

GDELT 기반 미국 대통령선거 예측

엄기홍 · 서대원 · 육태훈

I. 서론

본 연구는 미국 소재 언론사를 토대로 미국 언론이 본 Joe Biden 후보와 Donald Trump 후보의 기사 논조와 정책 보도 비중을, 전국 차원 그리고 주 차원에서 분석하고 있다. 특히, 본 연구는 일일 단위로 업데이트함으로써 미국 대통령선거의 추이를 시의적절하게 파악하리라 본다. 이를 토대로, 2024년 11월에 치러질 미국 대통령선거 당선자를 시의에 맞게 예측함과 동시에 후보의 전국 차원 이슈와 주별 이슈를 파악하는 데 목적을 두고 있다.

II. 데이터

1. 데이터 개요

본 연구는 Global Data on Events, Location and Tone(이하 GDELT) 프로젝트에 의해 생성되는 자료를 토대로 미국 대통령선거를 예측하는 데 목적이 있다. 특히, GDELT 데이터의 Global Knowledge Graph(이하 GKG)를 사용하여, 미국 전체 차원에서 그리고 주별 차원에서 조 바이든과 도널드 트럼프 두 후보에 대한 대통령선거의 결과를 예측하고자 한다.

GDELT는 미국의 국립과학재단(National Science Foundation)의 지원을 받아 구글과 대학교가 협력하여, 1979년부터 현재 시점까지 세계 각국의 언론 기사를 15분 단위로 수집하여 제공하고 있다(Leetaru and Schrodtt 2013; 이성우 · 정성희 2020).¹⁾ 특히, GDELT는 영어, 중국어, 한국어 등 약 100개의 언어로 출간되는 기사들을 수집한 후, 머신러닝 기술을 적용하여 영어로 번역하고, 이를 토대로 기사의 출처, 주요 인물, 주요 기관 정보를 비롯한 기사의 주제(Theme), 논조(Tone)를 추출하고 있다.

GDELT에 의해 생성된 데이터는 세계적 차원의 국가 간 갈등(이성우 2022a), 남북한 관계(이성우 2017; 이성우 · 정성희 2020; 이성우 · 임재형 2021; 이성우 2022b) 등 다양한 연구에 사용되고 있다. 예를 들어, 이성우 · 정성희(2020)는 2000년 1월부터 2020년 1월까지 약 20년 동안 일별 자료를 토대로 월별로 남북한 관계를 분석했다. 이성우와 정성희는 남북 간 협력 및 분쟁 관계를 협력 및 분쟁 각각의 빈도 및

1) GDELT Project(<https://www.gdelproject.org/data.html#rawdatafiles>, 방문일: 2023년 9월 13일). GKG 2.1은 15분 단위로 수집되어 GDELT 사이트에 보고되고 있다.

사건유형에 따른 가중치를 곱하여 합산한 수치를 이용하여 협력과 분쟁의 정도를 측정했다. 저자에 의하면, 북한은 남한에 대해 고강도 저빈도의 분쟁을 일으키며, 남한은 북한의 도발행위에 대해 저강도 고빈도의 대응을 하는 패턴을 발견했다. 김규철(2022)은 서울대 통일평화연구원의 ‘통일의식조사’ (2007년-2021년)와 GDLET 자료를 가공한 ‘남북관계지수’를 활용하여, 남북관계가 우호적일수록 남한 국민은 통일이 필요하다고 응답하고, 북한을 우호적인 대상으로 인식하는 등 통일의식이 개선된 것을 발견했다.

2. 데이터 수집 및 정제

본 연구는 Google Bigquery를 사용하여 GKG 2.1에 탑재된 데이터를 취합한다. 데이터 수집 단계에서 인물 정보가 담긴 Persons 키워드 값은 “Biden” 또는 “Trump”로 한정하였다. 따라서 본 연구의 예측치는 “Biden”만 언급한 기사 또는 “Trump”만 언급한 기사에 토대를 두어 개발되었다. “Biden”과 “Trump”를 동시에 언급한 기사의 경우, 기사가 누구를 긍정적 또는 부정적으로 평가하기 어렵다는 판단으로 인하여 분석에서 제외하였다.

수집한 데이터는 언론사 소재지가 미국인 언론으로 한정하기 위하여 전처리하였다. 전처리 방식은 GKG에서 기사 원문 링크를 담고 있는 SourceCommonName 필드의 언론사 URL이 연구진이 선정한 미국 언론사의 URL과 일치하는지이다. 연구진의 미국 언론사 선정은 SCImago의 SCImago Media Rankings²⁾와 Online News Papers³⁾에 근거한다. 연구진이 선정한 미국 소재 언론사는 총 5,131개이며, 해당 언론사는 전국 차원의 언론사인 Associated Press, Reuters 등과 주별 차원의 언론사로 구분된다. 예를 들어, URL이 kentucky.com인 Lexington Herald Leader는 미국 소재 언론사이며, 전국 차원의 언론사가 아닌 미국 켄터키 주 언론사에 해당한다. 이렇게 GKG 데이터는 전국 차원의 후보 동향과 주 차원의 후보 동향 분석이 가능한 형식으로 정제되었다.

각 후보에 대한 언론의 긍부정 정도는 기사 논조(Tone)를 활용하였다. GKG에는 기사 속 단어를 토대로 기사의 감성을 측정된 Positive Score와 Negative Score가 있다. Positive Score는 전체 단어 중 긍정 단어의 퍼센트이며, Negative Score는 전체 단어 중 부정 단어의 퍼센트이다. 각 Score의 범위는 0부터 100이며, 숫자가 클수록 감성의 정도가 크다는 것을 의미한다. Tone 필드에는 [Positive Score - Negative Score]로 계산한 Tone 값이 존재한다.⁴⁾ Tone 값은 이론적으로 -100부터 100까지이

2) SCImago Media(<https://www.scimagomedia.com/>, 방문일: 2023년 11월 30일).

3) Online News Papers 중 미국 주별 사이트(<https://onlinenewspapers.com/usstate/usatable.shtml>, 방문일: 2023년 11월 30일).

4) 여기서 유념할 점은 Tone 값이 0인 경우가 긍정과 부정 단어 비중 간의 상쇄로 인한 경우가 아닌, 중립 단어의 높은 비중으로 인한 경우일 수도 있다는 점이다. Tone 필드에는 이를 측정하기 위한 값인 Polarity가 존재한다. Polarity는 전체 단어 중 긍정·부정 단어가 차지하는 비중으로서 Polarity 값이 크고 Tone 값이 0에 가까울수록, 언론 기사가 긍부정 감성적 단어로 충만하나 긍정·부정이 균

나, 일반적으로 -10부터 10까지 변화하며, 0 값은 중립적인 기사를 의미한다. 본 연구는 Tone 필드의 값을 후보에 대한 언론의 논조로 간주하였다.

본 연구는 GKG의 Theme 필드를 기반으로 기사 주제를 분류하였다. GKG의 Theme 필드에는 GDELT 자체 분류체계를 비롯하여 World Bank 등의 분류체계를 제공하고 있다. 본 연구는 World Bank 분류체계를 활용하여, 후보 언급 기사와 동반된 기사의 주제를 분류하고 있다.⁵⁾ World Bank의 분류체계는 World Bank 관여 사업 및 정책의 성격을 분류하기 위해 구축됐다.

World Bank 분류체계는 24개 최상위 분류 개념(Top Concept)을 시작으로, 상위 개념(Broader Concepts)-하위개념(Narrower Concepts) 구조로 배치되어 있다.⁶⁾ 24개 최상위 분류는 Education(470), Poverty(695), Information and Communication Technologies(133), Social Protection and Labor(697), Social Development(134), Water(137), Transport(135), Environment, Natural Resources and Blue Economy(566), Climate Change(567), Jobs(2670), Macroeconomic and Structural Policies(439), Fragility, Conflict and Violence(2432), Macroeconomic Vulnerability and Debt(1104), Financial Sector Development(1920), Trade(698), Health, Nutrition and Population(621), Public Sector Management(696), Energy and Extractives(507), Agriculture and Food Security(435), Public-Private Partnerships(2206), Private Sector Development(1921), Urban Development(699), Gender(615), Economic Growth(471)이다.

World Bank 분류체계에 존재하는 3,890개의 개념은 최상위 분류 개념으로 환원된다. 예를 들어, Social Media(662)는 Enterprise Applications(658) → ICT Applications(652) → Digital Government(678)로 환원되어 최상 분류 개념인 Information and Communication Technologies(133)로 연결된다. 달리 말하면, 최상위 개념인 Information and Communication Technologies(133) 하위개념에는 Digital Government(678)와 ICT for Jobs(2420) 등이 있다. Digital Government(678) 하위개념에는 ICT Applications(652)가 있고, 그 하위개념에는 Enterprise Applications(658)가 있으며, 그리고 그 하위개념에는 Social Media(662) 등이 있다.⁷⁾ 달리 말하면, 678은 133과 직접적으로 연결된 한편, 2420과는 간접적으로 연결되어 있다.

본 연구는 GKG Theme 필드에 있는 World Bank 분류체계, 특히 최상위 분류개념인 24개 Theme을 토대로 각 후보 관련 기사의 정책 보도 비중을 계산했다. 이를 토대로 Biden 후보와 Trump 후보가 전국 차원에서 그리고 주 차원에서의 정책 보

형되게 존재하는 것을 의미한다.

5) 본 연구 분석 기간 중 World Bank 분류체계가 포함되지 않은 기사는 10%였으며, 이 기사는 분석에서 제외되었다.

6) World Bank 주제 분류체계(<https://vocabulary.worldbank.org/taxonomy.html>, 방문일: 2023년 12월 30일). World Bank 분류체계는 상호 배제적(mutually exclusive)이지 않아 주제가 중첩될 수 있다. GKG 데이터에서 World Bank 분류체계를 포함하지 않은 데이터는 분석 기간에 10%이었다.

7) Information and Communication Technologies 분류체계(<https://vocabulary.worldbank.org/taxonomy/133.html>, 방문일: 2023년 12월 30일).

도 비중을 계산하여 제공하고 있다.

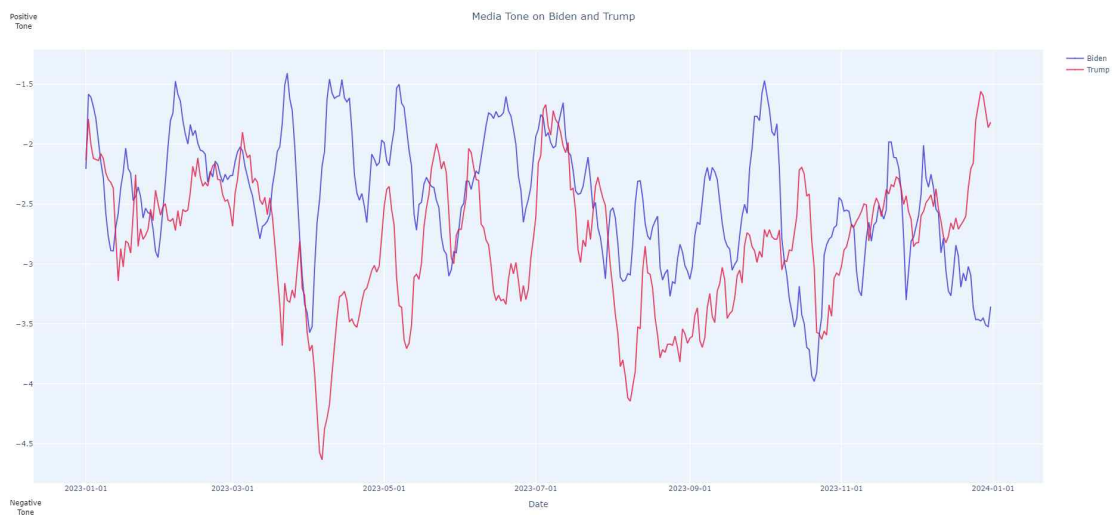
요컨대 본 연구는 GDELT 제공 자료를 토대로 미국 언론에서 언급된 Joe Biden 과 Donald Trump 기사 자료를 추출한 후, 전국 차원과 주 차원에서 이들 후보의 긍부정 정도와 정책 보도 내용을 분석하여 제공하고 있다. 이들 자료를 토대로 2024년 치러질 미국 대통령선거의 결과를 예측하고자 한다.

III. 예측 알고리즘

본 연구는 GDELT에서 제공하는 기사 논조에 토대를 두어 미국 대통령선거의 후보 동향을 일일 단위로 분석한다. 본 연구는 미국 대통령 선거제도의 특성을 고려하여 전국 차원의 예측치와 주 차원의 예측치를 각각 제공한다. 전국 차원의 예측치 계산에는 5,131개 언론사 모두가 활용되는 반면, 주 차원의 예측에 있어서는 전국 차원의 언론사는 사용되지 않은 채 주 차원의 언론사만 활용된다.⁸⁾

전국 차원의 예측치는 모든 언론사의 기사 논조에 토대를 두고 있다. 구체적으로 보면, 전국 차원의 후보 예측치는 Biden 또는 Trump를 언급한 기사를 대상으로 현재 기준 7일 이전 기사들의 논조에 대한 이동평균값(moving average)이다. 따라서, [그림 1]에 제시한 전국 차원의 예측치는 미국 대통령선거 후보에 대한 전반적인 경향을 측정한다.

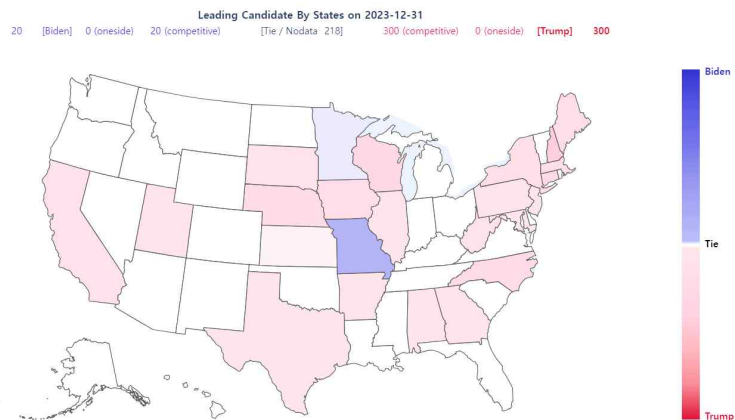
[그림 1] 전국 차원의 미 대통령선거 예측



8) 전국 차원의 언론은 240개이다. 또한, 주 차원의 언론이 여러 주에 걸친 경우도 있다. 이 경우, DocumentIdentifier의 세부 주소를 통해, 기사 제목 혹은 기사의 URL에서 주를 특정할 수 있는 요소를 추출하여, 주를 특정했다.

두 번째 예측치는 주별 언론사들의 기사에 토대를 두어 계산한다. 주별로 지난 7일간 후보 각각의 기사 논조에 대한 이동평균값을 계산하여, 기사 논조가 더욱 긍정적인 후보를 승자로 결정하여 주에 배정된 선거인단을 할당하였다. 예를 들어, [그림 2]에 제시한 것처럼 2023년 12월 25일 기준 캘리포니아 주 소재 언론사가 지난 7일 동안 Biden과 Trump 관련 기사를 보도했고([그림 2]의 competitive), Biden 관련 기사가 Trump에 비해 부정적으로 보도되었다면, 캘리포니아 주에 배정된 54명 선거인은 Trump로 배정되었다.⁹⁾ 만약 후보 한쪽에 대한 기사만 있을 경우([그림 2]의 oneside), 기사 논조에 따라 선거인을 배정했다. 가령 오하이오 주에 Trump 관련 기사만 있고 기사 논조의 이동평균값이 부정적일 경우, 오하이오주 선거인 배정된 17명을 Biden에 할당하였다. 끝으로 현재일 기준 7일 동안 기사가 없을 경우, 예측치를 제공하지 않고 Tie / Nodata로 분류하였다. 주 차원의 예측치는 미국 대통령 선거의 당락을 예측한다.

[그림 2] 주 차원에서의 미 대통령선거 예측



미국 대통령선거를 예측하는 것은 중요한 과제이다. 그러나, 미국 대통령선거를 단순 예측하는 것을 넘어, 전국 차원에서 그리고 주 차원에서 후보의 어떤 의제가 언론에 의해 보도되었는지를 제공하는 것도 한국 외교를 비롯한 다양한 분야에서 중요한 분석이 될 수 있다. 본 연구는 GDELT의 Theme 필드를 활용하여, 언론에 의한 후보 정책 보도 비중을 제공한다. 특히, Theme 필드에서 제공하는 다양한 분류체계 중 활용도가 높은 World Bank Theme을 활용하여 후보의 정책 보도 비중을 제공하고 있다.

후보의 정책 보도 비중은 전국 및 주 차원의 예측치 계산과 같이, 현재일 기준 7일간 최상위 분류 개념에 대한 이동평균 비중값이다. World Bank 분류체계에 존재

9) 이동평균을 계산할 때 데이터가 없는 일자리는 분모에서 제외하였다. 예를 들어, 지난 7일 중 5일간 기사가 있었다면, 이동평균의 분모는 7이 아닌, 5이다.

하는 3,889개의 개념은 24개 최상위 분류개념으로 환원하며, 따라서 정책 보도에 대한 이동평균 비중값은 24개 개념 중 어떤 개념이 주로 언론에 의해 보도되었는지를 보여준다. 다만, 환원 과정에서 한 개 기사는 2개 이상의 최상위 개념으로 분류될 수 있으므로 정책 보도 비중의 분모는 기사 총개수가 아닌 분류된 최상위 개념들의 총개수이다.

[그림 3]은 전국 차원에서의 정책 보도 비중을 보여준다. [그림 3]에서 파란색은 Biden의 정책 보도 비중이며, 빨간색은 Trump의 정책 보도 비중이다. 예를 들어, Biden 관련 정책 보도 비중은 2023년 12월 31일 기준 지난 7일 동안 Transport가 가장 높았으며(7.08%), 그 뒤를 이어 Fragility, Conflict and Violence(6.54%) 등이었다. Trump 관련 정책 보도 비중은 Public Sector Management(18.91%)가 압도적으로 높았으며, Fragility, Conflict and Violence(9.86%) 등이 뒤를 이었다.

[그림 3] 전국 차원에서의 정책 보도 비중



이렇게 계산된 정책 보도 비중은 Biden과 Trump로 구분하여 전국 차원과 주 차원으로 제시한다. 다만, 지면 제약상 전국 차원에서의 정책 보도 비중만을 [그림 3]에서 제시하였다. 이와 같은 정책 보도 비중은 한국 외교 등이 주목해야 할 영역을 전국과 주 차원에서 파악하는 데 기여하리라 본다.

IV. 자동화된 일별 예측 시스템

본 장에서는 미국 대통령선거 예측을 자동화하는 시스템 설계와 구현 방법을 설명한다. 본 연구에서 개발한 시스템은 데이터 추출, 저장, 정제 및 시각화까지의 전 과정을 포괄하며, 이를 통해 미국 대통령선거에 대한 일별 예측을 가능케 하는 데 목적이 있다.

○ 1단계: 데이터 추출, 저장 및 정제

본 시스템의 첫 번째 단계는 Google BigQuery에서 제공하는 GDELT GKG 2.1 데이터에 SQL 쿼리를 수행하여 필요한 데이터를 추출하는 것이다. 데이터는 II장에서 설명한 바와 같이 “Biden” 또는 “Trump” 키워드로 추출하며, CSV 파일 형태로 저장된다. 정제 과정은 불필요한 분류체계 제거, 이동평균값과 예측치 계산 등으로 구성된다. 이 모든 과정은 자동화된 Python 스크립트를 통해 수행되며, 일일 단위로 갱신되어 데이터를 효율적으로 반영할 수 있도록 설계했다.

○ 2단계: 데이터 베이스 구축

정제된 데이터는 PostgreSQL 관계형 데이터베이스에 저장한다. PostgreSQL은 안정적인 데이터 관리와 빠른 쿼리 수행능력을 제공하므로 GKG 데이터와 같이 대용량 데이터를 효율적으로 처리하는 데 적합하다. 또한, 안정적인 데이터 관리로 선거 예측치의 신뢰성을 높일 수 있다.

○ 3단계: 자동화 관리

전체 데이터의 자동화 과정은 Apache Airflow를 사용하여 자동화된다. Airflow는 각 단계에서 수행하는 작업의 실행 순서를 정의하고, 모든 작업이 매일 성공적으로 실행될 수 있도록 관리하는 기능을 제공한다. 이를 통해 매일 자동으로 데이터가 추출, 정제, 저장되며, 작업 실패 시 이메일 알림 기능을 통해 관리자가 이를 실시간으로 확인하고 오류를 수정할 수 있도록 설계했다.

○ 4단계: 데이터 시각화

최종적으로 처리된 데이터는 Django 프레임워크 기반 웹 애플리케이션을 통해 시각화된다. 이 과정에서는 plotly.js라는 Javascript 기반의 웹 시각화 라이브러리를 사용하여 대화형 그래프와 차트를 구현함으로써, 사용자가 데이터를 보다 직관적으로 이해할 수 있도록 구현했다.

○ 5단계: 미국 대통령선거 예측치 웹 페이지

본 연구에서 개발한 미국 대통령선거 예측치는 경북대학교 사회과학기초자료연구소 웹 페이지(<http://www.basicdata.kr/>)에 제공되며, 일일 단위로 업데이트가 된다. 다만, 주 단위에서의 예측치 추이 및 정책 보도 비중은 웹 페이지 제약 상 별도로 제공한다.

V. 결론

본 연구의 목적은 미국 대통령선거를 예측하고, 후보의 주요 정책 보도 비중을 파악하는 데 있다. 이를 통해 향후 한국 외교정책을 포함한 대미 정책 수립에 기여하고자 한다. 특히, 본 연구는 일일 단위로 선거 예측과 정책 보도 비중을 파악하여 학계 등 관련 분야와 공유하고자 한다.

본 연구는 미국 대통령선거 이후에도 의의가 있다고 본다. 예를 들어, 전국 및 주 차원에서 당선자의 언론 논조를 시공간적으로 파악함으로써, 유권자에 의한 대통령 지지율과 유사하게 언론이 바라본 대통령 지지율로도 활용할 수 있으리라 본다. 또한, 전국 및 주 차원에서 당선자의 정책 보도 비중을 살펴봄으로써, 전국 차원에서 그리고 주 차원에서 당선자가 주목하고 있는 정책을 언론의 눈을 통해 들여다볼 수 있을 것이다.

또한, 본 연구는 향후 한국 선거에도 활용할 수 있으리라 본다. 미국 대통령선거 예측과 같이, 한국 대통령선거 결과 예측과 더불어 후보의 정책 보도 비중을 파악하는 데 활용될 수 있으리라 본다. 또한, 미국 경우와 같이 당선자의 언론 지지율로도 활용할 수 있으리라 본다.

부록: World Bank 분류체계 최상위 개념의 Level 2¹⁰⁾

<p>Agriculture and Food Security(435)</p> <ul style="list-style-type: none"> • 농업 정책 • 농업 위험 및 보안 • 생물학적 자산 회계 • 생물 보안 • 기후 스마트 농업 • 상업적 농업 • 농부와 시장 연결 • 자연 자원 • 농촌 개발 • 추적 가능성 	<p>Macroeconomic and Structural Policies(439)</p> <ul style="list-style-type: none"> • 재정정책 • 거시경제 모델링 및 통계 • 거시경제 모니터링 • 거시건전성 정책 • 통화정책 • 국채관리 • 구조정책 및 개혁
<p>Climate Change(567)</p> <ul style="list-style-type: none"> • 기후변화 적응 • 기후변화법 • 기후변화 완화 • 기후금융 • 기후회복력 개발 • 기후과학 • 기후서비스 	<p>Macroeconomic Vulnerability and Debt(1104)</p> <ul style="list-style-type: none"> • 원조 의존 • 국제수지 • 부채와 거시경제 안정성 • 디플레이션 • 네덜란드병 • 경제 충격과 취약성 • 금융 부문 불안정성 • 금융 부문 부채 • 재정 균형 • 재정 정책과 불평등 • 재정 위험 • 인플레이션 • 거시 취약성 • 거시경제 불균형과 조정 • 분쟁과 취약성의 거시경제학 • 물가 변동성 • 소프트 버짓 제약
<p>Economic Growth(471)</p> <ul style="list-style-type: none"> • 경제성장의 결정요인 • 경제성장 분석 • 경제성장 진단 • 경제성장 정책 • 포용적 성장 • 혁신과 성장 • 공간적 성장 • 지속가능한 경제성장 	<p>Poverty(695)</p> <ul style="list-style-type: none"> • 빈곤 감소 및 형평성을 위한 행동 이니셔티브 • 증거 기반 정책 • 포용적 성장 • 불평등과 공유 번영
<p>Education(470)</p> <ul style="list-style-type: none"> • 교육의 경제학 • 모두를 위한 교육 • 교육 관리 및 행정 • 교육 수준 	

10) 분류체계 중 최상위 24개 개념의 Level 2 개념들은 World Bank 주제 분류체계 (<https://vocabulary.worldbank.org/taxonomy.html>, 방문일: 2023년 12월 30일)에 제시된 하위 개념(Narrower Concepts)이다. 각각의 하위 개념은 다음 수준의 하위 개념(Level 3)으로 이어지며, 그 다음의 하위 개념도 같은 방식으로 분류된다.

<p>Energy and Extractives(507)</p> <ul style="list-style-type: none"> • 탄소 포집 및 저장 • 에너지 접근 • 에너지 효율 • 에너지 금융 • 에너지 저장 • 추출 산업 및 교차 주제 • 광업 부문 • 광업 시스템 • 비재생 에너지 • 석유 및 가스 산업 • 석유 및 가스 시스템 • 전력 및 전기 부문 • 전력 시스템 • 재생 에너지 	<ul style="list-style-type: none"> • 일자리와 빈곤 • 빈곤 감소 및 공유 번영을 위한 시장과 제도 • 빈곤 측정 및 분석 • 충격과 빈곤에 대한 취약성 <p>Private Sector Development(1921)</p> <ul style="list-style-type: none"> • 경쟁 정책 • 경쟁력 있는 산업 • 혁신, 기술 및 기업가 정신 • 투자 환경 • 하위 국가 투자 • 상류 활동
<p>Environment, Natural Resources and Blue Economy(566)</p> <ul style="list-style-type: none"> • 보전 금융 • 생태계 관리 • 생태계 서비스 • 생태계 • 환경 공학 • 환경 관리 • 환경 보호 • 환경 지속 가능성 • 녹색 성장 • 자연 자원 관리 • 오염 관리 	<p>Public-Private Partnerships(2206)</p> <ul style="list-style-type: none"> • PPP 계약 관리 • PPP 재정 관리 • PPP 거버넌스 및 제도 • PPP 정책 및 전략 • PPP 조달 • PPP 프로젝트 선정 • PPP를 위한 민간 부문 개발 • 테마 기반 PPP <p>Public Sector Management(696)</p> <ul style="list-style-type: none"> • 공공부문의 거버넌스 • 정의 • 법적 및 규제적 틀 • 공공행정 • 공공재정
<p>Financial Sector Development(1920)</p> <ul style="list-style-type: none"> • 자본시장 • 금융 아키텍처 및 은행 • 금융 포용 • 금융 인프라 및 송금 • 금융 기관 • 금융 무결성 • 금융 기술 • 국제 금융 표준 	<p>Social Development(134)</p> <ul style="list-style-type: none"> • 지역사회 주도 개발 • 문화유산 • 사회적 책임 • 사회 분석 • 사회 통합 • 사회 포용 • 사회적 회복력과 기후변화 • 사회적 보호

<p>Fragility, Conflict and Violence(2432)</p> <ul style="list-style-type: none"> • 분쟁과 폭력 • 위기 위험 관리 • 인권 • 평화 작전 및 분쟁 관리 <p>Gender(615)</p> <ul style="list-style-type: none"> • 성별과 기후 변화 • 성별과 경제적 권한 부여 • 성별과 인간 개발 • 성별과 인프라 • 성별과 거시 경제 정책 • 성별과 농촌 개발 • 성별과 도시 개발 • 성별에 기반한 폭력 • 성별 모니터링 및 평가 • 글로벌 성별 정책 프레임워크 • 목소리와 대리인 <p>Information and Communication Technologies(133)</p> <ul style="list-style-type: none"> • 디지털 정부 • 일자리를 위한 ICT • ICT 혁신과 변혁 • 통신 및 브로드밴드 접속 • 통신, 미디어 및 기술 <p>Health, Nutrition and Population(621)</p> <ul style="list-style-type: none"> • 건강 • 영양 • 인구, 고령화 및 건강 • 공중 보건 • 출생, 모성 및 아동 건강 <p>Jobs(2670)</p> <ul style="list-style-type: none"> • 포용적 일자리 • 일자리와 발전 • 일자리 진단 • 일자리 전략 	<p>Social Protection and Labor(697)</p> <ul style="list-style-type: none"> • 노동 시장 • 사회 지원 • 사회 보험 • 사회 보호 및 노동 전달 시스템 • 사회 보호 및 노동 시스템 <p>Trade(698)</p> <ul style="list-style-type: none"> • 무역 경쟁력 및 다각화 • 무역 원활화 및 물류 • 무역 정책 및 통합 <p>Urban Development(699)</p> <ul style="list-style-type: none"> • 도시 이니셔티브 • 경쟁력 있는 도시 • 문화유산과 지속 가능한 관광 • 재난 위험 관리 • 결과 중심적 접근을 위한 글로벌 파트너십 • 녹색 도시 • 포용적인 도시 • 국가 도시 정책 • 회복력 있는 도시 • 도시 거버넌스와 도시 시스템 • 도시 고형 폐기물 관리 및 청소 서비스 • 도시 교통 • 도시 물 • 도시화와 성장 <p>Water(137)</p> <ul style="list-style-type: none"> • 농업용수 관리 • 에너지와 물 • 물 할당과 물 경제학 • 물 안보와 통합 자원 관리 • 물 공급과 위생 <p>Transport(135)</p> <ul style="list-style-type: none"> • 신형 교통 기술 • 교통 수단
--	---

	<ul style="list-style-type: none">• 농촌 교통• 교통 및 물류 서비스• 교통 및 지속 가능한 개발• 교통 경제학• 교통 장비• 교통 거버넌스• 교통 인프라• 도시 교통
--	--

참고문헌

- 김규철. 2022. “한국 국민의 통일의식은 남북관계에 영향을 받는가?” 『통일과 평화』 14집 2호: 39-78.
- 이성우. 2017. “남북한의 상호갈등 및 협력관계의 변화: 빅데이터를 이용한 시계열 분석 1979-2013.” 『분쟁해결연구』 15권 3호: 111-144.
- 이성우·정성희. 2020. 「빅데이터 기반 GDELT 활용 남북관계 시계열 분석: 협력과 분쟁의 역학」 경기연구원 기본연구 2020-05.
- 이성우·임재형. 2021. “남북한 간의 상호협력과 갈등의 역학: 2000-2020.” 『분쟁 해결연구』 19권 2호: 67-99.
- 이성우. 2022a. “사건계수자료(GDELT)를 활용한 한·중 양자관계 역학 분석.” 『아태연구』 29권 3호: 37-75.
- 이성우. 2022b. “GDELT를 활용한 한반도 평화지수 개발: 남북관계의 실증적 분석을 위하여.” 『국제지역연구』 26권 3호: 119-142.
- Leetaru, Kalev, and Philip A. Schrod. 2013. “GDELT: Global data on events, location, and tone, 1979-2012.” Paper presented at the ISA Annual Convention, San Francisco, April 2013.
- World Bank. 2016. “Theme Taxonomy and definitions.” Revised July 1, 2016 at <https://pubdocs.worldbank.org/en/275841490966525495/New-Theme-Taxonomy-and-definitions-revised-July-012016.pdf>.