

## 결과보고서 요약

<b>사업기관</b>	언론정보연구소		
<b>사업명</b>	통일기반 구축사업		
<b>사업책임자</b>	<b>성명</b>	<b>소속</b>	<b>직위</b>
	한규섭	언론정보학과	교수
<b>사업기간</b>	2019.4.1. ~ 2020.2.28	<b>사업비</b>	

### 1. 사업 목적

- 서처블한 데이터베이스와 분석도구를 함께 갖춘 초보적 형태의 북한 미디어 분석 엔진을 구축하고 디지털 북한학의 개척을 위한 초석을 제공
- 북한에 대한 객관적 자료를 생성하고 국내외 씽크탱크들과의 네트워크를 구축함으로써 대북정책 수립에 있어 한국의 주도권을 확보
- 세계북한 연구자 및 언론 매체를 위한 데이터베이스를 제공하고 이를 통해 북한에 대한 과학적 분석과 객관적 보도를 유도

### 2. 사업 내용

- 1950 ~ 2018년까지의 주제별 ‘노동신문’ 기사 약 16만 건 수집한 후 텍스트를 추출하고 자연어처리를 통해 데이터 정교화
- 기사에 대해 기계학습을 통한 논조분석을 수행한 후 역사적 변곡점을 분석
- 각 기사들의 주요 아젠다를 분석하기 위해 등장 단어에 대한 군집분석을 수행. 또 의미망 분석을 통해 주요 아젠다의 프레임을 분석
- 분석 결과를 바탕으로 미국 워싱턴 지역 북한 관련 씽크탱크들과 연구관련 협의를 진행

### 3. 사업 성과

- 노동신문 헤드라인 분석을 통해 한, 미, 일, 중에 대한 북한 보도 분석
- 기계학습 방식을 적용, 노동신문의 논조를 데이터 사이언스적 방법론으로 분석

# 언론정보연구소

## 1. 사업 배경

### □ 개요

- 사업명: 통일기반 구축사업
- 기관명: 언론정보연구소

### □ 추진 배경 및 목적

#### ○ 추진배경

- 통일에 대한 기대가 높아지고있으나 실제로 북한 사회와 체제에 대한 객관적 이해가 매우 부족한 상황임. 기존의 주관적이고 규범적인 연구를 넘어 북한의 현 상황을 객관적으로 파악할 수 있는 계량적 연구방법론을 도입할 필요가 있음. 특히 빅데이터 기술의 발달로 기존 북한 연구의 한계를 뛰어넘을 수 있는 환경이 조성됨.
- 국제학계에서 북한 관련 체계적 연구가 매우 부족한 상황. 기존 실증 데이터에 기반하여 북한 관련 지식의 확대에 매진해 온 해외 정책연구 기관(스팀슨센터, 38노스)도 한국어 구사 인력을 확보하기 어렵다는 점과 데이터 사이언스적 분석이나 대용량 데이터 수집능력이 부재하다는 점에서 한계를 가짐.

#### ○ 목적

- 「북한 데이터 포털」을 구축하여 여러 곳에 산재되어있는 북한 관련 데이터의 효율적인 활용을 가능하게 하고 이를 통해 '디지털 북한학'을 개척할 필요가 있음. 이 과제에서는 서처블한 데이터베이스와 분석도구를 함께 갖춘 초보적 형태의 북한 미디어 분석 엔진을 구축하고자 함.
- 구체적으로 1950년대부터 수집된 약 16만 건의 '노동신문' 기사에 대해 기계학습을 통한 논조분석과 의미망 분석, 토픽모델링 등을 수행하여 기간별 아젠다 및 논조분석을 수행. 또 변곡점 분석을 통해 아젠다 및 논조의 터닝포인트가 된 역사적 사건에 대해 연구.
- 향후 이러한 분석엔진을 확대함으로써 국내외 싱크탱크들과의 네트워크를 구

축하고 대북 정책 결정에 있어 한국의 주도권을 확보할 수 있을 것으로 기대. 또 전 세계 도서관 및 언론사에서 분석엔진을 구독할 수 있게 함으로써 장기적으로 북한에 대한 과학적 분석과 객관적 언론보도를 유도.

## 2. 사업 내용

### □ 사업 추진 내용

- 1950년 ~ 2018년까지의 주제별 ‘노동신문’ 기사 약 16만 건 수집
  - ‘미국’, ‘남한’, ‘중국’, ‘일본’, ‘핵’, ‘경제’ 등 여섯 가지 주제에 대해 키워드 75개를 선정한 후 이를 기반으로 관련 기사의 헤드라인 총 160,035개 수집.
  - 통일부 산하 북한자료센터의 데이터베이스를 활용
- 수집된 기사에서 텍스트 데이터를 추출하고 이에 대해 자연어처리(Natural Language Process: NLP)를 통해 데이터를 정교화함
- 주제별 기사에 대해 기계학습을 통한 논조분석을 수행하고(긍정,부정,중립), 그 빈도에서의 ‘변곡점’을 추정하여 이를 역사적 사건과 관련지어 해석
- 각 기사들의 주요 아젠다를 분석하기 위해 등장 단어에 대한 군집분석을 수행. 또 의미망 네트워크 분석을 통해 주요 아젠다가 어떻게 ‘프레임’되어 있는지 분석
- 분석 결과를 바탕으로 미국 워싱턴 지역 썬크탱크들과의 연구관련 협의를 진행
  - Wilson Center, CSPS, Schar School of Public Policy Studies at George Mason University, Stimson Center, Library of Congress 관계자와 협력 방안에 대해 논의

### □ 일정별 추진 사항

일 정	추진 내용
2019.03. ~ 2019.05	노동신문 헤드라인 수집 및 데이터 전처리
2019.06 ~ 2019.07	노동신문 기계학습 논조분석, 아젠다 분석, 의미망 네트워크 분석
2019.07	CSPS(Center for Security Policy Studies) 및 워싱턴 지역 싱크탱크들과 북한미디어 연구를 위한 협의
2019.11.27	서울대학교 연합학술대회

### 3. 사업 성과

#### □ 사업 성과

##### ○ 연구성과

- 본 사업 내용을 조지 메이슨 대학교 안보연구소와 공동으로 2020년 중 저서로 출판 예정
- 현재 본 과제의 연구성과물을 기초로 추가 연구비 수주 지원, 심사 진행중.
- 본 과제의 연구성과물을 기초로 북한데이터 포털 구축을 위한 연구 개시

##### ○ 네트워크 구축

- 조지 메이슨 대학교 안보연구소 (Center for Security Policy Studies) 방문
  - 일정: '19. 7. 6 ~ 7. 27
  - Allen Laipson 소장 및 Mark Rozell 학장과 수차례 미팅
  - 한국 세미나등 계획
- Stimson Center 회의
  - 일정: '19. 7. 17
  - 38North Jenny Town 연구원과 회의
  - MOU를 포함한 공동 연구 추진 협의
- Woodrow Wilson Center 회의
  - 일정: '19. 7. 10
  - Jean Lee 한국센터 소장과의 미팅
  - 한국 세미나등

- Library of Congress
  - 일정: '19. 7. 15
  - Sonya Lee 한국컬렉션 디렉터와 회의
  - 북한 미디어 데이터베이스 파악

## □ 관련분야 기여도

- 북한관련 연구의 체계화
  - 데이터가 부족한 북한에 대한 계량 연구 사례 제시
  - 북한 관련 데이터에 대한 데이터 사이언스적 접근방식 적용

## 5. 자체평가 및 건의사항

### □ 사업평가

- 사업의 적절성
  - 본 사업에서 제안한 계량연구를 충실히 수행함.
  - 북한 미디어 데이터를 수집하고 분석 방법 제시
- 사업의 효율성
  - 본 사업의 효율적 진행을 위해 해외 씽크탱크들과 협력
- 사업의 발전가능성
  - 본 연구를 기초로 현재 북한데이터 포털 구축을 위한 연구 개시 예정
  - 본 연구를 기초로 작성한 연구비 제안서 심사중

### □ 미흡한 점(한계) 및 개선(보완)할 점

- 구체적인 서비스 개발
  - 본 사업의 최종 목표는 북한 미디어 지표 발표와 북한데이터 포털 구축등 구체적인 서비스를 제공하는 것이나 본 연구는 데이터 분석 위주의 파일럿 성격의 연구였음.
  - 현재 서비스 구축을 위한 연구 추진중

### □ 향후 계획

- 1단계: 주간 북한 뉴스 지표 개발
  - 매주 북한 뉴스의 아젠다와 논조에 대한 다양한 지표를 구축하여 발표
  - 기계학습과 인공지능 방법론을 적용한 텍스트 및 이미지 분석 실시.

- 2단계: 북한 인물 지표 개발 (Image - Text 인물 인식)
  - 텍스트 및 영상물 출현 및 공동 출현 빈도에 기반한 북한 인물 연결망 지표 구축
  - 조선중앙통신 보도에 등장하는 이미지를 인공지능 기법으로 인식하여 북한 체제에서 중요한 위치를 차지하는 주요 인물들의 등장 빈도 등을 데이터화.
  - 기사에서 언급된 인물이 담긴 영상 이미지를 기사 내용과 연결하는 알고리즘을 개발
- 3단계: 북한 뉴스 DB 및 분석 엔진 구축
  - 노동신문과 조선중앙통신 뉴스 인덱스 DB 구축 및 분석 엔진 개발
  - 지속적인 업데이트를 통해 궁극적으로 세계각국의 연구기관 및 도서관 등에 제공
- 4단계: 북한 관련 타 데이터와 연계 분석
  - 뉴스 데이터와 다른 북한 관련 데이터 연계 분석
  - 장기적으로 다양한 북한 관련 데이터를 동시에 제공할 데이터 플랫폼으로 발전
- 5단계: 해외 싱크탱크들과의 교류
  - 다양한 북한 데이터를 분석하여 해외 싱크탱크들과 교류 및 협업 진행
  - 미국 등 이해 당사국들의 정책 결정에 객관적 자료 제공

## 6. 성과 관련 지표

### □ 대표 사업 실적

- 미국 조지 메이슨 대학 안보연구소와 함께 저서 출판 예정 (2020. 10월 발간 예정)
- 미국 조지 메이슨 대학 안보연구소와 함께 워싱턴 D.C. 싱크탱크들과 세미나 개최 예정 (2020년 중)

### □ 성과 자율 지표

- 사사 표기(예정) 저서(해외 출간) : 1편 예정