



Received: 2026/08/21
Revised: 2026/09/02
Accepted: 2026/09/16
Published: 2026/09/30

***Corresponding Author:**

Hye-Suhn Chang
Tel: +82-55-751-4649
E-mail: haesun@krit.re.kr

Abstract

본 연구는 2021년부터 2025년까지 발간된 국방기술기획서의 로드맵에 대한 연도별 변화를 텍스트 유사도 분석을 통해 정량적으로 제시하였다. 이를 위해 명사 기반 단어 빈도·역문서 빈도 가중치, 문자 n-gram, 상위 키워드 자카드 유사도 지표를 산출하였으며, 분석 결과 로드맵이 매년 일정한 방식으로 개정되기보다는 특정 시점에 변화가 집중된다는 점을 확인하였다.

This study quantitatively analyzed year-over-year changes in the roadmap chapters of the Defense Technology Planning Report from 2021 to 2025 using text similarity analysis methods. For this purpose, noun-based term frequency-inverse document frequency (TF-IDF), character n-gram, and Jaccard similarity of top keywords were computed. The results showed that changes were not evenly distributed across years but underwent changes concentrated in specific periods.

Keywords

국방기술기획서(Defense Technology Planning), 로드맵(Roadmap Analysis), 텍스트 유사도(Text Similarity), 텍스트 마이닝(Text-Mining), TF-IDF, 코사인 유사도(Cosine Similarity), 자카드 유사도(Jaccard Similarity),

국방기술기획서 로드맵의 연도별 변화에 관한 유사도 기반 분석

A Similarity-Based Analysis of Year-over-Year Changes in the Defense Technology Planning Report Roadmap

장혜선*

국방기술진흥연구소 함정기술팀 연구원

Hye-Suhn Chang*

Researcher, Korea Research Institute for Defense Technology Planning and Advancement

1. 서론

국방기술기획서는 국방 기술 분야에 대한 중·장기 기술개발 방향과 기술확보 로드맵 형태로 담아 매년 발간되는 정책 문서이다. 이때 로드맵이 매년 실질적인 변화를 담아내는지, 아니면 전년도 로드맵을 그대로 답습하는데 그치는지는 국방기술기획 체계가 실질적으로 작동하고 있는지를 판단하는 근거가 된다. 그러나 국방기술기획서의 연도별 로드맵을 서로 비교·분석한 연구 사례는 없다.

기존 데이터 기반 국방기술 연구로는 특히 마이닝을 통한 기술트렌드를 분석한 사례[1]나, 국방기술기획서의 국방전략기술과 세부 기술 항목을 대상으로 기술 구조를 분석한 사례[2] 등이 있다. 본 연구는 2021년부터 2025년까지 5년 간 발행된 국방기술기획서의 로드맵을 서로 비교하고, 이들 간 유사도를 텍스트 마이닝 기반의 정량 분석을 통해 제시하고자 한다.

2. 이론적 배경

국방기술기획서는 「국방과학기술혁신 촉진법」 및 「국방과학기술혁신 기본계획」에 근거해 매년 작성되는 정책 문서로 향후 국방과학기술 개발 방향과 기술확보 로드맵, 관련 과제 등을 제시한다. 공개된 국방기술기획서는 총론, 국방전략기술 기반 국방기술

확보전략(로드맵), 국방기술과제 추진계획, 과제카드 등으로 구성된다.

본 연구는 이 중 국방전략기술 기반 국방기술 확보 전략, 즉 로드맵을 2021년부터 2025년까지 5개년에 걸쳐 분석한다. 해당 로드맵이 기획서 일반본에 처음 제시된 것이 2021년에 발간된 '21-’35 핵심기술기획서이기 때문이다. ‘핵심기술기획서’는 2022년 ‘국방기술기획서’로 명칭이 변경된 이후 현재의 ‘25-’39 국방기술기획서로 이어져오고 있다([3], p. 364; [4], p. 422; [5], p. 341; [6], p. 252; [7], p. 258).

다만 해당 5년 동안 앞서 언급한 로드맵의 구성 방식에는 몇 차례 변화가 있었다. Fig. 1과 같이 기존의 ‘8대 전략기술 분야’가 2023년부터 ‘10대 국방전략기술 분야’로 개편되었으며, 로드맵의 시기 역시 연도에 따라 변화가 있었다. 로드맵의 연도별 변화는 이어지는 Table 1에 정리하였다.

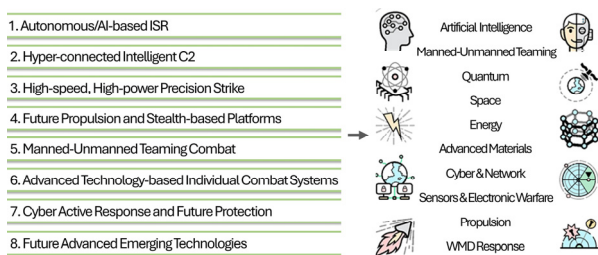


Fig. 1. Reorganization of strategic technology domains

Table 1. Changes in roadmap time-horizon structure

Time Bin	2021	2022	2023	2024	2025
Near-term	Short/Mid	Short/Mid	Short/Mid	Short	Short
Mid term		(~F+7)	(~F+5)		Mid
Far-term		Long/Long+	Long		Long
		(F+8~)	(~F+10)		
Far-term (extended)			Long+ (F+15~)		-

이와 더불어 로드맵에서 기술별 정보를 표기하는 방식도 2025년에 바뀌었다. 2024년까지는 기술별로 관련 과제가 매핑되는 구조였던 반면, 2025년에는 기술별로 세부목표가 직접 서술되는 구조로 바뀐 것이다. 이러한 변화가 유사도 결과에 미쳤을 영향은 결론에서 다시 논의한다. 아울러 본 연구는 공개된 일반본만을 분석 대상으로 하였으며, 비밀로 분류된 세부 내용은 연구 범위에 포함하지 않았다.

3. 연구 방법론

3.1 연구 방법

분석 대상은 2021년부터 2025년까지 발간된 국방기술기획서의 로드맵(제2장)이며, 모두 공개 자료이다. 앞서 살펴본 바와 같이 다섯 개 연도는 전략기술 분류와 로드맵의 시기 기준 등이 서로 상이하므로 개별 기술이나 분야, 시기 구간을 연도별 로드맵 간의 동일한 조건으로 둘 수 없다. 이에 본 연구는 세부 항목 간의 유사도를 분석하는 대신 로드맵 장 전체를 하나의 단위로 삼아 연도 간 유사도를 산출하는 방식을 택하였다.

3.2 분석 방법

연도별 기획서의 유사도를 산출하기 위하여 각각의 로드맵 장의 서술 전체를 하나의 텍스트로 집계하였다. 총론이나 과제카드 등 로드맵 외의 부분은 연도와 상관없이 같은 상용구가 반복되어 유사도를 왜곡할 수 있으므로, 각 문서의 내부 페이지 번호를 이용해 로드맵 장 범위만을 정확히 추출하였다. 이렇게 추출한 텍스트 중 페이지 번호, ①, ②와 같은 항목 번호 등 문서 형식에서 비롯된 부분은 정규식 기반으로 제거하였다. 이어지는 단계를 포함한 전체적인 분석 절차에 대해서는 Fig. 2에 정리하였다.

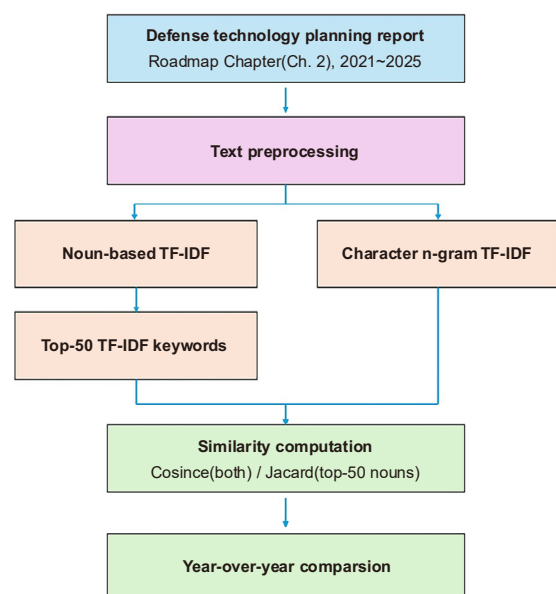


Fig. 2. Analysis process

3.2.1 명사 기반 TF-IDF 유사도

연도별 로드맵 간의 유사도를 측정하기 위해 먼저 각 문서에서 어떤 단어가 해당 문서를 대표하는지를 식별해야 한다. ‘기술’, ‘개발’ 처럼 5년간 로드맵에 반복해서 등장하는 단어는 특정 연도 로드맵의 특징을 보여주지는 못하지만, 어느 한 해의 로드맵에 한정하여 언급되는 단어는 해당 연도의 로드맵을 차별화해주는 단어라 할 수 있다. 본 연구는 이러한 단어의 중요도를 계량화하기 위해 명사 기반 TF-IDF(term frequency-inverse document frequency, 이하 TF-IDF)를 적용하였다.

TF-IDF는 한 문서 안에서 특정 단어가 얼마나 자주 등장하는지(TF)와, 그 단어가 전체 문서를 통틀어 얼마나 드물게 나타나는지(IDF)를 곱해 가중치를 매기는 방식으로, 문서 d 내 단어 t 에 대한 TF-IDF 가중치는 아래 수식 (1)과 같이 정의된다.

$$TF-IDF(t, d) = tf(t, d) \times \log\left(\frac{N}{df(t)}\right) \quad (1)$$

Where, $tf(t, d)$ denotes the frequency of term t in document d , N denotes the total number of documents, and $df(t)$ denotes the number of documents containing term t .

이 정의에 따라 5년간 공통으로 나오는 일반 어휘는 $df(t)$ 가 커서 가중치가 낮아지고, 특정 연도 로드맵에만 자주 등장하고 다른 연도에는 드물게 나타나는 용어는 가중치가 높아진다. 이렇게 산출된 가중치를 이용해 각 연도의 로드맵을 하나의 벡터로 표현할 수 있다. 로드맵에 등장하는 단어 각각이 좌표축이 되고, 그 단어의 TF-IDF 가중치가 좌표값이 되어 각 연도의 로드맵은 이 다차원 공간 위의 벡터로 표현된다. 두 연도의 벡터 A , B 사이의 유사도는 아래 수식 (2)와 같이 코사인 유사도로 계산하였다.

$$\cos(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (2)$$

Where, A and B denote the TF-IDF weight vectors of two roadmaps in a given year pair, $A \cdot B$ denotes the dot product of A and B , and $\|A\|$ and $\|B\|$

$\|A\|$ denote the Euclidean norms (magnitudes) of A and B , respectively.

코사인 유사도는 두 벡터 사이 각도의 코사인 값으로 1에 가까울수록 두 문서의 단어 구성 비중이 유사함을 의미하며, 0에 가까울수록 상이함을 뜻한다. 이 값은 벡터의 방향만 보고 문서의 분량과 같은 길이는 반영하지 않으므로 분량이 상이한 연도별 문서를 비교할 때도 왜곡이 적다. 이 방식은 서론에서 소개한 [2]에서 다른 방법론과 같다.

다만 이렇게 산출되는 유사도의 절대값은 로드맵 전반에 공통으로 쓰이는 어휘의 영향을 받으므로, 수치 자체보다는 연도 간 상대적인 비교로 해석하는 것이 적절하다.

3.2.2 문자 n-gram 기반 TF-IDF 유사도

명사 기반 방식은 형태소 경계를 기준으로 단어의 사용 빈도를 보기 때문에 같은 단어를 쓰더라도 문장 구조나 서술 방식이 크게 바뀔 경우는 잘 잡아내지 못한다. 이를 보완하기 위해 단어가 아니라 연속된 글자 n 개 단위(n -gram, $n=2\sim4$)로 텍스트를 나누어 같은 TF-IDF·코사인 유사도 절차를 적용하였다. n 이 1이면 개별 글자 하나의 빈도만 반영하여 의미있는 단위가 되기 어렵고, n 이 지나치게 크면 사실상 단어 단위와 다르지 않아 명사 기반 방식과 차별성이 사라지기 때문이다. 예를 들어 ‘무인체계’라는 글자는 2-gram 기준으로 {무인, 인체, 체계}, 3-gram 기준으로 {무인체, 인체계}로 나뉘며, $n=2$ 부터 $n=4$ 까지 생성되는 글자 조각을 각각 따로 산출해 평균하는 대신 하나의 벡터에 함께 담아 여러 수준의 변화를 동시에 반영할 수 있도록 하였다.

이 방식은 어순이나 조사, 문장 표현의 변화에도 반응하므로 같은 용어를 재사용하면서도 서술 자체를 새로 쓴 경우를 명사 기반 방식보다 낮은 유사도로 잡아낸다. 즉 명사 TF-IDF가 ‘다루는 용어가 유사한가’를 보는 지표라면, 문자 n -gram TF-IDF를 통해 ‘표현 자체가 유사한가’를 보는 지표에 가깝다. 두 지표를 함께 보면 전년도 대비 달라진 부분이 용어 교체에 그쳤는지 문장 서술까지 다시 쓰였는지를 구분할 수 있다.

산출된 두 벡터 집합에 대해 전년도 대비 유사도를

계산하였다. 유사도가 1에 가까울수록 서술이 유지되었음, 낮을수록 내용이 그만큼 달라졌음을 뜻한다. 아울러 2023년부터 2025년까지는 동일한 10대 국방 전략기술 분야 체계를 따르고 있으므로, 이 구간에 한해 분야 단위의 유사도도 함께 분석하였다.

3.2.3 상위 키워드 집합 자카드 유사도

TF-IDF 코사인 유사도는 단어의 상대적 가중치를 반영한 벡터 방향의 유사성을 측정하는 방식이다 보니 상위 몇몇 핵심 용어의 비중만 유지되면 나머지 어휘 구성이 상당히 바뀌어도 유사도가 높게 산출될 수 있다는 한계가 있다. 이를 보완하기 위해 코사인 유사도 계산에 사용한 것과 동일한 TF-IDF 벡터에서 가중치가 높은 상위 50개 명사 키워드만을 뽑아 집합으로 변환하고, 두 연도의 상위 키워드 집합 X, Y 사이의 자카드 유사도(Jaccard similarity)를 아래 수식 (3)과 같이 추가로 산출하였다.

$$J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad (3)$$

Where, X and Y denote the sets of the top-50 TF-IDF-weighted keywords for two roadmaps in a given year pair.

이는 두 집합의 교집합 크기를 합집합 크기로 나눈 값으로 가중치 크기는 반영하지 않고 두 연도의 대표 어휘 목록이 얼마나 그대로 유지되는지만을 본다. 코사인 유사도가 ‘핵심 용어의 비중이 비슷한가’를 보는 지표라면 자카드 유사도는 ‘대표 어휘 목록 자체가 얼마나 바뀌었는가’를 보는 지표에 가깝다. 그만큼 자카드 유사도는 코사인 유사도보다 어휘 교체에 더 민감하게 반응한다.

4. 분석 결과

4.1 연도별 로드맵 유사도

연도 간 로드맵 유사도는 Table 2와 같으며 세 가지 산출 방법 모두에서 5년간의 유사도가 매년 다르게 나타났다. 표의 각 열은 해당 연도와 직전 연도를 비교한 값이다. 이 때 ‘2023’열은 2022년 로드맵과

2023년을 비교한 결과를 의미한다. 또한 Table 2의 자카드 유사도 계산에 사용된 상위 50개 TF-IDF 키워드 집합에서 연도별로 빠지거나 새로 등장한 용어 일부는 이어지는 Table 3에 정리하였으며, 전체 목록은 지면 제약상 별도로 제시하지 않았다.

Table 2. Roadmap content similarity (2021~2025)

Method	2022	2023	2024	2025
Noun-based TF-IDF	1.00	0.92	0.90	0.99
Character n-gram TF-IDF	0.99	0.83	0.51	0.95
Jaccard Similarity	0.82	0.45	0.59	0.79

Table 3. Top-50 TF-IDF keyword changes by year

Year	Dropped Keywords	New Keywords
2021 → 2022	Future, Real-time, Guidance, Battlefield	UAV, Autonomous, Application, Optimal
2022 → 2023	Radar, Underwater, Robot, Autonomous, Control, Tracking, Identification, Imagery	Quantum, Space, Cyber, Network, Materials, Laser, Battlefield, Platform
2023 → 2024	UAV, Target, Threat, Signal, Situation, Recognition, Fusion, Platform	Energy, Electronic, Precision, Output, Response, Detail, Level, Research
2024 → 2025	Implementation, Multiple, Stage, Small-scale, System, Trend	Capability, Data, Analysis, Underwater, Ground, Support

2022년은 세 지표 모두에서 가장 높은 값(1.00, 0.99, 0.82)을 보였다. 이는 이 해의 로드맵이 전년도 내용을 가장 많이 이어받은 결과로 보인다. 반면 2023년과 2024년은 세 지표 모두에서 상대적으로 낮은 값을 보였다.

2023년은 국방전략기술 분야가 기존 8대 분야에서 10대 분야로 개편된 해로 자카드 유사도가 0.45로 4개년 중 가장 낮았다. 이는 2022년과 2023년 로드맵에 각각 등장한 상위 50개 키워드 가운데 겹치는 것이 31개에 그쳤다는 뜻으로, 분야 개편에 맞추어 로드맵을 대표하는 용어들이 큰 폭으로 교체되었음을 보여준다. 실제로 상위 TF-IDF 키워드를 살펴보면 2023년에 새로 등장한 용어 가운데 양자, 우주, 사이버 등이 해당 연도에 개편된 전략기술의 명칭과 그대로 맞닿아 있다는 것을 확인할 수 있다. 다만 명사 TF-IDF(0.92)와 문자 n-gram(0.83)은 상대적으로

높게 유지되었는데, 이 두 지표는 상위 몇 개 용어만이 아니라 문서 전체에 걸친 수천 개 단어의 사용 비중 분포를 종합적으로 비교하기 때문이다. 이는 로드맵을 대표하는 주요 용어는 새로운 분야 체계에 맞추어 바뀌었지만 문서 전반의 어휘 사용 패턴과 서술 방식 자체는 비교적 유지되어 코사인 유사도는 상대적으로 높게 산출된 것으로 보인다.

2024년에는 전년도와 반대의 양상을 보였다. 로드맵의 시기구분 표기가 바뀐 해로, 문자 n-gram 유사도가 0.51로 4개년 중 가장 낮았다. 반면 자카드 유사도는 0.59로 2023년보다 오히려 높아져, 새로 등장한 용어의 비중이 2023년만큼 크지는 않았음을 보여준다. 즉 다루는 용어 자체는 전년도와 크게 다르지 않았으나, 문장 표현이 상당 부분 새로 쓰이면서 문자 수준의 유사도가 가장 크게 떨어진 것으로 해석된다.

2025년은 앞선 두 해와 다른 양상을 확인할 수 있다. 이 해는 로드맵에서 기술별 정보를 표시하는 방식이 과제 매핑에서 세부목표 서술로 바뀐 해이다. 그럼에도 2025년의 유사도는 세 지표 모두 2022년 다음으로 높은 값을 보여, 로드맵의 형식은 바뀌었지만 실제 서술 내용은 전년도와 상당히 비슷하게 유지되었음을 보여준다.

4.2 분야별 로드맵 유사도

2023년부터 2025년까지의 국방기술기획서는 모두 10대 국방전략기술 분야를 기준으로 로드맵을 구성하고 있어, 이 구간은 분야 단위로 세분화된 비교가 가능하다. 분야별 비교에는 지면 제약을 고려하여 세

Table 4. Domain-level similarity of roadmap descriptions

Domain	2023→2024	2024→2025
Artificial Intelligence	0.84	0.97
Manned-Unmanned Teaming	0.89	0.97
Quantum	0.89	0.97
Space	0.88	0.98
Energy	0.90	0.97
Advanced Materials	0.83	0.95
Sensors & Electronic Warfare	0.80	0.95
Propulsion	0.87	0.96
WMD Response	0.83	0.96
Cyber & Network	0.85	0.97
Average	0.86	0.97

지표 중 문서 전체의 어휘 사용 비중을 종합적으로 반영하는 명사 TF-IDF를 대표 지표로 사용하였으며, 그 결과는 Table 4에 정리했다.

분야별로 나누어 보면 2023년에서 2024년으로 넘어가며 서술이 상당 부분 새로 쓰이고 2024년에서 2025년으로 넘어갈 때는 대부분 그대로 유지되었다는 앞선 결과가 10개 분야 전체에서 동일하게 나타난다.

한편 2023년에서 2024년으로 바뀌는 정도는 분야마다 달랐다. 센서·전자기전(0.80)과 첨단소재·WMD 대응(0.83), 사이버·네트워크(0.85) 분야는 상대적으로 많이 바뀐 편이었던 반면, 에너지(0.90)와 유·무인 복합·양자(0.89) 분야는 전년도 서술이 비교적 많이 남아 있었다. 이러한 분야 간 편차(0.80~0.90)는 앞서 살펴본 2021년에서 2022년으로 넘어가는 시기와는 다른 모습이다. 그때는 8개 분야 전체에서 명사·문자 유사도가 0.98 이상으로 고르게 나타나 분야와 상관 없이 대부분의 내용이 유지되었다. 반면 2023년에서 2024년으로 넘어갈 때는 일부 분야에서만 서술이 크게 바뀌었다.

5. 결론

5.1 주요 연구 결과 및 시사점

본 연구를 통해 확인한 사실은 국방기술기획서 로드맵이 매년 고르게 개정되는 것이 아니며 특정 시기에 크게 바뀐다는 것이다. 2021년부터 2025년까지 다섯 해 사이의 네 차례 변화 중 두 번은 전년도의 서술이 상당 부분 그대로 유지되었고 나머지 두 번에 큰 폭의 변화가 집중되었다. 이는 로드맵이 늘 그대로 답습되기도, 매년 새롭게 바뀌지도 않는다는 것을 보여준다.

변화가 집중된 두 시기는 서로 성격이 다르다. 2022년에서 2023년으로 넘어가며 기존 임무 중심의 8대 국방전략기술 분야가 기술 중심의 10대 분야로 개편되었는데, 이는 로드맵이 기술 분류의 틀 자체를 바꾼 것이다. 반면 2023년에서 2024년으로 넘어갈 때는 이 틀은 그대로 둔 채 서술 방식과 세부 구조가 다시 쓰였으며, 문자 수준 유사도가 0.51까지 떨어진 것이 이를 뒷받침한다. 즉 분류 틀을 바꾼 시기와 내용을 다시 쓴 시기가 서로 달랐다. 이어지는 2024년에서 2025년으로 넘어가는 시기는 기술별 정보를 담

는 방식이 과제 매핑에서 세부목표 서술로 바뀌었음에도, 세 지표 모두 가장 높은 유사도를 보였다.

5.2 한계 및 향후 연구 계획

본 연구는 몇 가지 한계를 지닌다. 우선, 로드맵 장 전체를 하나로 묶어 비교하기 때문에 분야나 기술 하나하나가 새로 생기거나 사라지는 것까지는 잡아내지 못한다. 다만 2023~2025 구간에 대해 수행한 분야별 분석이 이를 어느정도 보완한다.

더불어 본 연구가 다루는 자료가 매년 같은 조건에서 나오지 않는다는 점도 감안해야 한다. 국방기술기획서는 상위 계획을 근거로 매년 작성되는데, 상위 근거와의 연계 방식이나 반영되는 전략기술은 해마다 달라질 수 있다. 또한 로드맵을 서술하는 방식 자체도 고정되어 있지 않아서, 2024년에는 기술마다 관련 과제가 매핑되는 구조였다가 2025년에는 기술별 세부목표가 직접 서술되는 구조로 바뀌는 등 문서 형식이 변화하였다. 이러한 상위 근거와 서술 양식의 변화는 본 연구가 측정하는 유사도에 내용상의 실제 변화와 함께 반영되었을 가능성이 있어, 관찰된 유사도 변화를 로드맵 내용의 변화로만 해석하는 데에는 신중할 필요가 있다.

또한 본 연구에서 사용된 TF-IDF 및 자카드 유사도는 같은 개념이더라도 표기가 다르면 별개의 단어로 처리한다는 한계가 있다. ‘인공지능’과 ‘AI’처럼 동의어가 혼용된 경우가 이에 해당하며, 실제로 유사

도가 낮게 산출될 가능성이 있다.

이러한 한계를 보완하기 위해 후속 연구로는 연도별로 상이한 세부기술을 서로 대응시키는 크로스워크를 구축하여 개별 기술의 신규 생성과 소멸을 직접 추적하는 작업이 필요하다. 아울러 로드맵의 양식이 일치하는 구간별로 분석을 세분화하거나, 양식에 따른 정형화된 표현을 식별해 유사도 계산에서 구분함으로써, 자료의 배경 조건 변화와 실제 내용 변화를 분리하는 방법도 검토할 필요가 있다. 이와 함께 표기 변형에 따른 유사도 왜곡을 줄이기 위해 사전학습 임베딩 기반의 의미 유사도 측정을 병행하는 방안에 대해서도 향후 검토가 필요하다. 이를 통해 국방기술기획이 어떻게 개정되는지를 더 정밀하게 밝힐 수 있을 것이다.

참고문헌

- [1] Kim Kyung-Soo & Cho Nam-Wook, ‘A Study on Networks of Defense Science and Technology using Patent Mining,’ Journal of Korean Society for Quality Management, VOL. 49, NO. 1, 2021, pp. 97-112.
<https://doi.org/10.7469/JKSQM.2021.49.1.97>
- [2] 정대현·정경은·권오영, ‘미래 국방전략기술 예측 네트워크 분석: 국방기술기획서를 중심으로,’ 국방정책연구 제42권 제1호, 2026, pp.191-218. <https://doi.org/10.22883/jdps.2026.42.1.007>
- [3] 방위사업청. ‘21-’35 핵심기술기획서, 2021.
- [4] 방위사업청. ‘22-’36 국방기술기획서, 2022.
- [5] 방위사업청. ‘23-’37 국방기술기획서, 2023.
- [6] 방위사업청. ‘24-’38 국방기술기획서, 2024.
- [7] 방위사업청. ‘25-’39 국방기술기획서, 2025.