KISDI Perspectives

동향 March 2025 No.1

ICT 뉴스경기지수 구축방법 및 주요결과 소개

손녕선 연구위원
정보통신정책연구원 ICT통계정보연구실

KISDI 정보통신정책연구원
KOREA INFORMATION SOCIETY DEVELOPMENT INSTITUTE

KISDI Perspectives

동향

2025.03.26

ICT 뉴스경기지수 구축방법 및 주요결과 소개[1)]

손녕선 연구위원
정보통신정책연구원 ICT통계정보연구실, nsson@kisdi.re.kr

요 약

- 방대한 비정형 텍스트와 자연어처리(NLP) 기술의 발전에 따라 텍스트 자료를 활용한 분석 방법이 주목받고 있음
- 이러한 가능성에 착안하여, 뉴스 기사를 통해 ICT 산업의 동향 및 주요 트렌드를 살펴볼 수 있는 ICT 뉴스경기지수를 구축하였음
- ICT 뉴스경기지수 구축을 위해 ICT 관련 뉴스를 정의 및 수집하고, KoBERT 모델을 사용하여 감성분석을 수행
- 감성분석 결과를 표준화하여 월별 ICT 뉴스경기지수를 구축하고, 뉴스기사 제목을 활용한 네트워크 분석을 수행하여 주요 결과를 ITSTAT에 공개
- ICT 뉴스경기지수의 활용은 ICT 경기를 사전적으로 측정하고 트렌드의 변화를 분석할 수 있다는 점에서 가치를 지님
  - ICT 뉴스경기지수의 결과는 ICT 산업의 생산 및 수출값의 움직임과 유사한 추세를 보이며, BSI에 선행하는 경향을 나타냄
  - 2025년 1~2월 네트워크 분석을 통해 ICT 관련 뉴스에서 AI가 핵심 키워드에 위치하였으며, 딥시크의 파급력이 크다는 점을 확인

* 본고의 내용은 "ICT 주요이슈 분석체계 구축"(손녕선 외, 2024)에서 수행한 "ICT 뉴스경기지수 구축" 연구 결과를 요약 정리한 것이다. 분석 결과는 2025년 자료를 추가하여 작성하였다.

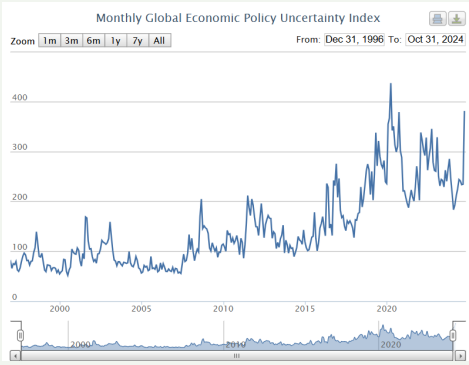
01 개요

배경

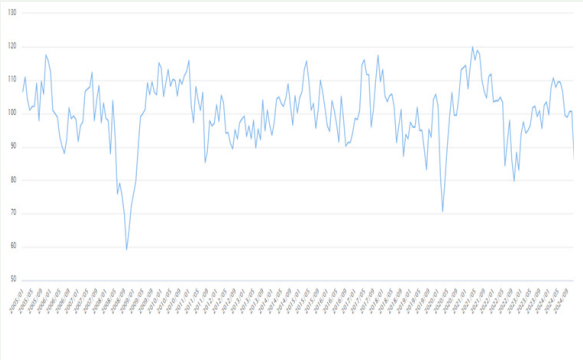
- 비정형 텍스트에 대한 관심이 높아지고 자연어처리(NLP) 기술이 발달함에 따라, 텍스트 자료가 다양한 분석의 도구로 주목받고 있음
  - 특히 뉴스기사에 대한 관심도가 높은데, 온라인 플랫폼에서 수집 가능하고 데이터의 양이 방대하다는 점과 다양한 주제를 포괄할 수 있다는 것이 장점
  - 뉴스 기사는 해당 시점의 트렌드와 해당 주제에 대한 사회적 반응을 담고 있기에 숫자로 표현할 수 없는 정보들을 제공
- 뉴스 기사의 특징을 활용하여 분석 및 전망을 시도한 다양한 연구가 존재
  - Baker et al.(2016)에서는 경제 정책과 관련된 기사를 수집하여 정책의 불확실성 정도를 분석
  - Shapiro et al.(2022)는 뉴스 기사를 활용한 감정 분석이 경제적 의사결정에 미치는 영향을 살펴보았고, 이 과정에서 감정 분석 모델의 개선 방향을 제시
  - Scott and Varian(2014)는 실업률과 물가의 단기 변동을 측정하는데 있어 뉴스 기사를 활용
  - 한국은행은 2021년부터 뉴스 기사에 기반한 뉴스심리지수를 개발하여 발표하고 있으며, IMF와 각 국의 중앙은행도 뉴스 기사를 활용하는 방안을 연구

그림 1 뉴스기사 활용 연구 예시

월별 미국 경제 정책 불확실성 지수



한국은행 뉴스심리지수



출처: https://www.policyuncertainty.com/(최종접속일: 25.1.8), 한국은행 경제통계시스템(https://ecos.bok.or.kr/)(최종접속일: 25.1.8)

- 뉴스 기사를 활용한 다양한 시도를 고려하면, 디지털 전환의 핵심이라 할 수 있는 ICT 산업의 동향을 시의성 있게 측정할 수 있는 지표를 개발하는 것은 의미가 클 것으로 판단됨
  - ICT 산업은 빠른 혁신과 융합이 일어나는 대표적 산업이며 그만큼 트렌드나 상황도 빠르게 변화하는 것이 특징
  - 전통적인 통계조사 방식은 신뢰성 높은 결과를 제공하지만, 조사에 걸리는 시간등의 이슈로 시의성 측면에서는 한계가 존재
  - ICT 산업의 빠른 변화를 신속하게 추적하여 동향 및 이슈를 파악할 수 있도록 돕는 지표가 필요
  - 뉴스 기사는 빠른 자료 제공 및 광범위한 주제 포괄이 장점인 만큼, 이러한 필요성을 충족시켜줄 수 있는 도구가 될 수 있음
- 본고에서는 뉴스 기사를 통해 ICT 산업의 동향과 주요 트렌드를 빠르게 파악할 수 있도록 돕는 ICT 뉴스경기지수를 소개하고 주요 결과를 설명

02 ICT 뉴스경기지수 구축 방법 소개

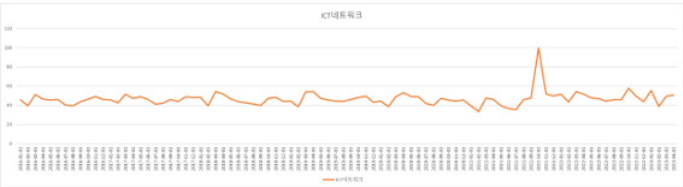
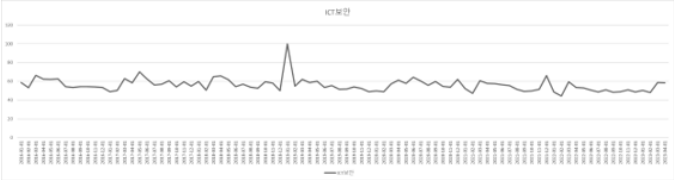
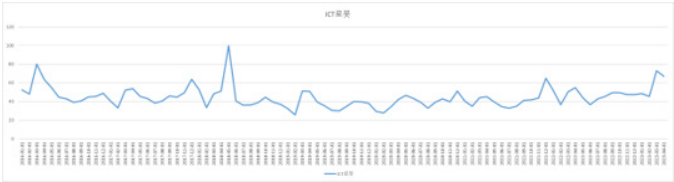
키워드 선정 및 뉴스 수집

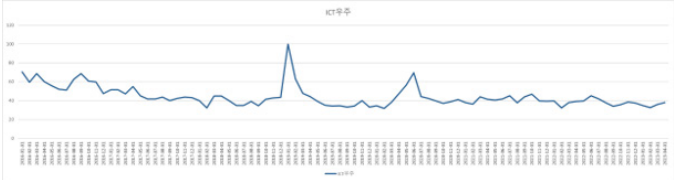
- ICT 뉴스경기지수를 만들기 위해 먼저 ICT 관련 뉴스를 정의하고 이를 수집할 필요가 있음
  - ICT 산업과 관련된 동향을 파악하는 것이 목적이기에 ICT 산업을 대표하는 다양한 키워드들을 뉴스 수집 과정에서 활용할 필요가 있음
  - 2가지 측면에서 키워드를 검토하였으며 ① ICT 분야를 큰 틀에서 상징하는 키워드 ② 최신 트렌드를 검토하여 세부적인 기술이나 영역을 상징할 수 있는 키워드를 선정하였음
- ICT 분야를 대표하는 키워드로 8개를 선정
  - 8개 키워드는 ICT 산업, ICT 기업, ICT 수출, ICT 품목, IT 산업, IT 기업, IT 수출, IT 품목을 포함
  - 이 단어들을 통해 수집한 뉴스 기사는 ICT의 전반적인 소식을 포함하며, 거시적 관점에서 ICT 산업의 흐름을 알려주는 자료로 활용
- 세부적인 기술 및 영역을 포함하기 위한 키워드를 선정하기 위해 참고문헌에 기반한 후보군을 도출
  - 가트너, 과학기술정보통신부 등 공신력 있는 기관 및 기업에서 발표한 ICT 트렌드 관련 보고서를 검토
  - 조사를 통해 공통적으로 도출되는 주요 키워드들을 선정하여 후보군으로 구성

- 후보군을 대상으로 2016년 1월 – 2023년 4월까지의 검색량 평균과 분산을 분석하고, 이를 기준으로 10개의 신규 키워드를 선정
  - 구글트렌드와 네이버 데이터랩에서 해당 키워드들의 검색량을 분석하여 조건에 맞는 키워드들을 ① 과거부터 지속적으로 언급된 ICT 키워드 ② 최근에 등장한 ICT 키워드 2개의 그룹으로 분류
  - 과거부터 지속적으로 언급된 ICT 키워드는 분석 기간 평균 검색량이 일정 수준*을 유지하는 키워드들을 선정

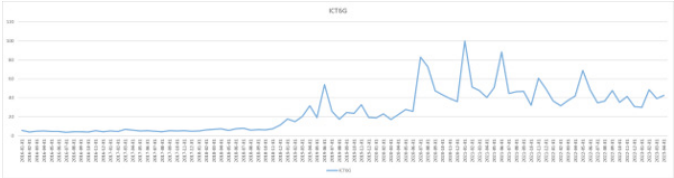
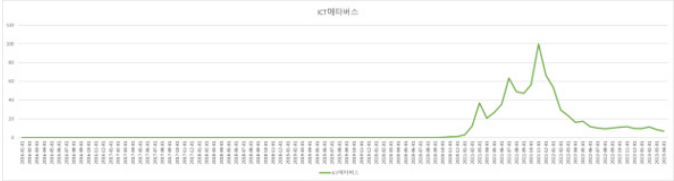
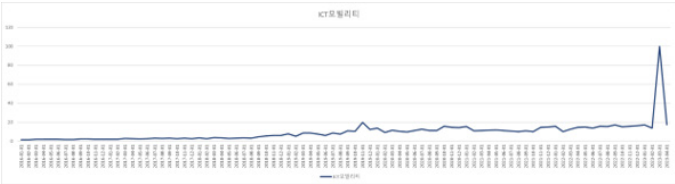
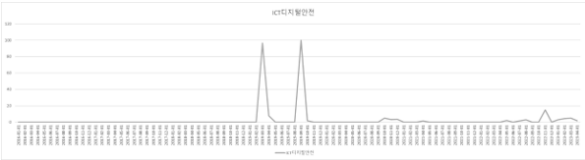
    * 꾸준하게 등장하는 키워드의 핵심 특징을 일정 수준의 검색량 유지와 낮은 변동성으로 판단하였으며, 검색량 수준을 비교하기 위한 기준으로 해당 기간 "ICT" 키워드의 검색량을 참고하였음
  - 최근에 등장한 ICT 키워드는 2010년대 대비 2020년에 검색량이 빠르게 증가하고 상대적으로 분산이 높은 키워드들을 선정
  - 이상의 작업을 거쳐 과거부터 지속적으로 언급된 키워드에는 ICT 네트워크, ICT 보안, ICT 로봇, ICT 우주, ICT 플랫폼을 선정
  - 최근 등장한 키워드에는 ICT 6G, ICT 메타버스, ICT 모빌리티, ICT 플랫폼 엔지니어링, ICT 디지털안전을 선정

| 표 1 | 키워드별 검색 트렌드: 과거부터 지속적으로 언급된 ICT 키워드

키워드명	검색 트렌드	평균	분산
ICT 네트워크		46.6	54.8
ICT 보안		56.1	49.9
ICT 로봇		44.6	126.1

키워드명	검색 트렌드	평균	분산
ICT 우주		44.3	122.2
ICT 플랫폼		39.5	135.7

| 표 2 | 키워드별 검색 트렌드: 최근에 등장한 ICT 키워드

키워드명	검색 트렌드	평균	분산
ICT 6G		25.7	484.8
ICT 메타버스		8.8	338.7
ICT 모빌리티		9.4	123.4
ICT 플랫폼 엔지니어링		3.7	239.5
ICT 디지털안전		2.9	218.0

- 최종적으로 선정한 18개의 키워드를 활용, 2008년 1월 1일부터 2025년 2월 28일까지의 뉴스기사를 수집하여 활용

| 표 3 | 키워드별 검색 트렌드: 최근에 등장한 ICT 키워드

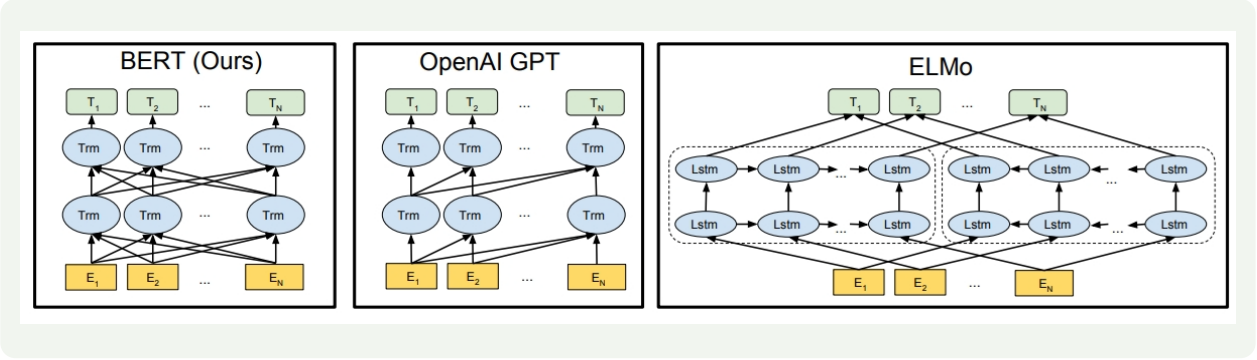
ICT 대표 키워드	ICT 산업
	ICT 기업
	ICT 수출
	ICT 품목
	IT 산업
	IT 기업
	IT 수출
	IT 품목
과거부터 꾸준히 언급된 ICT 키워드	ICT 네트워크
	ICT 보안
	ICT 로봇
	ICT 우주
	ICT 플랫폼
최근에 등장한 ICT 키워드	ICT 6G
	ICT 메타버스
	ICT 모빌리티
	ICT 플랫폼엔지니어링
	ICT 디지털안전

| 모델 구축

- ICT 뉴스경기지수는 수집된 뉴스 기사에 대한 감성 분석을 실시하고 이 결과를 바탕으로 ICT 산업의 동향을 보여주는 것을 목적으로 함
  - 감성 분석은 기사에서 등장하는 긍정 / 부정 단어의 빈도를 측정하는 것을 의미
  - 긍정 기사가 많은 기간에는 ICT 뉴스경기지수 값이 상승하게 되며, 이는 해당 기간 ICT 산업과 관련된 동향이 좋았음을 의미한다고 볼 수 있음
  - 반대로 부정 기사가 많은 기간에는 값이 하락하고, 해당 기간 ICT 산업과 관련된 환경에 부정적인 요소가 많았음을 추측 가능
- 뉴스 기사에 대한 감성 분석을 위해 KoBERT 모델을 활용
  - KoBERT는 SK 텔레콤에서 발표한 모델로 BERT에 기반을 두고 있음
  - BERT는 구글에서 2018년에 발표한 자연어처리(NLP) 방법의 한 종류이며, 단어 확률 예측 과정에서 양방향을 사용하여 성능을 개선한 것이 특징

- 뛰어난 성능으로 다양한 자연어처리 과정에서 BERT가 활용되었으나, 한국어에 대해서는 상대적으로 성능이 떨어지는 한계점이 존재
- 이러한 제약을 극복하기 위해, SK 텔레콤에서 수백만 건의 한국의 문장을 학습시켜서 개선한 것이 KoBERT 모델

| 그림 2 | BERT와 다른 자연어 처리 모델 비교



자료: 인공지능 신문, '인공지능(AI) 언어모델 BERT(버트)는 무엇인가'(https://www.aitimes.kr/news/articleView.html?idxno=13117)

- KoBERT를 활용하여 뉴스 기사의 긍정/부정을 판단하기 위해 사전 기반 라벨링을 우선적으로 실시
  - 사전 기반 라벨링은 단어를 긍정, 부정, 중립으로 미리 분류를 해둔 단어 사전(lexicon)을 사용하여 문장의 감성을 결정하는 것을 의미
  - 단어 사전 정보에 기반하여 해당 문장에 긍정 단어가 많다면 이를 긍정 문장으로, 부정 단어가 많다면 이를 부정 문장으로 분류
  - 사전 기반 분류 작업을 통해 학습용 뉴스 기사에 긍정/부정 라벨링을 달고, 이를 KoBERT에 학습시킨 후 학습된 모델을 ICT 뉴스경기지수 구축에 활용
- 사전 기반 라벨링 작업을 위해 한국은행의 긍부정 키워드, KNU 감성사전, 썸트렌드 등의 자료를 활용
  - 한국은행에서 수행한 감성사전 구축과 관련된 연구들에서 제시한 키워드들을 참고(김현중 외, 2019, 서범석 외, 2022, 전종준 외, 2020)
  - KNU 감성사전은 인간의 일반적인 감정 표현을 나타내는 긍정어 및 부정어 조합을 제공
  - 썸트렌드에서 제공하는 ICT 산업, ICT 기업, ICT 수출 관련 기사들의 긍정/부정 단어들을 활용
  - 기업 재무분석에 강점이 있는 감성 사전인 KOSELF도 참고하였음(조수지 외 2021)

- 단어 사전과 수집한 ICT 뉴스 기사를 통해 약 400만개의 학습 데이터를 구축하였고 이를 통해 감성 분석을 위한 KoBERT 학습을 진행
  - Training set 70%, Test set 30%의 비율로 작업을 진행하였음
  - 학습 결과 Training set에서는 98.6%, Tese set에서는 98.2%의 정확도를 보였으며 Accuracy Score는 98.3%를 기록하여 전반적으로 모델의 예측도가 높은 것을 확인

지수 산출 및 네트워크 분석 방법

- 학습한 KoBERT를 사용하여 수집한 ICT 뉴스 기사의 긍부정 분류를 진행
  - 기사의 본문을 문장 단위로 나누었으며, 각각의 문장에 대한 긍정/부정 라벨링을 진행
  - 기간 내 수집된 기사들의 긍정/부정 문장의 개수에 기반하여 지수값을 산출
- 산출을 위한 계산식은 한국은행의 뉴스심리지수(서범석 외(2022))를 참고하였으며 식은 다음과 같음

$$\text{뉴스경기지수} = \left(\frac{X-\overline{X}}{S}\right) \times 10 + 100$$

$$X = \frac{\text{기간내긍정문장수평균} - \text{기간내부정문장수평균}}{\text{기간내긍정문장수평균} + \text{기간내부정문장수평균}}$$

$$\overline{X} = \text{표준화구간평균}, S = \text{표준화구간표준편차}$$

  - 날짜별로 뉴스기사를 수집 후 KoBERT를 통해 긍정/부정을 분류하여, 해당 날짜의 긍정 문장 수와 부정 문장 수를 계산
  - 이후 해당 날짜를 포함한 과거 7일동안의 긍정 문장 수와 부정 문장 수를 평균하여 계산하고, 이를 표준화함*

    * 표준화구간은 2008년 1월 1일부터 2023년 4월 30일이며, 이 기간의 ICT 뉴스경기지수 평균값과 표준편차를 사용
  - 이렇게 계산한 ICT 뉴스경기지수는 100을 기준으로 하여 100보다 큰 경우에는 긍정적인 기사가, 100보다 작은 경우 부정적인 기사가 많았음을 의미
- 일별로 계산한 ICT 뉴스경기지수 값을 평균하여 월별 ICT 뉴스경기지수를 산출
  - 일별 단위 지수값은 해당 날짜 뉴스 기사의 수에 따른 변동성이 큰 편으로, 안정적인 결과 산출에 제약이 있음
  - 지수의 안정성을 높이기 위해 월별 값을 주요 결과로 작성

- 본고에서는 2015년 1월 ~ 2025년 2월 기간의 자료를 사용하여 ICT 뉴스경기지수를 구축한 결과를 소개
  - 2015년 이전 기간의 경우, ICT와 관련된 뉴스기사 수가 충분하지 않아 일별 지수의 변동성이 큰 편이며 이로 인해 월별 측정에도 제약이 존재
  - 기사 수가 일정 수준 이상으로 확보되는 2015년 이후부터의 기간을 활용
- ICT 뉴스기사에서 등장하는 주요 단어를 살펴보고 트렌드를 추적하기 위해 네트워크 분석을 시행
  - 월별로 뉴스기사의 제목을 활용하여 네트워크 분석을 실시하여 해당 월의 트렌드를 확인
  - 네트워크에서 노드(단어)의 중요도를 측정하기 위해 구글에서 발표한 PageRank를 활용
  - PageRank는 특정 노드가 다른 노드로부터 받는 링크가 많아질수록 중요도가 높다고 판단
    ※ 특히 중요도가 높은 노드로부터 링크를 받는 횟수가 많을수록 중요도는 더욱 증가
  - PageRank 기반으로 상위 60개 단어를 선정하여 네트워크 결과를 시각화
  - 본고에서는 최근 시점인 2025년 1월~2월 기간의 네트워크 분석 결과를 제시

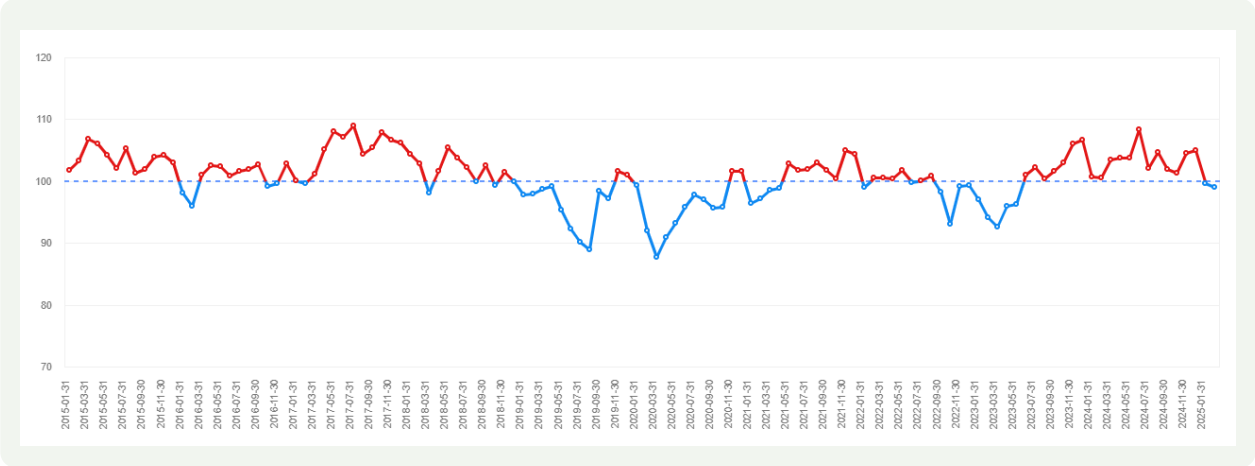
03 ICT 뉴스경기지수 구축 결과

1 뉴스경기지수

- ICT 뉴스경기지수로 살펴본 주요 결과는 다음과 같이 요약할 수 있음
  - 2015~2018년 기간에는 전반적으로 100 이상의 지수값을 가진 월이 많아, ICT와 관련하여 긍정적인 뉴스가 많았음을 확인
  - 2019년부터는 100 미만의 지수값이 빈번하게 등장하여, 부정적 이슈가 늘어났다는 것을 알 수 있음
  - 2021년 이후 상승 추세로 전환하였으나 2022년 9월 이후에 다시 부정적 이슈가 증가했음을 파악할 수 있음
  - 2023년 6월 이후에는 꾸준하게 지수값이 높은 상태를 유지하였으나, 2025년 들어 다시 하향 추세로 전환했음을 확인 가능

- ICT 산업의 생산과 수출 값을 보면 실제로 2015~2018년에 증가한 반면 2019년과 2020년 들어 역성장이 발생했다는 점[1], ICT 뉴스경기지수가 BSI에 선행하는 모습을 확인(손녕선 외(2023))하였다는 점 등을 고려하면 ICT 뉴스경기지수가 경기 측정에 있어 가치 있는 함의를 제공하는 것으로 판단

| 그림 3 | ICT 뉴스경기지수



주: 월별로 구축한 ICT 뉴스경기지수 결과. 100보다 큰 경우(붉은색) 기사에 긍정적인 내용이 많았음을 의미하며, 100보다 작은 경우(파란색) 기사에 부정적인 내용이 많았음을 의미
출처: ITSTAT(https://www.itstat.go.kr/itstat/kor/economic/economicList.html?menuId=2010136&rootId=2010003)

| 표 4 | ICT 뉴스경기지수와 BSI 간 상관관계

	ICT BSI	전산업 BSI
동기	33.3%	45.2%
1개월 선행	39.5%	51.3%
2개월 선행	34.8%	47.6%
3개월 선행	25.2%	40.3%

주: 2015년 1월 ~ 2023년 4월까지 자료를 활용하여 ICT 뉴스경기지수의 선행값과 각 BSI와의 상관관계를 비교
출처: 손녕선 외(2023)

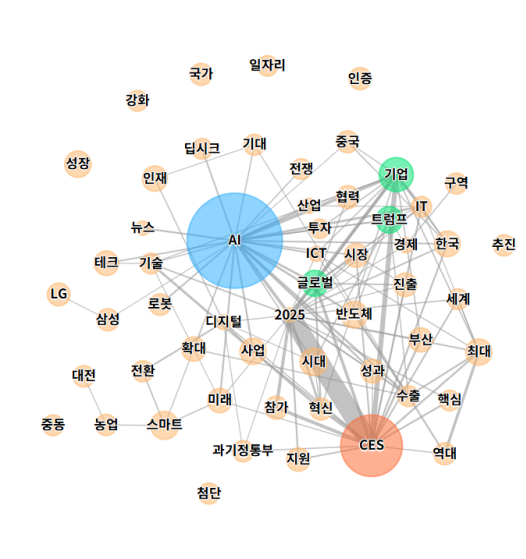
1 ICT 생산액은 과학기술정보통신부의 ICT 실태조사, ICT 수출액은 과학기술정보통신부·정보통신기획평가원의 ICT 수출입동향을 참고

네트워크 분석

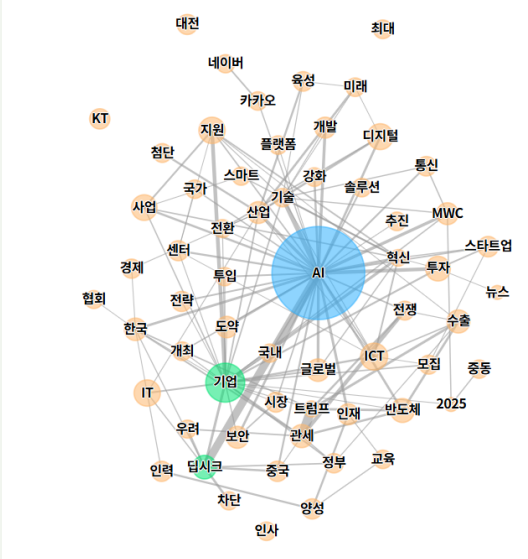
- 네트워크 분석 결과를 요약하면 다음과 같음
  - 1월과 2월 모두 AI가 핵심 키워드로 선정되어, ICT 내에서 AI 기술에 대한 관심도가 매우 높다는 사실을 확인 가능
  - 1월에 개최된 CES도 많은 주목을 받았으며, 그 외에 트럼프, 글로벌, 기업 등이 1월의 주요 키워드로 선정되었음
  - AI 외에도 반도체, 로봇 등의 키워드가 등장하였으며 반도체는 수출과 연관성이 큰 것으로 나타나 ICT에서 반도체 수출에 대한 관심이 높다는 점을 확인하였음
  - 흥미로운 키워드는 딥시크로, 1월에 등장 후 2월에는 주요 키워드로 입지가 상승하였음
  - 이는 딥시크의 등장이 생성형 인공지능 시장에 준 커다란 충격을 보여주는 것으로 풀이됨
  - 2월에 카카오, 네이버 등이 상위 키워드로 떠오른 것, 딥시크와 관련해서 차단, 우려, 보안 등이 관련된 것도 딥시크의 부상에 따른 영향인 것으로 풀이
  - 그 외에 3월에 개최된 MWC에 대해서도 사전적으로 많은 관심이 쏠린 것을 확인 가능

그림 4 네트워크 분석

2025년 1월



2025년 2월



참고문헌

과학기술정보통신부, ICT 실태조사

과학기술정보통신부 · 정보통신기획평가원, ICT 수출입동향

김현중, 임종호, 이혜영, 이상호(2019), 『온라인 뉴스 기사를 활용한 경제심리보조지수 개발』, 국민계정리뷰, 2019년 제2호, 한국은행.

서범석, 이영환, 조형배(2022), 『기계학습을 이용한 뉴스심리지수(NSI)의 작성과 활용』, 국민계정리뷰, 2022년 제1호, 한국은행.

손녕선·이은영·진정민·김재민·박재홍(2023), "ICT 주요이슈 분석체계 구축", 정책연구 23-28, 정보통신정책연구원.

손녕선·서영선·진정민·이은영·신현호(2024), "ICT 주요이슈 분석체계 구축", 정책연구 24-23, 정보통신정책연구원.

전종준, 안승환, 이문희, 황희진(2020), 『경제용어 감성 사전 구축방안 연구』, 국민계정리뷰, 2020년 제3호, 한국은행.

조수지, 김흥규, 양철원(2021), "기업 재무분석을 위한 한국어 감성사전 구축." 한국증권학회지 50.2, 135-170

Baker, S. R., Bloom, N., and Davis, S. J. (2016). "Measuring economic policy uncertainty." The Quarterly Journal of Economics, 131(4), 1593-1636.

Scott, S. L. and Varian, H. R. (2014). "Predicting the present with bayesian structural time series." International Journal of Mathematical Modelling and Numerical Optimisation, 5(1-2), 4-23.

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. Journal of econometrics, 228(2), 221-243.

《인공지능신문》(2019. 1. 3.), '인공지능(AI) 언어모델 BERT(버트)는 무엇인가'.

[웹사이트]

네이버 Datalab, https://datalab.naver.com/keyword/trendSearch.naver.

Google Trend, https://trends.google.co.kr/trends/

썸트렌드, https://some.co.kr/

Economic Policy Uncertainty, https://www.policyuncertainty.com/

한국은행 경제통계시스템, https://ecos.bok.or.kr/

ITSTAT, https://www.itstat.go.kr/

KISDI Perspectives 발간 내역

KISDI PERSPECTIVES는 국내 외 정보통신방송 관련 주요 정책 및 시장 동향을 분석한 리포트입니다.

문의 : 노희윤 전문연구원 (정보통신정책연구원 방송미디어연구본부, hyoooon@kisdi.re.kr, 043-531-4042)