

LDA 주제 모델링을 활용한 SKB와 POSIVA의 심층 처분 연구 동향 비교 분석

최민환¹ · 채병곤² · 최정해^{3*}

¹경북대학교 지구과학교육과 학부과정,

²한국지질자원연구원 국토안전연구본부 방폐물심층처분연구센터 박사, ³경북대학교 지구과학교육과 교수

Comparative Trend Analysis of Deep Disposal Research by SKB and POSIVA Using LDA Topic Modeling

Minhwan Choi¹ · Byung-Gon Chae² · Junghae Choi^{3*}

¹Student, Department of Earth Science Education, Kyungpook National University

²Principal Researcher, Deep Geological Disposal & Energy Storage Research Center, Geological Safety Division, Korea Institute of Geoscience and Mineral Resources

³Professor, Department of Earth Science Education, Kyungpook National University

Abstract

This study quantitatively analyzes the long-term research and development (R&D) trends of two leading companies in deep geological disposal, Sweden's SKB and Finland's POSIVA, using latent Dirichlet allocation (LDA) topic modeling combined with structural break analysis. The dataset consists of 2,520 technical report titles from SKB and 1,014 from POSIVA. The LDA results identify nine major R&D topics for SKB and six for POSIVA. This indicates a relatively diversified research portfolio at SKB (top three topics: 49.6% of the total publications) and a more concentrated portfolio at POSIVA (top three topics: 72.2%). Regression analysis of annual topic proportions (i.e., proportion of the total publications) and Wald tests reveal statistically significant ($p < 0.05$) structural breaks around the years when the firms applied for a construction license (2011 for SKB and 2012 for POSIVA) in five of nine topics for SKB and three of six topics for POSIVA. In particular, hydrogeological modeling R&D undertaken by SKB shows accelerated growth after the license application, whereas underground site characterization R&D undertaken by POSIVA shows a declining trend, indicating contrasting transitions in research focus. In addition, the average inter-topic similarity is lower for SKB (0.12) than for POSIVA (0.34), reflecting differences in the degree of thematic integration associated with multi-site and single-site strategies, respectively. By applying text mining techniques to the field of deep geological disposal, this study visualizes the evolution of research activities undertaken by leading companies in a data-driven manner. The results are expected to provide useful insights for latecomer countries, including the Republic of Korea, in formulating efficient R&D strategies.

Keywords: deep geological disposal, high-level radioactive waste, topic modeling, latent Dirichlet allocation (LDA), research trend analysis, structural break

OPEN ACCESS

*Corresponding author: Junghae Choi
E-mail: choi.jh@knu.ac.kr

Received: 30 March, 2026

Accepted: 20 April, 2026

© 2026 The Korean Society of Engineering Geology



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

초 록

본 연구는 심층 처분 분야 선도국인 스웨덴 SKB와 핀란드 POSIVA의 장기 연구개발(R&D) 경향을 정량적으로 비교·분석하기 위해, 잠재 디리클레 할당(latent Dirichlet allocation, LDA) 주제 모델링과 구조적 단절 분석을 적용하였다. 분석 대상은 SKB의 기술보고서 제목 2,520건과 POSIVA의 기술보고서 제목 1,014건이다. LDA 분석 결과, SKB는 9개, POSIVA는 6개의 주요 연구 주제가 도출되었으며, SKB는 상대적으로 분산된 연구 포트폴리오(상위 3개 주제 비중 49.6%)를, POSIVA는 보다 집중된 포트폴리오(상위 3개 주제 비중 72.2%)를 보였다. 연도별 주제 비중에 대한 회귀 분석과 Wald 검정 결과, 건설 허가 신청(SKB: 2011년, POSIVA: 2012년)을 기점으로 SKB는 9개 중 5개 주제, POSIVA는 6개 중 3개 주제에서 통계적으로 유의한 구조적 변화($p < 0.05$)가 관찰되었다. 특히 SKB의 수리지질학 모델링 연구는 허가 신청 이후 증가 속도가 가속화된 반면, POSIVA의 현장 특성화 연구는 감소 추세로 전환되어 두 기관 간 상이한 연구 단계 전환 양상이 확인되었다. 또한 주제 간 평균 유사도는 SKB (0.12)가 POSIVA (0.34)보다 낮아, 복수 부지 평가 전략과 단일 부지 집중 전략에 따른 연구 주제 통합도의 차이를 반영하는 것으로 해석되었다. 본 연구는 심층 처분 분야에 텍스트 마이닝 기법을 적용하여 선도국의 연구 진화 과정을 데이터 기반으로 가시화하였으며, 그 결과는 대한민국을 포함한 후발국의 효율적인 R&D 전략 수립을 위한 참고 자료로 활용될 수 있을 것으로 기대된다.

주요어: 심층 처분, 고준위 방사성폐기물, 주제 모델링, 잠재 디리클레 할당, 연구 경향 분석, 구조적 단절

서론

원자력 발전은 10만 년 이상의 장기 관리가 필요한 고준위 방사성폐기물(high-level radioactive waste, HLW)을 발생시킨다. 국제원자력기구(International Atomic Energy Agency, IAEA)는 안전한 심부 지질 환경에 폐기물을 격리하는 심층 처분(deep geological disposal)을 장기 안전 관리 방안으로 권고하고 있으나(IAEA, 2011), 이를 실제로 건설 단계까지 이행한 국가는 현재 스웨덴과 핀란드 두 국가에 불과하다.

핀란드의 POSIVA (Posiva Oy)는 2015년 세계 최초로 사용후핵연료 처분장 건설 허가를 획득하였고, 스웨덴의 SKB (Svensk Kärnbränslehantering AB)는 2022년 Forsmark 처분장 건설 허가를 획득하였다(Posiva Oy, 2015; SKB, 2022). 이로써 심층 처분 기술은 이론적 개념 검토 단계를 넘어 실증적 구현의 단계로 진입하게 되었다. 두 기관은 지난 수십 년간 수천 건의 기술보고서(technical report)를 발행하며 부지 특성 데이터, 공학적 방벽 설계, 지하연구시설 실험 결과 등 실증적 가치가 높은 지식 자산을 축적해 왔다.

그러나 이와 같이 장기간 축적된 연구 포트폴리오가 시간에 따라 어떻게 진화해 왔는지, 그리고 주요 정책적 이정표가 연구 전략과 주제 구성에 어떠한 영향을 미쳤는지에 대한 체계적인 정량 분석은 아직 미비한 상황이다. 최근에는 산사태, 지질공학 분야, 에너지 정책 등 다양한 분야에서 텍스트 마이닝과 주제 모델링 기법을 활용한 연구 동향 분석이 적용되고 있으나(Park and Yong, 2017; Liu et al., 2022; Kim et al., 2024), 심층 처분 분야에는 아직 적용된 사례가 많지 않다. 선도국의 연구 개발 방향을 정량적으로 분석하는 작업은 향후 심층 처분 사업을 추진해야 하는 후발국이 보다 효율적인 기술 확보 전략을 수립하는 데 중요한 참고 자료가 될 수 있다.

대한민국은 2016년 「고준위 방사성폐기물 관리 기본계획」 수립 이후 현재 연구용 URL 부지 선정 및 조사 단계에 있으며, 원전 부지 내 임시 저장시설이 2030년대 중반 포화될 것으로 예상됨에 따라 제한된 시간 내에 처분 기술을 확보해야 하는 과제에 직면해 있다(MOTIE, 2016). 이러한 상황에서 대한민국을 포함한 후발국이 선도국의 장기 연구개발(R&D) 전략과 그 진행 과정을 체계적으로 이해하는 것은 시간적 제약과 기술적 여건을 고려한 현실적인 전략 수립을 위해 중요한 의미를 가진다고 볼 수 있다.

이에 본 연구는 비지도 학습 기법인 잠재 디리클레 할당(latent Dirichlet allocation, LDA)을 활용하여 SKB (2,520건)와 POSIVA (1,014건)의 영문 기술보고서 제목을 분석하고, 두 기관의 장기 연구 경향을 정량적으로 비교·분석하고자 한다.

본 연구의 목적은 다음과 같이 정리할 수 있다. 첫째, 선도국의 핵심 연구 주제와 그 시간적 변화를 주제 모델링을 통해 정량적으로 도출한다. 둘째, 건설 허가 신청(SKB: 2011년, POSIVA: 2012년)이라는 정책적 사건을 전후로 연구 경향에 구조적 변화가 존재하는지를 통계적으로 검증한다. 셋째, 두 기관의 연구 전략 차이를 비교함으로써 부지 선정 전략의 차이가 R&D 포트폴리오 구성에 어떻게 반영되었는지를 분석한다.

이론적 배경

고준위방사성폐기물 심층 처분

스웨덴 SKB와 핀란드 POSIVA는 사용후핵연료 심층 처분 분야에서 선도적인 기관으로 평가된다. 두 기관 모두 구리-주철 용기, 벤토나이트 완충재, 결정질 기반암으로 구성된 KBS-3 (Kärnbränslesäkerhet-3) 다중방벽 개념을 채택하였다. 이 개념은 SKB의 Äspö 지하연구시설(1995년 운영 개시)과 POSIVA의 ONKALO (2004년 운영 개시)에서 수십 년간 검증되었으며, 수천 건의 기술보고서를 통해 문서화되어 왔다.

유사한 처분 개념에도 불구하고, 두 기관의 부지 선정 전략은 상이하였다. SKB는 Forsmark와 Oskarshamn 후보지를 병행 조사한 후 2009년 Forsmark를 선정한 반면, POSIVA는 초기 단계부터 Olkiluoto 단일 부지에 연구 역량을 집중하였다. 이러한 전략적 차이가 장기 R&D 포트폴리오 구성에 어떠한 차이를 야기했는지에 대해서는 제한적으로만 논의되어 왔다.

본 연구는 이러한 전략 차이가 실제 연구 주제의 분포와 시간적 진화에 어떻게 반영되었는지를 정량적으로 분석하고자 한다. 복수 부지 평가는 다양한 지질 환경에 대한 광범위한 연구를 요구하는 반면, 단일 부지 집중 전략은 특정 부지의 심층 특성화에 자원을 집중할 수 있다. 이러한 차이는 연구 포트폴리오의 다양성과 집중도에 구조적인 차이를 초래할 가능성이 있다.

잠재 디리클레 할당 모델링(latent Dirichlet allocation, LDA)

잠재 디리클레 할당(LDA)은 텍스트 코퍼스에서 잠재된 주제 구조를 확률적으로 추출하는 비지도 학습 기법이다(Blei et al., 2003). LDA는 하나의 문서가 여러 주제를 혼합하여 포함할 수 있다고 가정하므로, 열-수리-역학-화학(thermo-hydro-mechanical-chemical, THMC) 복합 거동, 안전성 평가, 재료과학 등 다학제적 성격을 지닌 심층 처분 연구의 주제 구조를 분석하는 데 적합하다.

LDA는 이미 미국국립과학원회보(Proceedings of the National Academic of Sciences, PNAS)에 발표된 과학문헌 연구 동향 분석에서 그 유효성을 널리 인정받은 바 있으며(Griffiths and Steyvers, 2004) 이후 다양한 지구과학 및 공학 분야에서 텍스트 마이닝 기법이 장기 경향 분석에 성공적으로 적용되어 왔다(Park and Yong, 2017; Liu et al., 2022; Kim et al., 2024).

본 연구에서는 많은 양의 기술보고서를 체계적으로 정리하고 분석의 효율성을 위해 기술보고서 초록이 아닌 제목을 분석 대상으로 삼았다. 보고서 제목은 핵심 연구 목적을 간결하게 표현하도록 작성되므로, 실험 세부사항이나 수치 정보가 포함되어 장기적인 주제 추세 분석을 어렵게 할 수 있는 초록에 비해, 거시적 연구 동향을 파악하는 데 보다 적합하다고 판단하였다.

이 접근법의 타당성을 검증하기 위해 POSIVA 최신 보고서 100건을 대상으로 제목과 초록 간의 키워드 일치도를 분석하였다. TF-IDF (term frequency-inverse document frequency) 가중치를 사용하여 각 초록에서 상위 5개 키워드를 추출한

결과, 전체 보고서의 84%에서 제목과 1개 이상의 키워드가 일치하였으며, 상위 5개 키워드 중 평균 1.7개(약 34%)가 제목과 공통으로 나타났다. 이러한 결과는 보고서 제목이 문서의 핵심 주제를 대표하는 데 충분한 정보를 포함하고 있음을 시사한다.

연구방법

본 연구는 SKB와 POSIVA 기술보고서를 대상으로 데이터 수집, 텍스트 전처리, LDA 모델링, 시계열 경향 및 구조적 단절 분석의 4단계 절차를 수행하였다(Fig. 1).

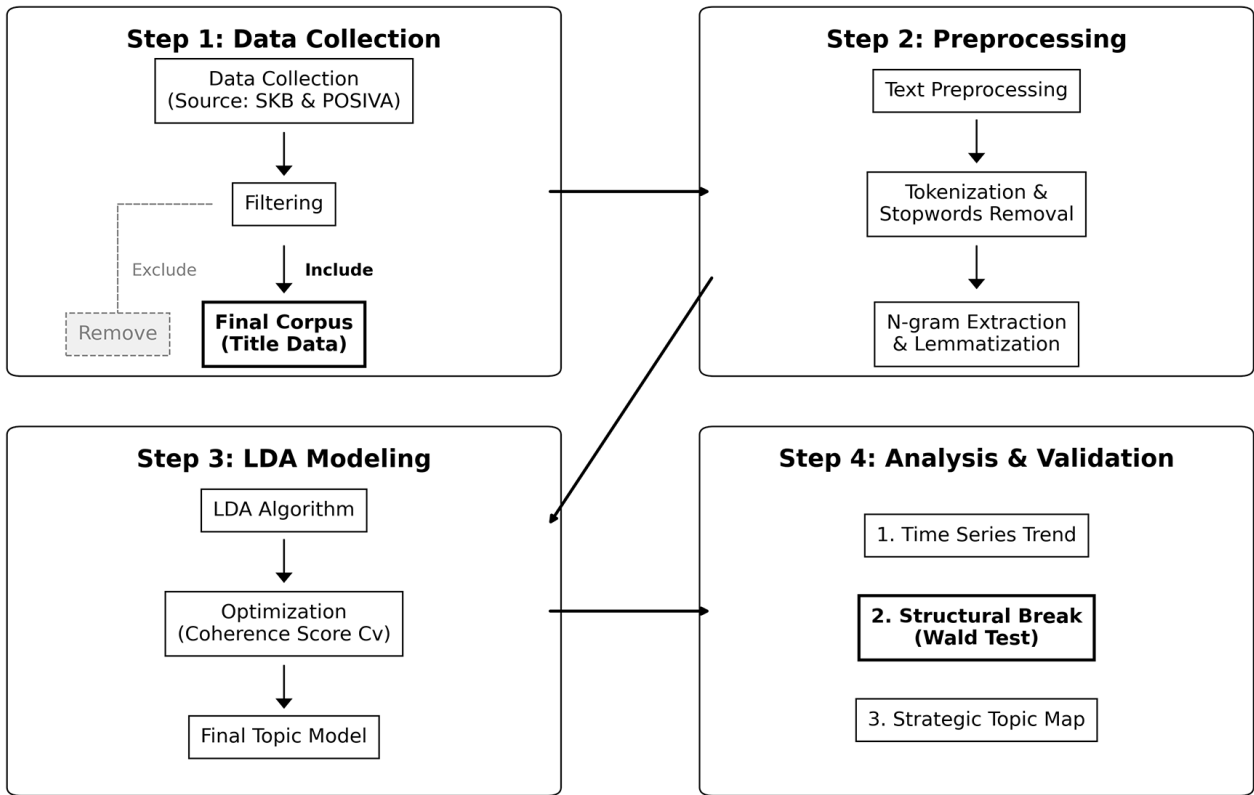


Fig. 1. Flowchart showing the overall research methodology.

데이터 수집 및 데이터셋 구축

SKB와 POSIVA가 설립 시점부터 2025년 11월까지 발행한 기술보고서를 각 기관의 공식 웹사이트(www.skb.com, www.posiva.fi)에서 수집하였다. SKB의 경우 공개된 전체 영문 기술보고서를, POSIVA는 POSIVA Reports, Working Reports, Posiva-SKB Reports 시리즈를 대상으로 하였다.

수집한 총 6,057건(SKB: 4,383건, POSIVA: 1,674건)의 보고서 중 심층 처분과 직접적 관련성이 낮은 보고서를 필터링 하여 제거하였다. 연례 보고서, 단순 시추공 데이터, 지표 생물권 보고서 등이 이에 해당하며, 제목에 특정 키워드(“annual report”, “monitoring well”, “biosphere monitoring” 등)를 포함한 문서를 제거하였으며, 최종적으로 SKB 2,520건, POSIVA 1,014건의 보고서 제목을 분석에 사용하였다.

POSIVA의 경우 시계열 분석의 통계적 안정성을 확보하기 위해 연간 발행 건수가 10건 미만인 2024-2025년 자료를 제외하고 2023년까지를 분석 대상으로 하였다. 연간 표본 수가 지나치게 적을 경우 주제 비중 추정치의 분산이 커져 회귀 분석 결과의 신뢰도가 저하될 수 있기 때문이다.

Table 1은 두 기관의 데이터 특성을 비교한 것이다. SKB가 더 긴 연구 기간(1977-2025)과 더 많은 보고서 수를 가지며, 제목당 평균 단어 수(6.36 vs 5.26)도 높은 것으로 나타났다. 구조적 단절 분석을 위한 기준 시점(break point)은 각 기관의 건설 허가 신청 연도(SK: 2011년, POSIVA: 2012년)로 설정하였다.

Table 1. Descriptive statistics of analyzed technical report titles

Category	POSIVA (Finland)	SKB (Sweden)
Study period (analysis)	1996 - 2023	1977 - 2025.11
Number of reports	1,014	2,520
Average words per title	5.26	6.36

단어 전처리

수집된 텍스트 데이터는 LDA 모델 학습의 안정성과 성능을 향상시키기 위해 전처리 과정을 거쳤다. 먼저 모든 텍스트를 소문자로 변환한 후, 일반적인 영문 불용어(the, a, an 등)와 함께 ‘report’, ‘study’, ‘analysis’, ‘result’, ‘data’ 등 문서 주제와 직접적인 관련이 없는 일반 어휘를 제거하였다. 또한 명사(NOUN), 고유명사(PROPN), 형용사(ADJ)만을 분석 대상으로 하여, 의미 정보가 적은 품사는 제외하였다.

의미 단위를 보존하기 위해 Gensim 라이브러리의 Phrases 모듈을 활용하여 연어(collocation)를 추출하였다. “Safety Assessment”나 “Groundwater Flow”처럼 의미 단위로 결합된 복합 명사를 식별하기 위해 NPMI (normalized pointwise mutual information) 기반 점수와 임계값 0.5를 적용하였다.

NPMI는 두 단어의 공출현 빈도를 개별 빈도로 정규화한 지표로, -1 (전혀 공출현하지 않음)에서 +1 (항상 공출현) 범위를 가진다(Bouma, 2009). 임계값 0.5는 비교적 강한 결합 관계를 나타내며, 기준을 충족하는 단어 조합은 Bi-gram 및 Tri-gram 수준까지 단일 토큰으로 변환하였다. 이를 통해 모델이 개별 단어가 아닌 기술 개념 단위로 학습하도록 하였다. 마지막으로 spaCy 라이브러리를 이용한 표제어 추출을 적용하여 단어의 품사를 추출하였다.

LDA 모델링 및 하이퍼파라미터 최적화

전처리가 완료된 문서-단어 행렬에 Python Tomotopy 라이브러리를 사용하여 LDA 분석을 수행하였다. Tomotopy는 확률적 샘플링 방법인 Gibbs sampling 알고리즘을 기반으로 하여 대규모 코퍼스 처리에 효율적이다. 모델 학습은 최대 20,000회 반복하였으며, 연속 100회 반복 간 단어당 로그 우도 변화율이 0.001 미만으로 감소하는 시점을 수렴으로 판단하였다.

LDA 모델의 최적 주제 수(K)는 주제 일관성(topic coherence, C_v) 지표로 결정하였다. C_v 는 각 주제 내 상위 단어들 간의 의미론적 유사도를 정량화한 지표로, 인간의 주관적 해석 가능성과 높은 상관관계를 갖는 것으로 알려져 있다(Röder et al., 2015). K 를 5에서 15까지 변화시키며 평가한 결과, SKB는 $K=9$, POSIVA는 $K=6$ 에서 C_v 점수가 최대값을 보였다(Fig. 2). 각 주제의 비중 분포와 상위 키워드의 해석 가능성을 함께 검토한 결과 이를 최적값으로 선택하였다.

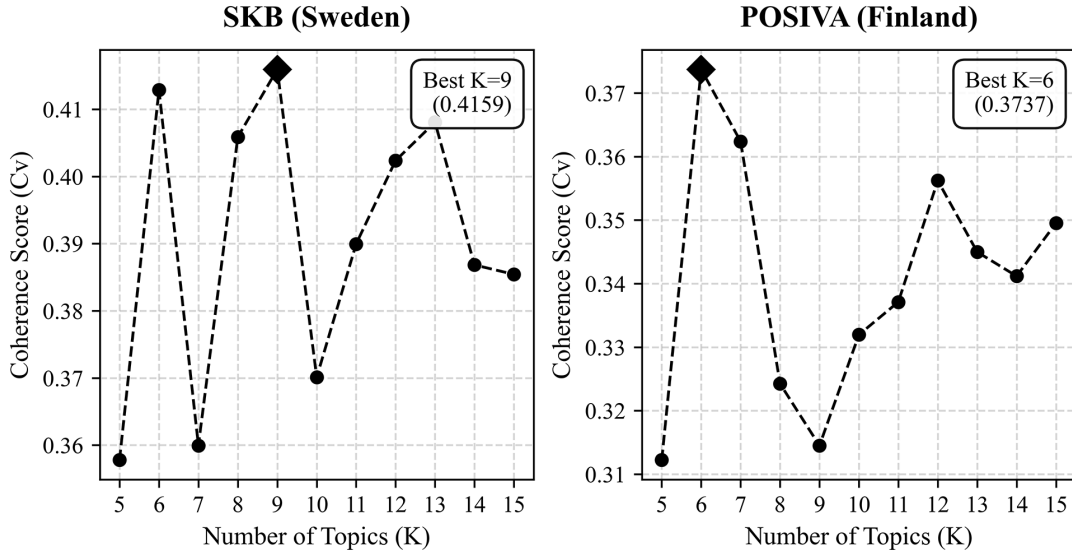


Fig. 2. Topic coherence scores for model selection.

LDA의 주요 하이퍼파라미터인 α (문서-주제 분포)와 η (주제-단어 분포)는 모델의 희소성(sparsity)과 C_v 를 고려하여 선정하였다. α 는 문서가 여러 주제를 혼합하는 정도를 조절하는 매개변수로, 값이 낮을수록 문서가 소수의 주제에 집중되는 경향을 보인다. η 는 각 주제가 여러 단어를 포함하는 정도를 조절하며, 값이 낮을수록 주제가 소수의 핵심 단어로 구성된다. 다양한 조합($\alpha = [0.01, 0.05, 0.1, 0.2]$, $\eta = [0.01, 0.05, 0.1]$)을 테스트한 결과, 두 기관 모두 $\alpha = 0.1$, $\eta = 0.01$ 조건에서 주제 간 구분이 가장 명확하게 드러났다. 이 설정은 각 문서가 평균적으로 2-3개의 주요 주제를 포함하면서도, 각 주제는 소수의 핵심 단어로 특징지어지는 해석 가능한 구조를 생성하였다는 것을 의미한다.

추세 및 구조적 변화분석

도출된 각 주제의 시계열적 변화를 분석하기 위해, 연도별 주제 비중의 평균을 산출한 후 단순 선형 회귀 분석을 수행하였다. 연도 t 에서 주제 i 의 평균 비중을 P_{it} 라 할 때, 전체 기간에 대한 회귀 모델은 식 (1)과 같이 표현된다.

$$P_{it} = \beta_0 + \beta_1 \cdot t + \epsilon_{it} \tag{1}$$

여기서, β_1 은 연간 변화율을 나타낸다. 회귀 계수의 부호와 통계적 유의성($p < 0.1$)을 기준으로 각 주제를 성장형(emerging, $\beta_1 > 0$), 쇠퇴형(declining, $\beta_1 < 0$), 안정형(stable, $p \geq 0.1$)으로 분류하였다. 유의수준 $p = 0.1$ 은 개별 주제의 장기적 변화 경향을 비교적 폭넓게 포착하기 위한 기준으로 설정하였다.

건설 허가 신청(SKB: 2011년, POSIVA: 2012년)을 기점으로 연구 경향의 구조적 변화가 발생했는지 검증하기 위해 Wald 검정을 수행하였다. 각 주제 i 에 대하여 회귀 모형을 식 (2)와 같이 설정하였다.

$$P_{it} = \beta_0 + \beta_1 \cdot Year_t + \beta_2 \cdot D_t + \beta_3 \cdot (Year_t \times D_t) + \epsilon_{it} \tag{2}$$

여기서, P_{it} 는 연도 t 에서 주제 i 의 평균 비중(0-1 범위의 연속 변수)이며, $Year_t$ 는 연도를 나타내는 연속 변수이다. D_t 는 건설 허가 신청 시점 이후를 1로, 그 이전을 0으로 하는 더미 변수이다. β_0 는 절편, β_1 은 건설 허가 신청 이전 주제 비중의 연간 변화율이다. β_2 는 건설 허가 신청 이후 주제 비중의 수준 변화(level shift)를, β_3 는 건설 허가 신청 이후 증가 속도 변화(slope change)를 나타낸다. ϵ_{it} 는 오차항이다.

Wald 검정을 통해 귀무가설($H_0: \beta_2 = \beta_3 = 0$)을 검증함으로써, 건설 허가 신청 시점을 전후하여 주제 비중의 수준 또는 증가 속도에 구조적 변화가 존재하는지를 판별하였다. 구조적 변화의 통계적 유의성은 유의수준 $p = 0.05$ 를 기준으로 판단하였다.

분석결과 및 토의

보고서 출판 동향

SKB와 POSIVA의 연간 출판량은 사업 단계에 따라 뚜렷한 변화를 보였다(Fig. 3). SKB는 2000년대 중반 Forsmark와 Oskarshamn 부지 특성화가 본격화되면서 연간 출판량이 400건 이상으로 급증하였다. POSIVA 역시 2004년 ONKALO 굴착이 시작된 이후 2010년경까지 연간 최대 약 100건의 보고서를 발행하였다.

두 기관 모두 건설 허가 신청(SKB: 2011년, POSIVA: 2012년) 이후에는 출판량이 감소하거나 안정화되는 경향을 보였다. 이러한 변화는 현장 데이터 수집 중심의 단계에서, 확보된 자료를 바탕으로 한 해석 및 안전성 평가 중심의 단계로 연구 초점이 이동한 것을 반영하는 것으로 해석될 수 있다.

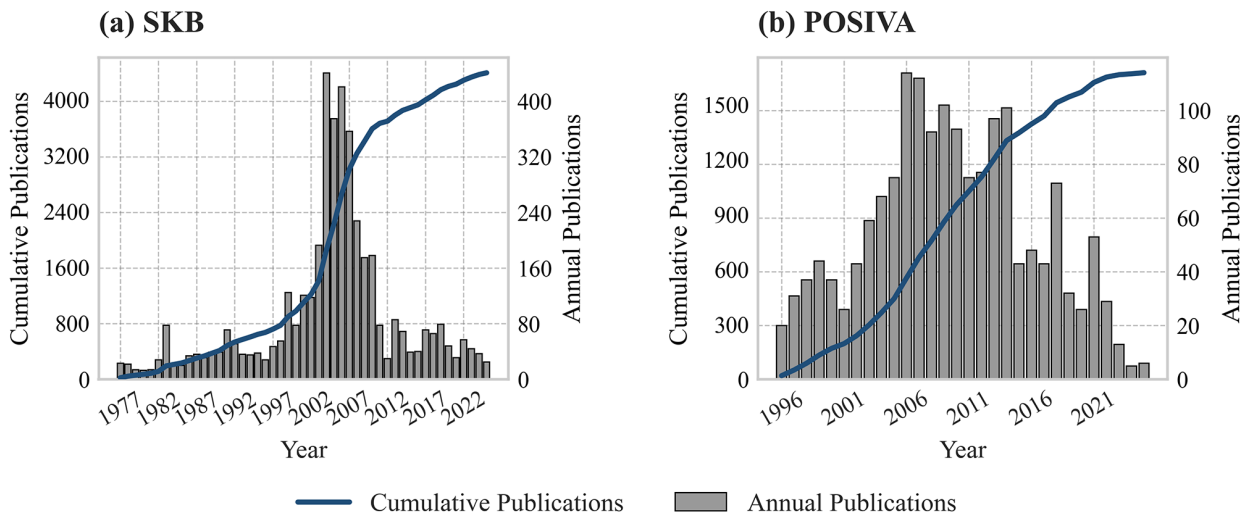


Fig. 3. Trends in the number of publications by SKB and POSIVA.

SKB 주제 분석

SKB의 2,520건 보고서 제목에 대한 LDA 분석 결과 9개의 주요 주제가 도출되었다(Table 2). 가장 높은 비중을 차지한 주제는 Topic 1 (Hydrogeological Modelling, 20.0%)이며, Topic 2 (URL Experiments, 15.0%)와 Topic 3 (Engineered Barrier System, 14.6%)이 뒤를 이었다. 상위 3개 주제가 전체의 49.6%를 차지하여, 수리지질학 모델링, 지하연구시설 실험, 공학적 방벽 시스템을 중심으로 전개되어 왔음을 시사한다.

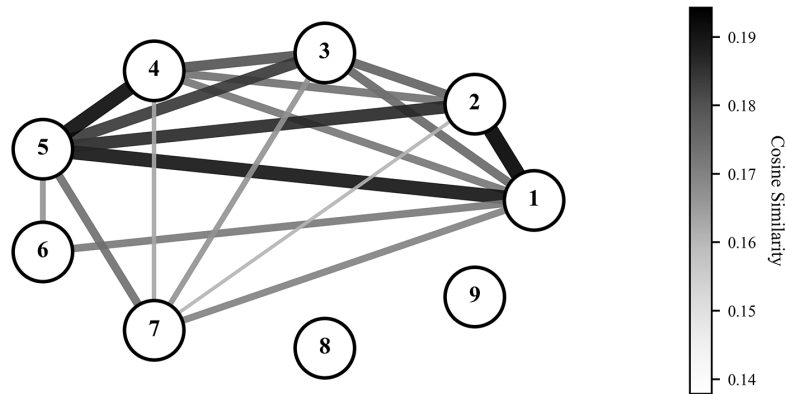
Table 2. Top keywords and R&D topics for SKB

Topic ID (proportion)	Label	Top 10 keywords
Topic 1 (20.0%)	Hydrogeological Modelling	modelling, model, site, groundwater_flow, fracture, site_descriptive, transport, numerical, hydrogeological, simulation
Topic 2 (15.0%)	URL Experiments	hard_rock_laboratory, modelling, experiment, measurement, rock, model, HRL, investigation, site_investigation, evaluation
Topic 3 (14.6%)	Engineered Barrier System	test, bentonite, buffer, water, backfill, laboratory, rock, material, modelling, canister
Topic 4 (13.4%)	Canister & Fuel Stability	copper, corrosion, fuel, canister, groundwater, water, bentonite, nuclear_fuel, effect, creep
Topic 5 (13.3%)	Radionuclide Transport	transport, radionuclide, model, calculation, flow, migration, assessment, clay, diffusion, influence
Topic 6 (7.7%)	Safety Assessment	site_investigation, SFR, safety_assessment, final_repository, site, climate, nuclear_fuel, bedrock, seismic, process
Topic 7 (6.7%)	Repository Engineering	repository, hard_rock_laboratory_prototype, condition, groundwater, construction, evolution, tunnel, deep, prototype_repository, design
Topic 8 (5.8%)	Seismology & Geophysics	investigation, earthquake, swedish_national_seismic_network, site, interpretation, site_investigation, geological, geophysical, lineament, characterization
Topic 9 (3.5%)	Site Monitoring	site_investigation, monitoring, hydrological, water, surface, borehole, ice, water_sampling, temperature, post

LDA 모델링 결과 각 보고서는 평균적으로 1.8개의 주제를 포함하는 것으로 나타나, 개별 연구가 단일 주제에 국한되기 보다는 여러 연구 주제를 복합적으로 다루는 경향이 있음을 보여준다. 주제 간의 상호 연관성을 정량적으로 파악하기 위해 주제 분포 간 코사인 유사도(cosine similarity)를 산출하고 네트워크 분석을 수행하였다. 그 결과, 주제 간 유사도의 평균값은 0.12, 최대값은 0.19로 전반적으로 낮은 수준을 보였다. 이는 각 주제가 비교적 명확히 구분되어 있으며, 모델이 주제 간 중복을 효과적으로 최소화하였음을 시사한다.

개별 주제 간의 연결성을 살펴보면, 가장 높은 유사도를 보인 조합은 Topic 1 (Hydrogeological Modelling)과 Topic 2 (URL Experiments)로, 유사도 값은 0.19이다. 이는 SKB가 수리지질학적 모델링 연구를 Äspö 지하연구시설에서 수행된 현장 실험과 긴밀히 연계하여 수행해 왔음을 보여준다. 다음으로 강한 연결성을 보인 조합은 Topic 4 (Canister & Fuel Stability)와 Topic 5 (Radionuclide Transport)로, 역시 유사도 값은 0.19이다. 이는 용기 부식 및 연료 안정성 연구가 단순한 재료공학적인 문제에 국한되지 않고, 핵종 이동과 연계된 안전성 평가 맥락 속에서 다루어졌음을 시사한다. 이러한 주제 간 관계를 나타낸 네트워크 구조는 Fig. 4에 제시하였다.

Network Analysis Result: SKB



T1: Hydrogeological Modelling

T2: URL Experiments

T3: Engineered Barrier System

T4: Canister & Fuel Stability

T5: Radionuclide Transport

T6: Safety Assessment

T7: Repository Engineering

T8: Seismology & Geophysics

T9: Site Monitoring

Fig. 4. Results of topic network analysis (SKB).

POSIVA 주제 분석

POSIVA의 1,014건 보고서 제목에 대한 LDA 분석 결과 6개의 주요 주제가 도출되었다(Table 3). 상위 3개 주제가 전체의 72.2%를 차지하여 SKB (49.6%)에 비해 높은 집중도를 보였다. 가장 높은 비중을 차지한 주제는 Topic 1 (Underground Characterization, 30.4%)이며, Topic 2 (Site Modelling, 21.5%)와 Topic 3 (Disposal System Design, 20.3%)이 뒤를 이었다. 이러한 결과는 POSIVA가 초기 단계부터 단일 부지(Olkiluoto)에 연구 역량을 집중하고, ONKALO 시설을 활용한 현장 자료 확보와 부지 적합성 검증에 중점을 두어 왔음을 시사한다.

Table 3. Top keywords and R&D topics for POSIVA

Topic ID (proportion)	Label	Top 10 keywords
Topic 1 (30.4%)	Underground Characterization	ONKALO, fracture, tunnel, seismic, rock, underground, water, evaluation, deposition_hole, effect
Topic 2 (21.5%)	Site Modelling	site, model, groundwater, investigation, bedrock, modelling, description, evolution, geological, interpretation
Topic 3 (20.3%)	Disposal System Design	repository, fuel, disposal, design, KBS-3H, assessment, system, modelling, canister, disposal_canister
Topic 4 (10.6%)	Engineered Barrier System	bentonite, material, buffer, backfill, test, development, scale, interaction, effect, testing
Topic 5 (10.6%)	Rock Properties & Testing	rock, measurement, test, sample, property, situ, laboratory, gas, ONKALO_experiment, modelling
Topic 6 (6.6%)	Geochemistry	sorption, copper, condition, groundwater, kaolinite, experiment, diffusion, solution, UO ₂ , solubility

POSIVA 보고서는 평균적으로 2.3개의 주제를 포함하는 것으로 나타났으며, 주제 간 평균 유사도는 0.34로 SKB (0.12)에 비해 높은 값을 보였다. 이는 POSIVA의 연구 주제들이 상호 간에 보다 긴밀하게 연계되어 있음을 의미한다. 특히 Topic 1 (Underground Characterization)과 Topic 2 (Site Modelling)는 유사도 0.42로 가장 높은 연관성을 보여, ONKALO 현장에서 획득된 자료가 부지 모델링에 직접적으로 활용되었음을 시사한다. 또한 Topic 2 (Site Modelling)와 Topic 3 (Disposal System Design) 역시 유사도 0.39를 보여, 부지 특성화 연구와 처분 시스템 설계가 통합적으로 수행되어 왔음을 나타낸다. 이러한 주제 간 관계를 나타낸 네트워크 구조는 Fig. 5에 제시하였다.

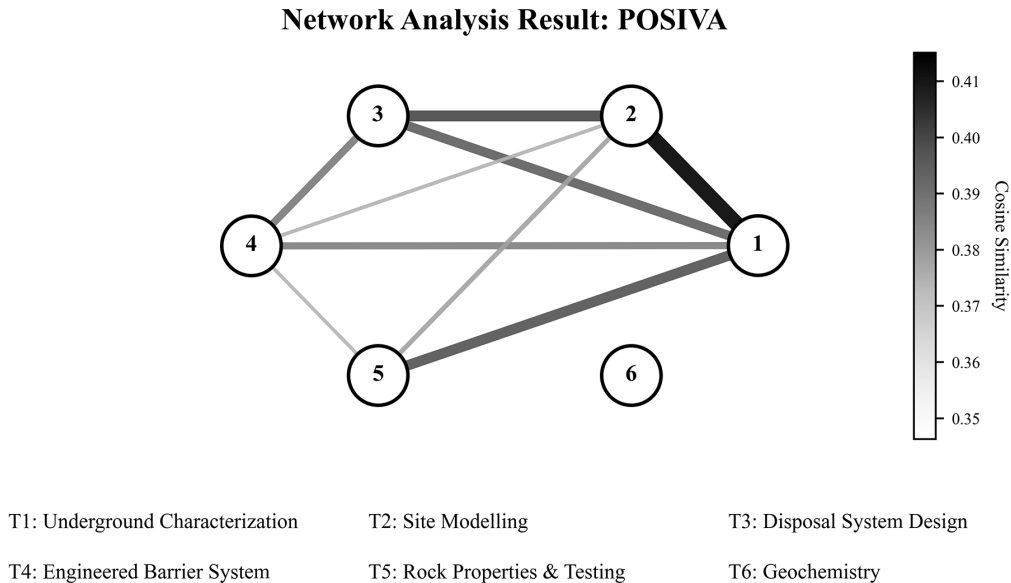


Fig. 5. Results of topic network analysis (POSIVA).

SKB와 POSIVA 연구 포트폴리오 비교

SKB와 POSIVA의 연구 포트폴리오는 두 기관의 부지 선정 전략 차이를 반영하는 것으로 나타났다. SKB는 9개의 주제로 구성된 상대적으로 분산된 포트폴리오(상위 3개 주제 비중 49.6%)를 보인 반면, POSIVA는 6개의 주제로 구성된 보다 집중된 포트폴리오(상위 3개 주제 비중 72.2%)를 보였다. SKB의 주제 간 평균 유사도(0.12)는 POSIVA (0.34)보다 낮게 나타났다. 이는 SKB의 분석 대상 문서 수가 POSIVA보다 약 2.5배 많아 주제가 더 세분화된 영향도 있겠으나, 복수 후보지를 병행 조사하는 과정에서 다양한 지질 환경과 연구 주제를 비교적 독립적으로 다루어 왔음을 시사한다.

POSIVA의 최대 단일 주제(Topic 1, 30.4%) 비중은 SKB의 최대 주제(Topic 1, 20.0%)보다 높게 나타났으며, 이는 Olkiluoto 단일 부지에 집중하여 ONKALO 시설을 기반으로 한 현장 특성화 연구에 연구 자원이 집중되었음을 반영하는 것으로 해석될 수 있다. 반면 SKB는 수리지질학(Topic 1), 지하연구시설 실험(Topic 2), 공학적 방벽(Topic 3), 용기 및 연료 안정성(Topic 4), 핵종 이동(Topic 5) 등 여러 핵심 주제에 비교적 고르게 연구 역량을 배분하여, 후보지 평가 단계에서의 유연성을 확보하는 전략을 취한 것으로 보인다.

이러한 차이는 부지 선정 시점과도 관련이 있다. POSIVA는 1999년에 Olkiluoto를 최종 부지로 선정할 이후 해당 부지에 대한 심층 연구를 지속한 반면, SKB는 2009년에 Forsmark를 최종 선정하기까지 여러 후보지를 병행 평가하였다. 본 연구의 정량적 분석 결과는 이러한 부지 선정 전략의 차이가 각 기관의 연구개발 포트폴리오 구성에 반영되어 왔음을 시사한다.

SKB 연구 주제의 구조적 변화분석

2011년 건설 허가 신청을 기준으로 수행한 Wald 검정 결과, SKB의 9개 주제 중 5개(Topic 1, Topic 2, Topic 3, Topic 4, Topic 5)에서 통계적으로 유의한 구조적 단절($p < 0.05$)이 관찰되었다(Table 4, Fig. 6).

Table 4. Summary statistics for SKB structural break analysis

Topic	Label	Pre-slope	Post-slope	Slope change	Wald p -value	Result
1	Hydrogeological Modelling	0.003569	0.0062	0.002631	1.73E-05	Significant break
2	URL Experiments	0.003997	0.002791	-0.00121	0.007588	Significant break
3	Engineered Barrier System	0.001198	-0.0024	-0.00359	0.048758	Significant break
4	Canister & Fuel Stability	-0.00266	-0.00352	-0.00086	0.000285	Significant break
5	Radionuclide Transport	-0.0068	0.000139	0.006936	0.000348	Significant break
6	Safety Assessment	0.001769	0.00053	-0.00124	0.743236	No break
7	Repository Engineering	0.000288	-0.00159	-0.00187	0.099466	No break
8	Seismology & Geophysics	-0.00264	-0.00343	-0.00078	0.313768	No break
9	Site Monitoring	0.001279	0.001267	-1.2E-05	0.076065	No break

Pre-slope: annual change rate before 2011 (percentage points/year); Post-slope: annual change rate after 2011; Slope change: Post-slope minus Pre-slope. Bold rows indicate statistically significant structural breaks ($p < 0.05$).

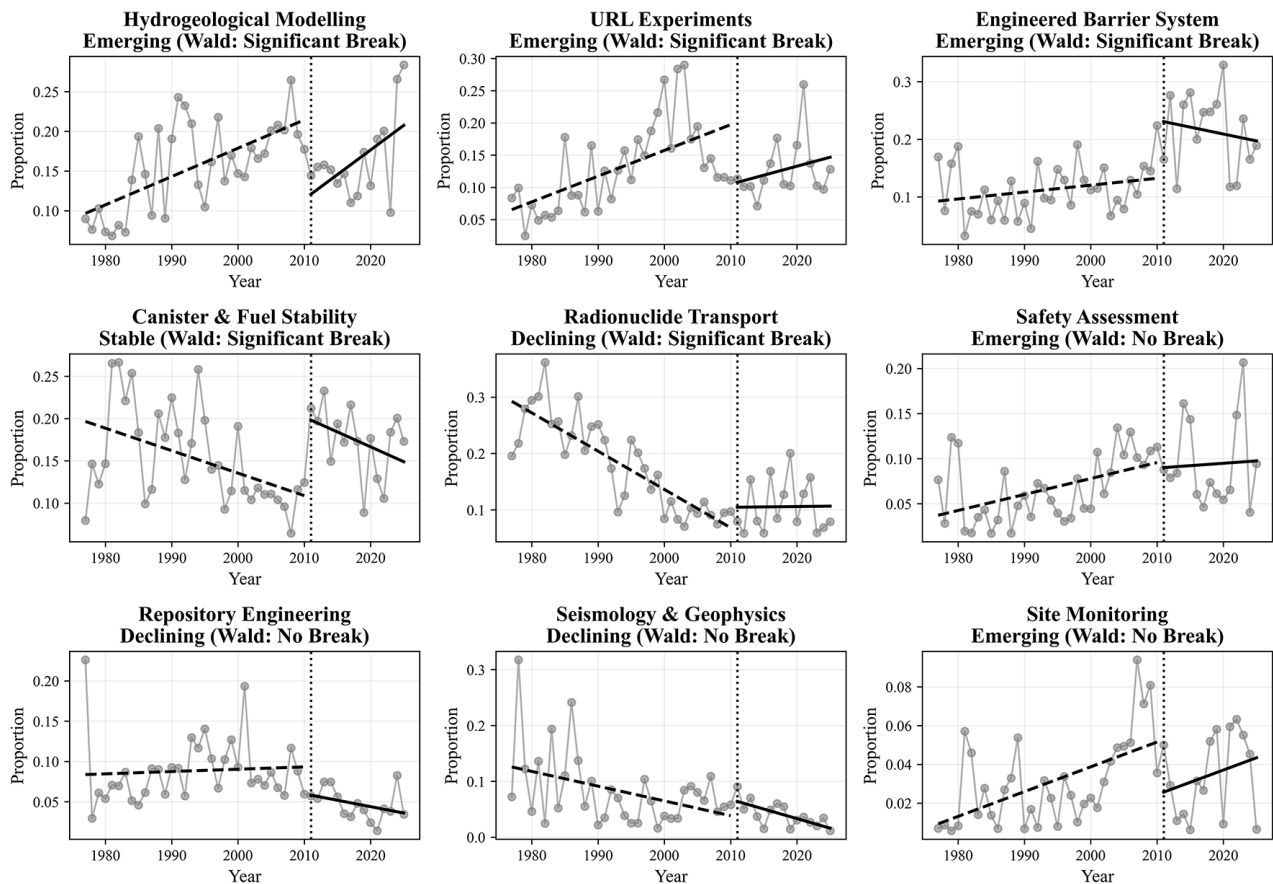


Fig. 6. Time-series trends and results of structural break analysis for SKB R&D topics (break point: 2011).

Topic 1 (Hydrogeological Modelling)은 2011년 이전 연간 0.36%p 증가하였으나, 이후에는 +0.62%p로 증가 속도가 가속화되었다. 이는 건설 허가 심사 과정에서 부지 수리지질학 모델의 정교화 및 불확실성 평가가 더욱 중요해지면서, 관련 모델링 연구의 비중이 확대된 것과 관련이 있는 것으로 해석될 수 있다(Geier, 2012). 반면 Topic 2 (URL Experiments)는 2011년 이전 연간 0.40%p 증가하였으나, 이후 +0.28%p로 둔화되었다. 이는 Äspö HRL에서의 실험 중심 연구에서, 점차 데이터 해석 및 모델 검증 중심의 연구로 무게중심이 이동한 흐름을 반영하는 것으로 보인다.

Topic 3 (Engineered Barrier System)과 Topic 4 (Canister & Fuel Stability)는 건설 허가 신청 이후 음의 성장률을 기록하였다. Topic 3은 이전 +0.12%p/yr에서 이후 -0.24%p/yr로 전환되었으며, Topic 4는 -0.27%p/yr에서 -0.35%p/yr로 감소 추세가 심화되었다. 이는 공학적 방벽 및 용기 관련 기술이 일정 수준의 성숙 단계에 도달함에 따라, 개별 기술 개발보다는 안전성 평가 체계 내에서의 통합적 해석과 모델링 연구로 연구 초점이 이동했음을 시사한다.

Topic 5 (Radionuclide Transport)는 2011년 이전 연간 -0.68%p로 비교적 급격한 감소 추세를 보였으나, 이후에는 거의 변화가 없는 수평 추세(+0.01%p/yr)로 전환되었다. 이는 초기 개념 검토 중심의 연구 단계를 지나, 안전성 평가 체계 내에서 보다 안정적인 역할을 차지하게 되었음을 반영하는 것으로 해석될 수 있다.

구조적 단절이 관찰되지 않은 주제(Topic 6, Topic 7, Topic 8, Topic 9)은 건설 허가 신청 전후 일관된 추세를 유지하였다. 특히 Topic 6 (Safety Assessment)는 전체 기간 동안 낮은 비중을 유지하면서도 완만한 증가 추세를 보여, 안전성 평가 연구가 특정 시점에 집중되기보다는 장기간에 걸쳐 지속적으로 수행되어 왔음을 시사한다.

Fig. 7은 SKB의 주제별 전략적 위치를 요약한 주제 맵을 나타낸다. X축은 주제의 성장성(선형 회귀 기울기), Y축은 중요도(연도별 주제 비중의 평균)를 나타낸다. Topic 1, Topic 2, Topic 3은 제1사분면(고중요도-성장)에 위치하여 향후에도 지속적인 연구 투자가 이루어질 가능성이 높은 핵심 영역으로 해석된다. Topic 4, Topic 5는 제2사분면(고중요도-쇠퇴)에 위치하여 성숙 단계에 진입한 연구 주제로 분류된다. Topic 7, Topic 8은 제3사분면(저중요도, 쇠퇴)에 위치하여, 사업 초기에는 중요했으나 현재는 상대적 비중이 감소한 연구 주제로 해석될 수 있다. Topic 6은 제4사분면(저중요도-성장)에 위치하여 향후 중요성이 확대될 가능성이 있는 신흥 연구 영역으로 분류된다.

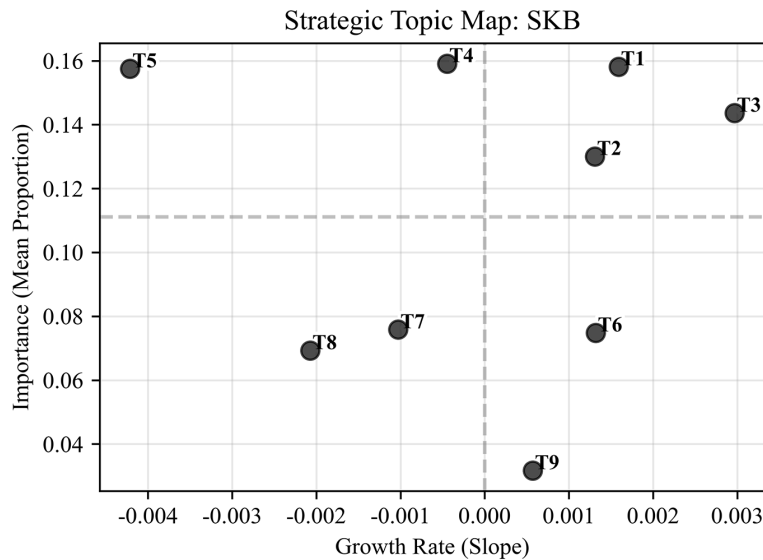


Fig. 7. Strategic topic map for SKB publications.

POSIVA 연구 주제의 구조적 변화 분석

2012년 건설 허가 신청을 기준으로 수행한 Wald 검정 결과, POSIVA의 6개 주제 중 3개에서 통계적으로 유의한 구조적 단절이 관찰되었다(Table 5, Fig. 8).

Table 5. Summary statistics for POSIVA structural break analysis

Topic	Label	Pre-slope	Post-slope	Slope change	Wald p -value	Result
1	Underground Characterization	0.007963	-0.00339	-0.01135	0.000949	Significant break
2	Site Modelling	-0.00265	0.00088	0.003529	0.265836	No break
3	Disposal System Design	0.000283	-0.00647	-0.00675	0.439708	No break
4	Engineered Barrier System	0.00497	0.003371	-0.0016	0.018321	Significant break
5	Rock Properties & Testing	-0.00347	0.004761	0.008232	0.075899	No break
6	Geochemistry	-0.0071	0.000843	0.007939	0.024918	Significant break

Pre-slope: annual change rate before 2012 (percentage points/year); Post-slope: annual change rate after 2012; Slope change: Post-slope minus Pre-slope. Bold rows indicate statistically significant structural breaks ($p < 0.05$).

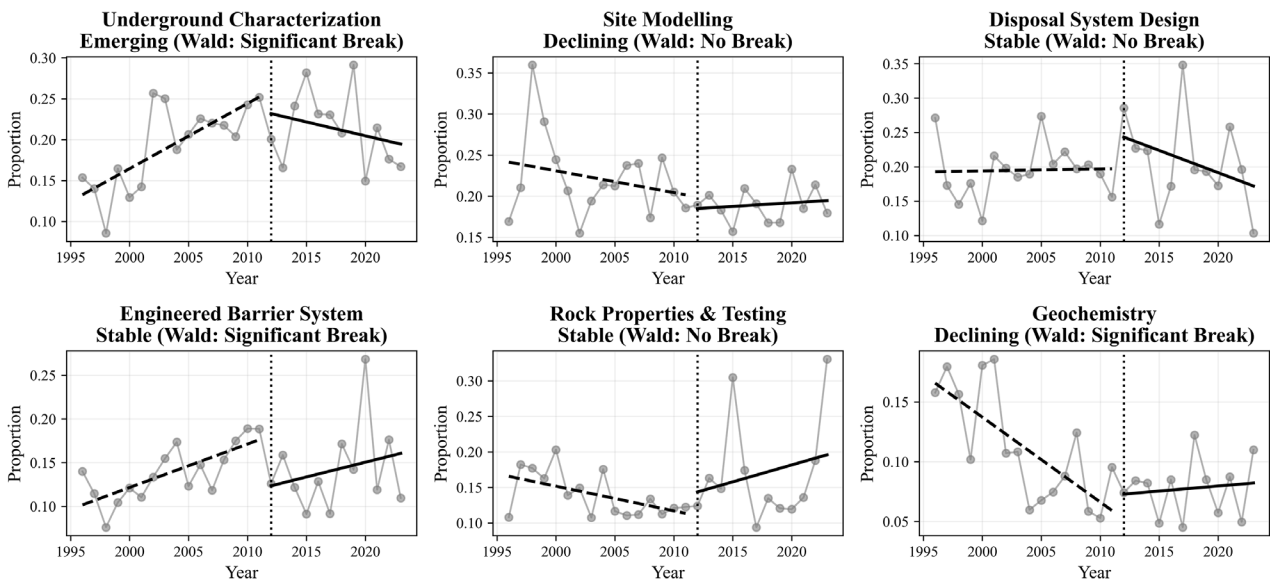


Fig. 8. Time-series trends and results of structural break analysis for POSIVA topics (break point: 2012).

Topic 1 (Underground Characterization)은 전체 기간 동안 최대 비중(30.4%)을 차지한 핵심 주제이지만, 2012년 이전에는 연간 +0.80%p로 증가하던 추세가 이후에는 -0.34%p로 감소하는 방향으로 전환되었다. 이는 ONKALO 건설과 현장 데이터 수집이 집중되었던 2000년대 후반을 지나, 건설 허가 신청 단계에 이르러 연구의 초점이 자료 축적에서 자료 통합 및 평가로 이동하였음을 반영하는 것으로 해석될 수 있다. 실제로 핀란드 규제기관 STUK는 2015년에 Olkiluoto 부지의 ONKALO 특성화가 충분한 수준에 도달하였다고 평가한 바 있으며(STUK, 2015), 이러한 평가는 현장 특성화 연구 비중이 감소하는 경향과도 일치한다.

Topic 4 (Engineered Barrier System)는 2012년 이전에는 연간 +0.50%p, 이후에는 +0.34%p로 증가 추세 자체는 유지되었으나, 증가 속도가 통계적으로 유의미하게 둔화되었다. 이는 공학적 방벽 관련 기술이 일정 수준의 성숙 단계에 도달

함에 따라, 실증 실험 중심의 연구에서 장기 성능 평가 및 해석 중심의 연구로 점진적으로 연구 초점이 이동하고 있음을 시사한다.

Topic 6 (Geochemistry)은 2012년 이전에는 연간 $-0.71\%p$ 로 감소 추세를 보였으나, 이후에는 거의 변화가 없는 수평 추세($+0.08\%p/yr$)로 전환되었다. 이는 초기 단계에서 수행된 부지 지구화학 특성화 연구가 일정 수준 완료된 이후, 안전성 평가를 위한 지구화학 모델링이 비교적 안정적인 비중으로 유지되고 있음을 반영하는 것으로 해석될 수 있다.

Topic 2, Topic 3, Topic 5는 건설 허가 신청 전후에 통계적으로 유의한 변화가 관찰되지 않았다. 특히 Topic 3 (Disposal System Design)은 전체 기간 동안 20.3%의 높은 비중을 유지하면서 비교적 일관된 추세를 보여, 처분 시스템 설계 연구가 사업 단계 변화와 무관하게 지속적으로 핵심 연구 분야로 유지되어 왔음을 시사한다.

Fig. 9는 POSIVA의 주제별 전략적 위치를 요약한 주제 맵을 나타낸다. Topic 1과 Topic 3은 제1사분면(고중요도-성장)에 위치하나, Topic 1의 경우 전체 기간 평균으로는 성장형으로 분류되나 2012년 이후에는 감소 추세로 전환되었다는 점에서 해석에 주의가 필요하다. Topic 2는 제2사분면(고중요도-쇠퇴)에 위치하여 해당 연구 주제가 성숙 단계에 진입하였음을 시사한다. 제3사분면에는 Topic 6이 위치하며, 낮은 비중과 과거의 감소 추세를 고려할 때 초기 단계의 기초 연구 성격을 지닌 주제로 해석될 수 있다. 제4사분면에는 Topic 4와 Topic 5가 위치하여, 중간 수준의 비중을 유지하면서 증가 추세를 보이는 향후 중요성이 확대될 가능성이 있는 연구 영역으로 분류된다.

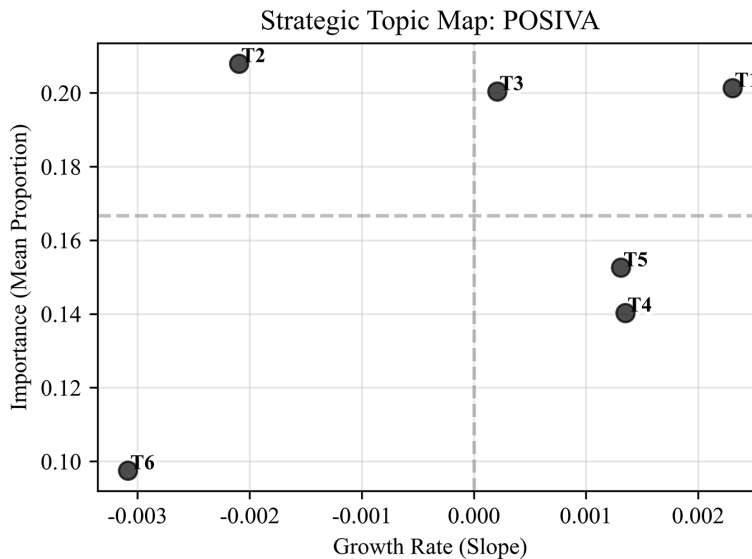


Fig. 9. Strategic topic map for POSIVA publications.

결론

본 연구는 LDA 주제 모델링과 구조적 단절 분석을 결합하여, SKB (2,520건)와 POSIVA (1,014건)의 기술보고서 제목을 대상으로 심층 처분 분야의 장기 연구 경향을 정량적으로 분석하였다. 이를 통해 두 기관의 연구 포트폴리오 구성과 시간에 따른 변화 양상이 부지 선정 전략 및 사업 단계와 어떻게 연관되어 있는지를 비교/분석하였다.

분석 결과는 다음과 같이 요약될 수 있다. 첫째, 부지 선정 전략은 각 기관의 R&D 포트폴리오 구성과 밀접한 관련을 보였다. 복수 부지를 병행 평가한 SKB는 9개 주제로 구성된 상대적으로 분산된 포트폴리오(상위 3개 주제 비중 49.6%)를

보인 반면, 단일 부지에 집중된 POSIVA는 6개 주제로 구성된 보다 집중된 포트폴리오(상위 3개 주제 비중 72.2%)를 보였다. 둘째, 건설 허가 신청은 연구 경향에 구조적 변화를 동반한 중요한 전환점으로 나타났다. SKB는 9개 주제 중 5개, POSIVA는 6개 주제 중 3개에서 통계적으로 유의한 구조적 단절이 관찰되었다($p < 0.05$). 특히 SKB의 수리지질학 모델링 연구는 허가 신청 이후 오히려 증가 속도가 가속화된 반면, POSIVA의 현장 특성화 연구는 뚜렷한 감소 추세로 전환되어, 두 기관 간 상이한 연구 단계 전환 양상을 보여주었다. 셋째, 주제 간 평균 유사도의 차이(SKB 0.12, POSIVA 0.34)는 부지 선정 전략에 따른 연구 주제 간 통합도의 차이를 반영하는 지표로 해석될 수 있다.

본 연구는 심층 처분 분야에 텍스트 마이닝 기법을 적용하여, 선도국들의 연구 활동이 시간에 따라 어떻게 진화해 왔는지를 데이터 기반으로 가시화하였다는 점에서 방법론적 의의를 가진다. 또한 이러한 분석 결과는 대한민국을 포함한 후발국이 자국의 제도적/기술적 여건을 고려한 효율적인 연구개발 전략을 수립하는 데 참고 자료로 활용될 수 있을 것으로 기대된다.

사사

이 논문은 2024년도 정부(원자력안전위원회)의 재원으로 사용후핵연료관리핵심기술개발사업단 및 한국원자력안전재단의 지원을 받아 수행된 연구사업임(RS-2021-KN066110).

References

- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003, Latent Dirichlet allocation, *The Journal of Machine Learning Research*, 3, 993-1022.
- Bouma, G., 2009, Normalized (pointwise) mutual information in collocation extraction, In: Chiarcos, C., Eckart de Castilho, R., Stede, M. (Eds.), *From Form to Meaning: Processing Texts Automatically*, Proceedings of the Biennial GSCL Conference 2009, Gunter Narr Verlag, Tübingen, 31-40.
- Geier, J.E., 2012, Hydrogeological conditions at the Forsmark site, Technical Note 2012:41, Swedish Radiation Safety Authority (SSM), 34p.
- Griffiths, T.L., Steyvers, M., 2004, Finding scientific topics, *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5228-5235.
- IAEA (International Atomic Energy Agency), 2011, Geological disposal facilities for radioactive waste, IAEA Safety Standards Series No. SSG-14, 104p.
- Kim, T.Y., Lee, H.R., Yang, M.J., 2024, Analysis of research trends in *The Journal of Engineering Geology* (1991-2024): Latent Dirichlet allocation and network analysis, *The Journal of Engineering Geology*, 34(3), 429-445 (in Korean with English abstract).
- Liu, L.L., Zhang, J., Li, J.Z., Huang, F., Wang, L.C., 2022, A bibliometric analysis of the landslide susceptibility research (1999 – 2021), *Geocarto International*, 37(26), 14309-14334.
- MOTIE (Ministry of Trade, Industry and Energy), 2016, Basic plan for the management of high-level radioactive waste, 26p (in Korean).
- Park, C., Yong, T., 2017, Prospect of Korean nuclear policy change through text mining, *Energy Procedia*, 128, 72-78.
- Posiva Oy, 2015, Posiva is granted construction licence for final disposal facility of spent nuclear fuel, Retrieved from https://www.posiva.fi/en/index/news/pressreleasesstockexchangereleases/2015/posiva_is_granted_construction_licence_for_final_disposal_facility_of_spent_nuclear_fuel.3225.html.

- Röder, M., Both, A., Hinneburg, A., 2015, Exploring the space of topic coherence measures, In: Cheng, X., Li, H., Gabrilovich, E., Tang, J. (Eds.), Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Association for Computing Machinery, New York, NY, 399-408.
- SKB (Swedish Nuclear Fuel and Waste Management Co.), 2022, The government approves SKB's final repository system, Retrieved from <https://skbinternational.se/nyhet/the-government-approves-skbs-final-repository-system/>.
- STUK (Radiation and Nuclear Safety Authority), 2015, STUK's statement and safety assessment on the construction of the Olkiluoto encapsulation plant and disposal facility for spent nuclear fuel, STUK-B 196, 91p.