춘식. 2013a. 공공갈등 해결의 대안적 기제로서 공론조사. 한국지방정부학회 학술 발표논문집 2013권 1호 pp. 480-493. 한국지방정부학회.
- 김춘식. 2013b. 숙의(熟議) 여론조사 기법으로서 『시민참여 여론조사』와 2가지 실행 사례([www.hrc.co.kr/renewal/newsletter/data/131203.pdf](http://www.hrc.co.kr/renewal/newsletter/data/131203.pdf)).

- 김형호·송경재. 2011. 한국적 타운미팅의 가능성과 한계에 관한 연구: 평택 채소마을 사례를 중심으로. 지방행정연구 제25권 제3호 통권 86호 pp.269-288.
- 더아이엠씨. 2014. 도홈페이지 및 포털사이트 빅데이터 분석을 통한 『정부3.0 도민 맞춤형 계획 수립』 용역 최종보고서.
- 라도삼. 2014. 2030 서울 문화플랜 수립을 위한 기본 구상. 서울연구원.
- 모효정·권복규. 2009. 사회여론 수렴방법으로서의 시민 합의회의: 이종이식 문제를 중심으로. 생명윤리정책연구 제3권 제1호 pp.91-109.
- 문신용. 2006. e-거버넌스의 활성화를 위한 과제와 발전 방향 - 시민참여를 중심으로. 한국행정연구원.
- 문화체육관광부. 2013. 빅데이터 분석을 통해 본 2013 국민 인식 변화. 문화체육관광부.
- 미래창조과학부·한국정보화진흥원·빅데이터전략센터. 2014. 빅데이터 활용 단계별 업무절차 및 기술 활용 매뉴얼(Version 1.0). 미래창조과학부·한국정보화진흥원.
- 박상호. 2012. SNS의 여론형성과정과 참여행태에 관한 고찰. 한국언론정보학보 통권 58호 pp.55-73.
- 박한우. 2012. ‘저비용 고효율’ SNS로 여론 읽기. 미디어포럼.
- 변미리. 2012. 빅데이터를 활용한 시정 모니터링 활용방안(내부자료).
- 서울시정개발연구원. 2012. 서울시민 100인, 2030년 미래서울을 말하다 시민원탁토론 최종보고서(내부자료).
- 서원석·조세현·송희준. 2012. 정부신뢰와 소통제고를 위한 Public Relations 시스템 구축사업: 소셜 미디어를 이용한 정부 PR 체계 구축 방안. 한국행정연구원.
- 솔트룩스. 2008. 텍스트 마이닝홈: 텍스트 마이닝의 정의. 솔트룩스.
- 신수정. 2014. “글에서 감정을 읽다” 감성 분석의 이해. IDG Tech Report.
- 심재용. 2003. 공론조사의 소개 및 그 실행의 장단점과 문제점. KOSOMAR 세미나자료집.
- 심재용. 2013. 정책과정과 공론조사의 활용. KIPA 조사포럼 vol.4. 한국행정연구원 사회조사센터.
- 안전행정부. 2013. 국민여론조사 및 정책수요 조사결과 보고서.
- 울산광역시 북구. 2013. 울산 북구 주민참여정책 추진성과 분석 및 제도화 방안.

- 이석주·연지윤·천승훈. 2013. 빅데이터를 이용한 교통정책 개발 및 활용성 증대방안. 한국교통연구원.
- 이영주·임은선·차미숙·이재용·이호상·박재희·이지원. 2013. 공간 빅데이터 기반의 복지·공간 융합모델 시범연구: 영유아 복지공간 진단을 통한 복지정책 지원 방안. 국토교통부.
- 이원태·차재권·신호철. 2012. 스마트 미디어 환경에서 SNS이용과 정책참여 활성화 방안 연구. 방송통신위원회.
- 이윤희. 2014. 국내 SNS의 이용 현황과 주요 이슈분석. Focus 4.
- 이혜정. 2007. 미래 유비쿼터스 정부의 진화, Government 3.0. 유비쿼터사회연구시리즈 제29호. 한국정보사회진흥원.
- 전채남. 2011. SNS를 활용한 지방자치단체 커뮤니케이션 방안에 관한 연구: 대구시의 SNS 활용을 중심으로.
- 조성은·박지영·박한우. 2011. 웹2.0 시대 소셜 미디어를 이용한 정책 홍보 전략: 농림수산물식품부의 트위터 활동을 중심으로.
- 진설아·허고은·정유경·송민. 2013. 트위터 데이터를 이용한 네트워크 기반 토픽 변화 추적 연구. 정보관리학회지 제30권 제1호 pp.285-302.
- 최광선. 2012. 비정형 Big-Data 활용을 위한 아키텍처 및 시스템 환경 구축 사례.
- 최선화·배병걸. 2013. 소셜 빅데이터 재난관리 운영방안 및 이슈탐지 기법연구. 국립재난안전연구원.
- 최선화·배병걸·다음소프트. 2013. 비정형 빅데이터 수집 및 분석 모델 개발. 국립재난안전연구원.
- 최진명. 2012. 빅데이터 시대의 데이터 활용과 전략, 지역정보화 동향분석 제3호 (2012.4월). 한국지역정보개발원.
- 하수욱·남광우·류근호. 2012. 마이크로 블로그기반의 공간 지식 추출 기법연구. 한국공간정보학회지 제20권 제2호 129~136.
- 한국대통령학연구소. 2008. 외교통상 정책결정에서의 의견수렴과정에 관한 연구.
- 한국전산원. 2004. e-Governance 구현을 위한 과제와 전략.

- 한국정보화진흥원·빅데이터전략연구센터. 2012. 빅데이터 시대! SNS의 진화와 공정정책. IT & Future Strategy 제 13호.
- Addis, M., Taylor, S., Nasser B. I., Joshi, S., Karamagioli, E., Wandhoefer, T., Fallon, F., Fletcher, R., Wilson, C. 2010. New Ways for Policy-makers to Interact with Citizens Through Open Social Network Sites – A Report on Initial Results. Internet, Politics, Policy 2010: An Impact Assessment (IPP2010).
- American Association for Public Opinion Research. 2014. Social Media in Public Opinion Research: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research.
- Andrienko, G., Andrienko, N., Bosch, H., Ertl, T., Fuches, G., Jankowski, P., and Thom, D. 2013. Thematic Patterns in Georeferenced Tweets Through Space-Time Visual Analytics. Computing in Science & Engineering May/June: 2-12.
- Blei, D.M. 2012. Probabilistic Topic Models. Communications of the ACM 55(4): 77-84.
- Chew, C. and Eysenbach, G. 2010. Pandemics in the Age of Twitter: Content Analysis of Tweets During the 2009 H1N1 Outbreak. PLoS One 5(11) e14118.
- Comber, A., Brunsdon, C., Batty, M., Neuhaus, F., and Gray, S. 2012. Geographical Analysis of Social Network Data. Proceedings of the AGILE'2012 International Conference on Geographic Information Science 293-392.
- Commission on Global Governance. 1995. Our Global Neighbor.
- Denhardt, J. V. and Denhardt, R. B. 2011. The New Public Service: Serving, not Steering.
- Donsbach, W. and Traugott, M. W. (Eds.) 2008. The SAGE Handbook of Public Opinion Research. Los Angeles·London·New Delhi·Singapore: SAGE Publications.

- Ghosh, D. and Guha, R. 2013. What Are We ‘Tweeting’ About Obesity? Mapping Tweets with Topic Modeling and Geographic Information System. *Cartography and Geographic Information Science* 40(2): 90–102.
- Goes, P. B. 2014. Big Data and IS Research. *MIS Quarterly* 38(3): iii–viii.
- Kaplan & Haenlein. 2010. Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons* 53(1): 59~68.
- Kavanaugh, A.L., Fox, E.A., Sheetz, S.D., Yang, S., Li, L.T., Shoemaker, D.J., Natsev, A., Xie, L. 2012. Social Media Use by Government: From the Routine to the Critical. *Government Information Quarterly* 29: 480–491.
- Leetaru, K. H. 2011. Culturomics 2.0: Forecasting Large-scale Human Behavior Using Global News Media Tone in Time and Space. *First Monday* 16(9).
- Lukensmeyer, C. J. 2005. PARTICIPATORY METHODS TOOLKIT A practitioner’s manual([www.kbs-frb.be](http://www.kbs-frb.be) or [www.viWTA.be](http://www.viWTA.be)).
- Ma, D. 2013. Visualization of Social Media Data: Mapping Changing Social Networks. Master Thesis. Univ. of Twente.
- Neuhaus, F. 2013. Urban Rhythms: Habitus and Emergent Spatio-temporal Dimensions of the City. PhD Dissertation. UCL.
- OECD. 2001. Citizens as Partners: Information, Consultation, and Participation. Paris.
- Paul, M. J. and Dredze, M. 2011. You Are What You Tweet: Analyzing Twitter for Public Health. Fifth International AAAI Conference on Weblogs and Social Media: 265~272.
- Sobkowicz, P., Kaschesky, M., Bouchard, G. 2012. Opinion Mining in Social Media: Modeling, Simulating, and Forecasting Political Opinions in the Web. *Government Information Quarterly* 29: 470–479.



## SUMMARY

Keywords: Public Opinion Monitoring, Spatiotemporal Analysis,  
Unstructured Big Data

Recently, a lot of efforts have been made to identify policy demands hidden in public opinions created through the Internet social media (ISC). Nonetheless, methodological challenges still remain to differentiate delicate changes in the trends of public opinions across space and time. This study aims to overcome such challenges by developing a new model for spatiotemporal analysis of unstructured big data from ISC and directing their uses in public policies.

Conceptually, our model seeks to combine the applications of text mining, social network analysis, spatiotemporal analysis to unstructured big data from ISC so that it can become a data-centric method for public opinion monitoring. The ultimate goal of this model is to support timely identification of policy demands tailored to the needs of the public and regions by enabling the continuous analysis and monitoring of spatiotemporal patterns in public opinions from ISC.

Methodologically, our model consists of three stages for the collection, preprocessing, and spatiotemporal analysis of unstructured big data. The first stage of our model involves the incessant gathering, storage, and management of messages that include discussions of the public on ISC. Oftentimes, this stage

necessitates the use of open APIs, and free or commercial software agents for the crawling of documents from ISC. The second stage of our model involves three tasks: 1) removal of advertisements, swearwords, and redundant messages; 2) linking messages to their associated locations/places (geocoding or spatialization); and 3) conversion of individual messages into a set of substantive terms. The final stage of our model focuses on exploring spatiotemporal patterns in the contents and creators of public opinions through the mixed applications of spatiotemporal analysis, time series data analysis, text mining, social network analysis, and etc. In specific, the contents of public opinions can be understood by analyzing spatiotemporal variations in the locations of message posting, spatio-temporal clusters of term usage, and topical and emotional trends hidden in a corpora of messages/documents. Meanwhile, the analyses of the locations and social networks of message authors could reveal the spatiotemporal distributions and interactions among the creators of public opinions.

The model proposed in this study contributes to the literature of public opinion research by providing a new analytical framework that can detect and visualize spatiotemporal patterns in the contents of public opinions as well as interactions among the creators of such opinions. For a policy standpoint, the proposed model is expected to develop into a software infrastructure for the shared use by various public institutions for the purpose of public opinion monitoring.

## 부록1. 시민참여형 여론모니터링 방법과 사례

### 1. 사회/여론 조사

#### 1) 개요

사회/여론 조사는 최근 “도시지표”, “도시정책지표”, “삶의 질 지표”, “사회지표” 등의 이름하에 다양한 정부기관에서 실시하고 있다. 오늘날 사회구조가 빠르게 변화하고 있고 이를 반영하여 국가 및 지역의 정책 방향을 설정하고 관련 의사결정을 내리기 위해서는 근거가 될 만한 기초자료들이 필요한데, 사회/여론조사는 이들 자료를 수집·제공할 수 있는 효과적 정책수단이기 때문이다.

사회조사의 대표적인 예는 시민서베이다. 1970~80년대 처음 등장했을 때만 해도 대표성이 낮고 체계적이지 못했으나, 여러 차례의 시행착오 과정을 거쳐 시민들의 수요를 대변하는 주요 자료원의 역할을 하고 있다. 우리나라 최초의 사회조사는 1977년 통계청에서 실시한 「한국의 사회지표」 조사인데, 이 때 조사항목은 사회적 관심사항, 삶의 질 관련 사항 등 시민의 주요 관심사였고 그 결과는 정부의 사회개발정책 수립을 위한 기초자료로 활용되었다. 이후 가족, 복지, 사회참여, 노동, 소득과 소비, 문화와 여가 등 영역으로 조사항목을 확대하고 있으며 2008년에는 사회조사로 그 명칭을 공식 변경하였다.<sup>15)</sup> 통계청의 사회조사를 포함한 시민서베이는 시민의 생각과 의견을 묻는데 유용하며 정량적 정보를 지속적으로 생산한다는 장점이 있다. 하지만, 조사설계가 잘못될 경우 결과가 왜곡될 수 있으며, 일회성 서베이의 경우 여론의 빠른 변화를 반영하기 어렵다는 단점도 있다(Denhardt & Denhardt, 2011).

15) <http://kostat.go.kr/survey/society/index.action> (2015년 1월 9일 오후 5시 44분 최종접속)

사회조사와 더불어 여론조사는 시민의 생각과 의견을 파악하고자 할 때 자주 활용된다. 최근 소셜미디어의 대중화로, 우편, 전화, 인터넷 기반 여론조사의 대안으로써 소셜 여론조사가 실시되고 있는데(American Association for Public Opinion Research, 2014), 질병확산, 선거캠페인, 산업계의 의사결정 등이 주요 활용 분야이다(박한우, 2012). 아래 표는 일반 여론조사와 소셜 여론조사의 특성이 어떻게 다른지를 보여준다.

<일반 여론조사와 소셜 여론조사의 비교>

| 구분       | 일반 여론조사                  | 소셜 여론조사                                  |
|----------|--------------------------|------------------------------------------|
| 모집단      | 알려져 있음                   | 매체에 따라 다름                                |
| 표본대상     | 일반인 전체                   | 인터넷 및 SNS 이용자                            |
| 표본방식     | 확률적 샘플링                  | 매체에 따라 다름<br>(대개 비확률적, 편의적전문가 샘플링)       |
| 조사주제     | 모든 이슈 가능                 | 사회적 흥미를 끌거나 주목받는 이슈                      |
| 분석단위     | 개인단위의 조사<br>(자기 응답식 데이터) | 개인보다는 집합적 수준<br>(이용자가 남긴 메시지 및 링크데이터)    |
| 분석데이터 특성 | 객관적 응답이 많음               | 대화형부터 링크까지 다양한 형태                        |
| 조사대상의 특성 | 알려져 있음                   | 매체에 따라 다름                                |
| 대상별 가중치  | 동등함                      | 조사 목적 및 분석 기법에 따라 다름                     |
| 대상자의 수집  | 조사자가 선별                  | 매체에 따라 다르나, 자발적 참여자가 많음                  |
| 조사방법     | 집 전화위주<br>(휴대전화 보완 시도)   | API를 이용한 데이터의 자동수집<br>(최근에는 설문조사로 보완 시도) |
| 여론 수집방식  | 조사자의 요청                  | 응답자의 자발적 동기                              |
| 조사의 비강제성 | 낮음                       | 높음                                       |
| 조사 비용    | 높음                       | 낮음                                       |
| 조사 규모    | 수백명                      | 최대 수십억 개의 메시지 및 링크                       |
| 조사 간격    | 월별~연도별                   | 일별~주별                                    |

출처: 박한우, 2012

## 2) 활용 사례

독일, 미국, 일본 등 해외 여러 국가들에서 국가 및 지역 정책을 지원하기 위해 사회/여론조사를 실시하고 있다. 가령 독일의 경우, 연방 교통건설부의 지원으로 정통한 여론조사 기관이 독일자전거 클럽과 공동으로 교통수단과 선호도, 교통수단 이용현황, 자전거 용도와 대중교통 연계 등에 대한 설문조사를 실시하고, 그 결과를 토대로 자전거가 미래 교통수단으로 부상할 것을 예측하고 자전거길 확장 등에

대한 정책을 수립하였다(머니바이크, 2013년 1월 18일자). 미국과 일본에서는 국가 단위에서의 사회변화를 모니터링하고 다른 국가와의 비교 자료를 구축하기 General Social Survey(GSS)를 실시하고 있다.<sup>16)</sup> 또한, 뉴욕과 동경 등의 도시 단위에서도 도시의 발전현황을 점검하고 장단기적 발전전망에 맞는 정책입안을 위해 도시정책지표를 모니터링하고 있다. 더불어, 미국의 휴스턴은 1982년부터 교육, 범죄, 이동성과 교통, 환경문제에 대한 평가 등 시민태도 변화를 조사하고 있다.

국내에서도 각 지역사회의 사람들이 무슨 생각으로 어떻게 살아가고 있는지를 모니터링하기 위해 다양한 사회조사를 실시하고 있다. 다음의 표 <사회조사를 시행 중인 국내지역>에 정리된 것처럼, 서울을 비롯한 부산, 인천, 광주, 울산 등의 광역기초 단체를 비롯해 세종특별자치시와 제주 특별자치도, 서울의 강남구, 마포구, 중구, 경상북도의 구미시, 전라남도의 순천시, 여주시, 광양시 등의 기초자치단체에서도 지역사회조사를 실시하고 이를 토대로 다양한 지표를 생산하고 있다. 이 중 서울시는 2003년부터 서울서베이 기본계획을 수립하여 매년 2만 가구를 대상으로 가구현황, 주거와 생활, 세대와 계층, 복지, 교통, 환경 등 다양한 분야를 조사하고 있으며, 그 결과를 기반으로 1인가구, 베이비부머, 중산층, 시민행복 등에 대한 지역정책 수립의 기초자료를 제공하고 있다. 또한 2006년부터는 마포구를 시작으로 총 7개의 자치구가 자체적으로 사회조사를 실시하여 주민들의 정책수요를 파악하고 지역정책 수립의 기초 자료로 활용하고 있다.

사회조사를 지역정책과정에 적극 활용하고 있는 또 다른 국내 지역은 대구시이다. 대구시의 경우, 사회조사를 활용하여 지역민이 처해있는 현실을 파악함과 동시에 SWOT분석을 통하여 이를 지역정책에 반영하고자 하는 노력을 기울이고 있다. 2011년부터 대구사회조사를 시작하였고, 설문을 통해 대구지역 현안에 관한 주민생

16) GSS는 미국 시카고대학교 NORC(National Opinion Research Center)에서 1972년부터 매년 시행되고 있는 국제적인 사회과학조사이다. 한국은 한국종합사회조사(Korean General Social Survey; KGSS)란 이름으로 2003년부터 이 조사에 참여하고 있으며, 2012년 현재 한국을 포함한 미국, 영국, 중국 등 49개국이 정치 및 사회의식, 일상생활 등의 내용을 다룬 반복핵심모듈과 특별한 주제 연구를 위해 개발된 모듈들을 그 설문 내용으로 하고 있다([http://www.kossda.or.kr/w02\\_01e.asp](http://www.kossda.or.kr/w02_01e.asp) 2014년 12월 13일 오후 1시 15분 최종 접속).

각을 조사하고 한 걸음 더 나아가 조사된 자료를 이용하여 다양한 지표를 산출하고 있다. 대표적 예가 주민행복지수와 청년희망지수로, 현재의 사실 뿐만 아니라 미래의 발전가능성까지를 분석하여 고용기회 개선, 고용불안정 개선, 세대간 갈등 해소 등의 정책을 제언하였다(김규원·신형진·권기욱, 2013). 대구시와 서울시의 사례를 종합해보면, 결국 사회조사는 단기적으로 현안이슈에 대한 기초자료를 제공하지만, 장기적으로는 시정운영을 위한 정책판단의 근거를 제공할 수 있음을 알 수 있다.

〈사회조사를 시행 중인 국내 지역〉

| 구분                         |       | 해당 지역                                              |
|----------------------------|-------|----------------------------------------------------|
| 광역기초단체                     |       | 서울, 부산, 대구, 인천, 광주, 대전, 경기, 강원, 충북, 충남, 전남, 전라, 경남 |
| 특별자치시(도)                   |       | 세종, 제주                                             |
| 기<br>초<br>자<br>치<br>단<br>체 | 서울특별시 | 종구, 강남구, 마포구, 광진구, 강동구, 중랑구, 성북구                   |
|                            | 대전광역시 | 대덕구, 중구, 유성구                                       |
|                            | 경기도   | 용인시, 성남시, 부천시, 군포시, 하남시, 양평군, 여주군, 오산시             |
|                            | 강원도   | 태백시, 춘천시, 원주시                                      |
|                            | 경상남도  | 진해시, 울주군                                           |
|                            | 경상북도  | 구미시, 상주시                                           |
|                            | 전라남도  | 여수시, 광양시, 순천시                                      |
|                            | 전라북도  | 전주시, 군산시                                           |
|                            | 충청남도  | 계룡시, 아산시                                           |
|                            | 충청북도  | 청주시, 천안시, 옥천군                                      |

사회조사와 함께 여론조사도 국내 지역정책을 지원하는데 활발히 이용되고 있다. 광주시의회는 사업추진이 지연되고 있는 도시철도 2호선 건설 여부에 대해서 시민여론조사를 실시하고, 공청회나 TV토론회 등도 개최하여 정책추진 여부에 대한 최종판단을 내리겠다고 결정한 바 있다(news1뉴스, 2014년 11월 5일자). 또한, 안전행정부는 2013년 정부가 추진했던 정책에 대한 인식수준, 성과평가, 향후 중점 추진 분야 등에 대해 일반국민, 공무원, 전문가를 대상으로 여론을 파악하고 2014년 업무계획 수립의 기초자료로 사용하였다(안전행정부, 2013).

## 2. 공론조사

### 1) 개요

공론조사는 James S. Fishkin 교수가 제안한 과학적 여론조사와 소집단토의가 접합된 공론수렴의 새로운 방식으로 일반 여론조사에 대한 대안으로 제시되었다(심재웅, 2003). 일반 여론조사는 조사와 관련된 이슈를 시민이 충분히 생각할 겨를 없이 조사 상황에서 반사적으로 떠오르는 의견을 표출하거나 언론보도의 영향을 반영하는 경향이 있다(김춘석, 2013b). 즉, 일반적으로 시민들은 정치, 사회 현안들에 대해 많은 정보를 접하지 않은 상태로 자신의 의견을 형성하게 된다. 세부적인 사항을 면밀히 살펴본 뒤 의견을 형성하기엔 너무 많은 시간과 노력이 들며, 그러한 투자 대비 시민 개개인에게 돌아오는 이익은 눈에 보이지 않기 때문이다. 결국 시민들은 사운드바이트<sup>17)</sup>와 헤드라인에 의존하고 이데올로기나 당파성에 기대어 자신의 의견을 형성하게 되는 경우가 많다. 이에 새롭게 등장한 방법이 공론조사로, 특정주제에 대해 정보를 폭넓게 제공하고 토의를 통해 그 주제에 대한 이해도와 관심도를 제고함으로써 보다 의미있는 응답을 적극적으로 이끌어 내는 조사방법이다. 공론 조사와 일반 여론조사의 구체적 차이점은 다음의 표에 정리하였다.

### 2) 활용 사례

공론조사는 1994년 4월 영국에서 ‘늘어나는 범죄문제에 대한 대책’을 주제로 처음 실시되었다(심재웅, 2003). 미국의 경우, 1996년 1월 ‘국정의제토론회’에서 공론조사를 채택했으며, 이후 외교정책, 의료보험, 교육정책, 미래정책 등에 대한 주제에 관한 논의에서도 실시되었다. 또한, 중국 저장성 원링현은 지역예산 할당과

---

17) 인상적인 한마디 혹은 짧은 코멘트를 지칭하며, 화제 인물의 짤막한 육성 발언이나 연설, 성명의 한 구절 등이 그 예이다. 방송, 특히 뉴스프로에 인용 삽입되는 경우가 많다 (<http://endic.naver.com/enkrEntry.nhn?sLn=kr&entryId=878e2c9f24254d22b4d5631914b7767c&query=soundbite>; 2015년 1월 9일 오후 5시 52분 최종접속).

운영을 위해 3번의 공론 조사를 시행하였으며, 마카오에서는 2012년에 언론법과 언론인 윤리에 대한 공론조사가 실시되었다.

<일반 여론조사와 공론조사의 비교>

| 구분                          | 일반 여론조사                     | 공론조사                             |
|-----------------------------|-----------------------------|----------------------------------|
| 목적                          | 'What the public thinks' 파악 | 'What the public would think' 파악 |
| 참여자 특성                      | 고립된 개인                      | 공론장(public sphere)의 주체로서의 개인     |
| 참여자 관심도<br>(attention)      | 낮음                          | 높음                               |
| 제공되는 정보<br>(information)    | 적음                          | 많음                               |
| 참여자 숙의 정도<br>(deliberation) | 낮음                          | 높음                               |
| 참여방식                        | 수동적                         | 능동적                              |
| 커뮤니케이션 방식                   | 일방향                         | 쌍방향                              |

출처: 김춘석, 2013a

국내에서는 2003년 북한산국립공원 관통 서울외곽순환도로 등과 같은 사회적 갈등과 관련해서 처음 공론조사가 도입될 예정이었으나 무산되었고, 2005년 재정경제부의 8.31 부동산 대책에 대한 논의를 시작으로 공론조사가 본격 활용되었다. 이후, 2007년 부산항만공사가 부산 북항 재개발 마스터플랜 선정을 위해, 2011년 경찰청에서 3색 신호등 도입에 따른 갈등해소를 위해 공론조사를 실시한 바 있다(심재웅, 2003; 김춘석, 2013a).

### 3. 시민참여형 토론

#### 1) 개요

시민참여형 토론은 정책관계자들이 서로의 말에 귀를 기울일 수 있는 토론의 장을 마련하여 단순한 정보교환 이상의 효과를 거두고자 할 때 이용 가능하다. 구체적으로, 지역문제에 대해 토의하는 과정을 통해 정책관계자들은 서로에게 배울 수 있는 기회를 얻게 되고 친밀감을 형성할 뿐만 아니라, 지역에 대한 소속감을 강화할 수 있다(Denhardt & Denhardt, 2011; 서울시정개발연구원, 2012). 시민참여형

토론은 타운홀미팅(Town Hall Meeting), 워크아웃 타운미팅(Workout Town Meeting), 오픈스페이스(Open Space), 월드카페(World Cafe), 시민원탁회의(The 21<sup>st</sup> Century Town Meeting) 등 다양한 방식으로 진행 가능하다. 시민참여 토론 방법들의 특징을 정리하면 아래 표와 같다.

<시민참여 토론 방법들의 특징>

| 구분       | 타운홀미팅                                                | 워크아웃<br>타운미팅                                             | 오픈스페이스                                                 | 월드카페                                                             | 시민원탁<br>회의            |
|----------|------------------------------------------------------|----------------------------------------------------------|--------------------------------------------------------|------------------------------------------------------------------|-----------------------|
| 토론<br>방식 | 지역사회<br>구성원 전체가<br>마을회관(타운<br>홀)에 모여<br>토론하고<br>의사결정 | 조직의<br>불필요한<br>업무를 없애기<br>위해<br>현장직원들이<br>참가하는 토론<br>워크숍 | 자기가 원하는<br>발언을 하고<br>관심있는<br>사람들이 모여<br>자유로운<br>토론을 진행 | 주최측에서<br>테이블마다 서로<br>다른 주제를<br>선정하여<br>토론자리를<br>옮겨다니는<br>방식으로 진행 | 소규모 집단과<br>기술을 결합한 포럼 |
| 실시<br>연도 | 17세기 초반                                              | 1986                                                     | 1985                                                   | 1995                                                             | 1995                  |
| 참여<br>규모 | 마을단위                                                 | 소수~200명                                                  | 소수~200명                                                | 최소 20명 이상                                                        | 500명~5,000명           |

출처: 서울시정개발연구원(2012)

먼저, 타운홀미팅은 영국 식민지 시대 미국 뉴잉글랜드 지역의 통치시스템에서 유래된 것으로, 최근 미국 대선 토론회에서 이용되면서 대중들에게 많이 알려졌다. 타운홀 미팅을 통해 지역 사람들은 한 자리에 모여 토론을 한 후 대표를 선출하고, 투표를 통해 법과 정책, 행정절차에 대해 결정을 내렸기에 미국의 민주주의를 발전시킨 기초가 되었다는 평가를 받기도 한다(김형호·송경재, 2011). 오늘날의 타운홀미팅은 정책결정권자 또는 선거입후보자가 주민들을 초대하여 정책 또는 주요 이슈에 대해 설명하고, 의견을 듣는 비공식적 공개회의로 지역사회에 정보를 제공하고 토론 결과를 피드백 하는 것을 목적으로 한다. 최근에는 정치인뿐 아니라 기업의 최고경영자들도 회사와 관련된 현안 또는 상황을 임직원과 공유하는 의사소통의 장으로 타운홀미팅을 활용하기도 한다.

타운홀 미팅은 오프라인 뿐만 아니라 온라인을 통해서도 이루어지며 이를 e-타운홀 미팅(e-town hall meeting)이라고 한다. 예를 들면, 미국의 백악관에서 실시하는

타운홀미팅은 실제 오프라인에서 평균적으로 100여명 안팎의 국민이 참여하지만, 이를 백악관 홈페이지를 통해 인터넷으로 생중계하여<sup>18)</sup> 훨씬 많은 국민들이 인터넷을 통해 타운홀미팅에 참여할 수 있다. 또한, 참여자들은 궁금한 사항에 대해서는 온라인을 통해 실시간으로 대통령에게 질문을 할 수 있고, 이견이 있는 내용에 대해서는 토론을 할 수 있기 때문에 큰 호응을 얻고 있다(이원태·차재관·신호철, 2012).<sup>19)</sup>

둘째, 워크아웃 타운미팅은 1986년 토마스 에디슨이 시작하였으며, 1980년대 미국 GE의 잭 웰치가 재활성화하였다. 이 방법의 핵심은 효과적인 의사소통과 문제해결을 위해, 마을 회의를 하듯이 자유스럽고 거침없는 분위기에서 토론을 진행하는 것으로, 필요한 경우 옷차림 규정을 없애거나 상관의 배석을 금지하기도 한다. 궁극적으로 워크아웃 타운미팅의 목표는 모든 구성원이 공유할 수 있는 워크아웃 핵심과제를 도출하고, 함께 문제를 해결하면서 조직의 능률 향상은 물론, 구성원간의 신뢰와 협업 능력을 제고하는 것이다.

셋째, 오픈스페이스는 북미의 조직 전문가인 Harrison Owen이 커피 브레이크에서 영감을 얻어 제시한 방법으로, 틀과 격식을 최소화하여 자유로운 분위기 속에서 효과적인 논의결과를 끌어내는 토론 방식이다. 오픈스페이스에서 회의 주최자는 ‘시간’과 ‘장소’와 같은 뼈대만 준비하고, 참여자가 스스로 행사내용과 진행자를 채워가는 참여자 주도의 토론회이다. 진행 방법은 세션별로 개요와 주요 논제를 정하고, 참석자는 참가한 세션에 대한 관심이 떨어지거나, 더욱 생산적인 토론을 원한다면 언제든지 자유롭게 세션을 바꿀 수 있다(서울시정개발연구원, 2012).<sup>20)</sup>

넷째, 월드카페는 Juanita Brown과 David Isaacs가 제안한 것으로 카페와 같은 공간에서 창조적인 토론을 통해 지식의 공유나 생성을 유도하고자 하며, “지식과 지혜는 딱딱한 회의실에서 만들어지는 것이 아니라, 열린 공간에서 이루어지는 사람들 간의 토론을 통해 형성된다”라는 모토를 가지고 있다. 진행 방법은 최소 20명 이상의

---

18) <http://whitehouse.gov/live> (2015년 1월 9일 오후 5시 52분 최종접속)

19) <https://www.facebook.com/maumceo/posts/473467326037377> (2015년 1월 9일 오후 5시 52분 최종접속)

20) <http://www.openspace.kr> (2015년 1월 9일 오후 5시 52분 최종접속)

참여자들 대상으로, 4-5인의 소그룹을 구성하여 각 소그룹 별로 상이한 주제로 토론을 하다가 정해진 시간이 되면 그룹대표만 남겨두고 모두 다른 그룹으로 이동한다. 이런 그룹이동을 몇 차례 거치면 월드카페의 모든 참석자 간에 지식이 공유되는 효과를 얻을 수 있다(서울시정개발연구원, 2012).<sup>21)</sup>

마지막으로 21세기 타운미팅은 복잡한 공공정책 문제<sup>22)</sup>에 대한 의사결정과정에서 시민의 직접적이고 개별적인 목소리를 반영하기 위한 수단으로서 미국의 아메리카스픽스에서 개발하였다. 이 방법에서는 시민들이 숙의과정을 통해 의제에 대해 다양한 의견을 낼 수 있고, 토론 후에는 구체적이고 실질적인 실행계획을 만들어 의사결정과정에 토론결과가 반영될 수 있는 구조를 갖추고 있다. 토론 과정은 맥락설정, 가치의 명확화, 권장사항 개발, 레포팅이라는 4가지 단계로 이루어지며, 한 가지 의제를 두고 모든 참가자가 동시에 토론할 수 있다. 한 테이블은 1명의 퍼실리테이터(facilitator)와 최대 10명의 토론자로 구성되며, 퍼실리테이터는 토론자의 발언내용을 발언자의 언어로 정확하게 노트북에 입력하고, 중앙에서는 실시간으로 입론내용을 취합하여 테마 별로 분석한 후, 권장할만한 추진사항들을 추출하고 이들의 우선순위와 선결추진 여부를 무선투표 시스템으로 현장 투표하여 결정하는 방식이다.

## 2) 활용 사례

시민참여형 토론방법들은 해외 주요도시에서 지역정책 수립시 적극 활용하고 있다. 대표적인 예로는 2012년 시카고 문화플랜(Chicago Cultural Plan 2012), 멜버른의 예술전략(Melbourne Arts Strategy), 뉴욕의 세계무역센터 재건축계획 수립 등이 있다. 우선 시카고 문화플랜은 시카고 미래 문화와 경제 성장을 위한 체계를 만들고 시카고 예술분야에서 창의성, 혁신, 우수성을 가진 글로벌 중심지로 거듭나는데

21) <http://theworldcafe.com> (2015년 1월 9일 오후 5시 52분 최종접속);

<http://userexperience.tistory.com/258> (2015년 1월 9일 오후 5시 52분 최종접속)

22) 21세기 타운미팅에 적합한 공공업무의 3가지 범주로는 1) 계획 및 경제개발, 2) 예산과 자원 할당, 3) 정책형성이 있다(Lukensmeyer, 2005).

중추적 역할을 하는 것을 목표로 하였다. 이 목표 달성에 있어서 계획수립 과정 자체도 중요시하여, 시민과 함께 문화플랜의 비전을 창조하는데 초점을 맞추었다(라도삼, 2014). 시카고 문화플랜은 크게 연구분석, 시민참여, 비전 및 방향 설정의 세 단계로 나누어 추진되었는데, 시민참여 단계에서는 미래의 예술과 문화를 위해 총 4,700명이 넘는 시카고 시민들 8번의 타운홀 미팅, 20여번의 지역사회와의 대화, 소셜미디어를 통한 시민과의 지속적인 교류, 10번의 부분별 미팅 등을 통해 향후 25년의 문화 분야의 발전방향을 제시하였다.

<2012년 시카고 문화플랜 수립을 위한 타운홀 미팅>



출처: City of Chicago, 2012

뉴욕시의 세계무역센터 재건축이 시민형 참여 토론의 두 번째 활용사례이다. 뉴욕시는 9.11테러로 폐허가 된 세계무역센터를 재건축하는 과정에 시민들을 참여시키고 그들의 생각을 재건축 계획에 반영하였다. 이 과정에서 가장 혁신적인 활동은 “Listening to the City”라는 시민포럼으로, 다양한 배경을 가진 약 4,300명의 시민이 참여하여 세계무역센터 재건축에 관한 생각을 말하고 경청하였다. 이 포럼은 시민 간의 상호 배움의 장이 되었을 뿐만 아니라, 공무원들도 시민들의 고충과 조언을 들을 수 있는 상호 이해의 기회가 되었다(Denhardt & Denhardt, 2011).

시민참여형 토론 활용의 세 번째 해외사례는 호주의 멜버른시이다. 이 도시는 2013년 창조도시 구현을 위한 방법으로 멜버른 예술 전략(Melbourne Arts Strategy) 보고서를 발표하였는데, 이를 위해 시와 예술가, 시민들이 협력하여 예술전략을 제시하였다. 예술가와 시민 등은 포럼이나 라운드테이블 등을 통해 전략을 수립하는데 참여하였으며 약 1천여명의 사람들이 토론에 참여하여 1,200여개의 의견을 제시하였다(라도삼, 2014).

시카고, 뉴욕, 멜버른 외의 해외도시에서도 시민참여형 토론방법을 활용하여 다양한 지역문제를 해결하고 있다. 미국 콜로라도주는 다양한 이해관계(인디언원주민, 연방정부, 지방정부 등)가 복잡하게 얽혀 있던 15억 달러 예산의 고속도로 건설 문제를 해결하기 위해 오픈스페이스를 개최하였고 총 225명의 이해 당사자들이 3일간의 회의를 통해 문제해결방안을 만들어냈다.<sup>23)</sup> 또한, 미국 켄터키주의 오웬스보로(Owensboro)에서는 지역경제 및 교육 문제의 해결을 위한 21세기 타운미팅을 개최하여, 복잡한 갈등상황 속에서도 시민과 함께 합의를 도출해내는 민주적 의사결정체제를 시범적으로 보여준 바 있다(서울시정개발연구원, 2012; Lukensmeyer, 2005).

국내에서도 서울시와 부산시에서 시민참여형 토론을 정책과정에 적극 적용하고 있다. 서울시의 경우 시민수요조사, 원탁토론, 정책토론회 등 다양한 방식으로 시민들의 여론을 수렴하고 그 결과를 정책수립에 반영하고자 지속적인 노력을 기울였다. 가령, 2011월 11월 25일부터 12월 4일까지 서울거주 19세 이상 시민 540여명을 대상으로 가구방문 면접조사와 집단 심층면접 조사(Focus Group Interview)를 실시하여 시민들이 원하는 서울시정의 중점과제가 무엇인지, 10대 희망공약의 분야별 우선 추진과제는 무엇인지를 파악하여 이를 중장기 계획의 기초자료로 활용하였다. 또한, ‘서울시민 100인, 2030미래서울을 말하다’라는 주제로 2012년 6월 서울시민 100인이 한자리에 모여 2030년의 미래 서울의 모습에 대해 이야기 나누고, 미래의 서울이 가져야 할 도시기능이 무엇인지 토론하고 기획하였다. 2030년 미래서울을 살아갈 시민과 함께 서울의 행정과 정책을 담당하는 행정가, 전문가, 그리고 시민사회

23) <http://www.openspace.kr> (2015년 1월 9일 오후 5시 52분 최종접속)

단체가 모두 토론회에 참가함으로써 미래 서울에 필요한 도시기능(공정성, 창의성, 포용성, 안전성)에 대해 공론화하고 보다 살기 좋은 서울이 되기 위해서 도시의 다양한 기능들이 어떻게 조화를 이루어야 할지에 대해 토론하였다. 이 외에도 교육도시 서울 플랜을 위해 500인 원탁회의를 개최하고, 도시기본계획인 ‘2030 서울플랜’을 우리나라 최초로 시민참여형으로 수립한 바 있다(변미리, 2012).

부산시의 경우 2007년 부산역 주변의 항만시설 개발에 대한 시민들의 의견을 수렴하기 위해 토론회를 개최하였다. 부산시는 패널로 선정된 1,000명의 시민에게 부산항 일대의 도시개발에 관한 두 가지 대안에 대해 자세하고 중립적인 정보를 담은 자료를 우편으로 송부하고 이 내용을 지역방송의 토론프로그램을 통해서도 전달하였다. 다음으로, 1,000명의 패널 중 100명을 선정하여 총 10개의 소그룹으로 나눈 뒤, 항만시설 개발에 관한 두 개의 대안을 두고 각 소그룹에서 토론을 하도록 하였다. 토론 후, 소그룹들은 두 가지 대안과 관련하여 전문가들에게 질의하는 종합토론을 과정을 거치고 종합적 의견을 제시하였다(심재웅, 2013; 김춘석, 2013a).

국토연 2014-10

여론 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구

지 은 이 황명화

발 행 인 김경환

발 행 처 국토연구원

출판등록 제25100-1994-2

인 쇄 2014년 12월 31일

발 행 2014년 12월 31일

주 소 경기도 안양시 동안구 시민대로 254

전 화 031-380-0114

팩 스 031-380-0470

ISBN 979-11-85948-22-5

한국연구재단 연구분야 분류코드 B170300

홈페이지 <http://www.krihs.re.kr>

© 2014, 국토연구원

이 연구보고서의 내용은 국토연구원의 자체 연구물로서  
정부의 정책이나 견해와는 상관없습니다.