Journal of Nanotechnology and Innovative Materials

Research Article

Creating a Term Database for Patentability Search in the Field of Nanotechnology

Alikhan I. Rzaguliev

Department of Information Technologies, School of Advanced Technologies and Innovation Engineering, Western Caspian University, Ahmad Rajabli 17A, III Parallel, AZ1001 Baku, Azerbaijan

Received: 25.04.2026 Accepted: 26.05.2026 Published: 29.05.2026

https://doi.org/10.54414/WUSO3256

Abstract

The present study examines the design of a multilingual database of technical terms that could be used in automated searches for patents in the area of nanotechnologies. Patents in the field of nanotechnology present several challenges due to terminology issues, documentation in several languages, and different terminological expressions used in science. The database suggested includes the organization of technical terminology into noun-based categories in three languages (Azerbaijani, English, and Russian) and also incorporates synonym management techniques. The study explains the use of structured vocabularies of technical terms for automated patent searching and presents the development of a multilingual database in order to aid semi-automatic patentability evaluation.

Keywords: patentability search, nanotechnologies, automation, knowledge-based systems, decision support, technical database

1. Introduction

The patentability search in the domain of nanotechnologies faces a number of difficulties associated with the specificity of technical terms, high pace of technological changes, and multilinguality of patents. Inventions related to nanotechnologies are commonly represented using certain scientific language, including scientific expressions, acronyms, and specific technical terms. Thus, the traditional keyword-based patent search may fail to identify the relevant prior-art documents if synonyms and other linguistic representations of technical terms are used.

Meanwhile, modern patent offices apply more and more automatic and semi-automatic search tools in order to increase efficiency and facilitate patent examination. Still, the success of such searches heavily depends on the structure and content of technical term databases used to generate search queries and semantic analysis. Current universal dictionaries are not always effective in relation to particular domains, e.g., nanotechnologies.

The goal of this study is to elaborate a structured multilingual technical term database for automated patentability search in the nanotechnology domain.

2. Literature Review

Currently, in the Patent and Trademark Centre under the Intellectual Property Agency of the Republic of Azerbaijan, patent search is conducted in multiple databases. In each database, the examiner conducts a manual search using Boolean Algebra operations (such as AND, OR, NOT, etc). The databases include: Azerbaijani Patent database where examiner searches in Azerbaijani and Russian [1], FIPS where the examiner searches in (mainly for Azerbaijani practice) Russian [2], EAPATIS where the examiner searches in (mainly for Azerbaijani practice) Russian [3], PATENTSCOPE where the examiner searches in (mainly for Azerbaijani practice) English [4], Espacenet where the examiner searches in (mainly for Azerbaijani practice) English [5],

USPTO where the examiner searches in (mainly for Azerbaijani practice) English [6], Google Patents where the examiner searches in (mainly for Azerbaijani practice) English [7].

According to the patent law of the Republic of Azerbaijan, the subjects of patents can be inventions, utility models, and industrial designs. In the Republic of Azerbaijan, utility models can only be devices. While inventions must satisfy three patentability requirements, utility models are required to satisfy only two criteria. Those are novelty and industrial applications. The devices for novelty conditions must-have new characteristics. If the device can be made or used in any field of industry and economy, then the invention is considered applicable [8].

Unlike some other national or regional patent legislations, in the Republic of Azerbaijan, examiners also provide supplementary examinations for utility models. The publication of the utility model occurs only after the patent search. Depending on the patent search, the examiner decides whether to grant or reject the application [8].

The area of patent retrieval and automated patent searching has gained significance because of the fast expansion of patent documentation across the globe. Currently, research efforts in the field mainly concentrate on semantic similarity, natural language processing, and machine learning techniques to improve patent searching.

Another study proposes an automated patent search approach by using machine learning and NLP to compare the full texts of patents, contrary to manual keyword search. The approaches used by the authors in this study include vector-based similarity and Cosine similarity. They use two different datasets for their research: a dataset of 2.4 million cited or random pairs and an expert-annotated dataset to extract relevant documents. The authors represent documents in numeric vectors, including Bag-of-Words, word2vec, doc2vec, LSA, and KPCA [9].

Different sources describe patent information retrieval methods. Where the authors mention ways of semantic patent retrieval, transformer models, patent embeddings, etc. The authors use European Patent Office search reports as a training data sample [10].

Another study presents a model for improving search query accuracy on specialized websites using text correction and reconstruction. The authors train an intelligent model to correct user-entered queries and reconstruct them for improved retrieval results. According to the authors, the system was implemented in Go (Golang) for backend efficiency on concurrent operations. The model achieved a very high accuracy of 99.85% after 15 epochs of training, with a testing accuracy of about 88% on a validation dataset of heavily altered queries. The focus was on the query input stage of search systems for improving user query quality [11].

The article addresses the important challenge of query reconstruction for specialized retrieval systems. The paper does not cover the downstream stages of search execution, similarity scoring, multilingual input handling, or classification. Sometimes, for conducting patentability, search logic might be required. And logic-based patentability evaluation is absent in this paper. The current research differs from the existing ones because it concentrates on creating a multilingual technical term database for automated patentability analysis.

In the proposed automated patentability search system, patent retrieval is performed using a multi-step search model. Unlike the previously proposed model, which employed a binary (0–1) scoring scale, the present approach adopts a 0–10 scoring scale to provide a more intuitive interpretation of similarity scores. The three-step patent search framework is retained, while the definitions of the X and Y similarity labels are further clarified. In addition, the proposed model incorporates International Patent Classification (IPC) similarity into the overall similarity assessment. IPC similarity is evaluated on a 0–10 scale according to the hierarchical distance between IPC codes and is integrated into the similarity scoring functions (S_X and S_Y), requiring corresponding modifications to the original formulas [12].

$$S_Y = \frac{S_N + S_{noun} + S_{adj} + S_{IPC}}{4} \quad (1) \quad [12]$$

$$S_X = \frac{S_N + S_{noun} + S_{adj} + S_{IPC} + S_C}{5} \quad (2) \quad [12]$$

$$S_{IPC\ N} = CD \times 2 \quad (3) \quad [12]$$

$$S_{IPC} = IPC_1 \times IPC_{w1} + IPC_2 \times IPC_{w2} \ldots IPC_N \times IPC_{wn} \quad (4) \quad [12]$$

In Equations (3) and (4), S_{IPC} represents the similarity score of the International Patent Classification (IPC) codes. IPC similarity is determined according to the hierarchical distance between two IPC classifications. For example, the IPC code F16H 7/06 can be divided into hierarchical levels (F, F16, F16H, F16H7, and F16H7/06). The similarity score increases with the number of matching hierarchical levels, corresponding to values of 2, 4, 6, 8, and 10, respectively. For a single IPC code, $SIPC_n$ denotes the similarity score, where CD represents the number of common hierarchical levels. When multiple IPC codes are assigned to a patent, the overall IPC similarity (SIPC) is calculated as the weighted sum of the individual IPC similarity scores, where the weights satisfy the condition $w_1 + w_2 + \ldots + w_n = 1$ [12].

3. Databases

To simplify the analysis, only polymer-related nanotechnology patents were used as sample data. Sample noun and adjective keywords were selected from PATENTSCOPE with the search query "EN_TI:(Polymer AND nanotechnology)". The search returned four patent documents [13].

Sample data for configuration keywords were selected from published inventions or utility models from 2022 and onward in Azerbaijan [14].

Table 1. English noun keyword database.

Term №	Noun keyword	Term №	Noun keyword	Term №	Noun keyword
1	equipment	12	pMHC	22	D-Tartrate
2	multimer	13	modification	23	DBCO
3	DNA	14	SpyTag	24	nitride
4	Carbon	15	Emission	25	milliequiv
5	Composition	16	SpyCatcher	26	polystyrene
6	N- succinimidyl-S-acetylthiopropionate	17	glucocerebrosidase	27	oxide
7	pole	18	polycarbonates	28	carbide
8	pH	19	boride	29	polyacrylates
9	Aircraft	20	polymer	30	Isofagomine
10	polyethylene	21	alloy	31	polypropylene
11	N-succinimidyl-S-acetylthiopropionate				

Table 2. English noun adjuncts keyword database (nouns used as adjectives).

Term №	Noun ad. keyword	Term №	Noun ad. keyword	Term №	Noun ad. keyword
1	ratio	12	Drag	23	molding
2	salt	13	Flight	24	net
3	drug	14	Nanoparticle	25	C linking
4	fusion	15	clay	26	amino fluid
5	chain	16	coating	27	metal
6	capture	17	solvent	28	emulsifying
7	Surface	18	protein	29	Cream
8	Powder	19	strand	30	Liquid
9	Nanocomposite	20	molecular chaperone	31	micron
10	Agent	21	Fuel	32	plastic
11	Value	22	Injection	33	water

Table 3. English adjective keyword database.

Term №	Adj. keyword	Term №	Adj. keyword	Term №	Adj. keyword
1	PE-modified	7	insulating	13	Molar
2	rectangular	8	organic	14	polyurethane
3	magnetic	9	silanol	15	conducting
4	submicron	10	reactive	16	non-destructive
5	Invisible	11	Aprotic	17	Nanoscale

6	electrically	12	HLB

Table 4. English configuration keyword database.

Term №	Configuration keywords	Term №	Configuration keywords	Term №	Configuration keywords
1.	Installed	2.	Connected	3.	Passes through the center
2.	Mounted	4.	Opposite	6.	Fixed
3.	Settled	6.	Lined up	9.	Inclusive
4.	Placed	8.	Attached	12.	Linked

Table 5. Azerbaijani noun keyword database.

Term №	Noun keyword	Term №	Noun keyword	Term №	Noun keyword
1	avadanlıq	12	pMHC	22	D- tartrat
2	multimer	13	modifikasiya	23	DBCO
3	DNT	14	SpyTag	24	nitrid
4	Karbon	15	emissiya	25	milliekviv
5	Tərkib	16	SpyCatcher	26	polistirol
6	N- suksinimidil-S-asetiltiopropionat	17	qlükoserebrozidaza	27	oksid
7	Qütb	18	polikarbonatlar	28	karbid
8	pH	19	borid	29	poliakrilatlar
9	Təyyarə	20	polimer	30	izofagomin
10	polietilen	21	ərinti	31	polipropilen
11	N- suksinimidil-S-asetiltiopropionat				

Table 6. Azerbaijani noun adjuncts keyword database (nouns used as adjectives).

Term №	Noun ad. keyword	Term №	Noun ad. keyword	Term №	Noun ad. keyword
1	Nisbət	12	Sürükləmə	23	qəlibləmə
2	duz	13	Uçuş	24	tor
3	dərman	14	Nanohissəcik	25	C birləşdirici
4	birləşmə	15	Gil	26	amino mayesi
5	zəncir	16	Örtük	27	metal
6	tutma	17	Həlledici	28	emulsiya
7	Səth	18	zülal	29	krem
8	Toz	19	Tel	30	maye
9	Nanokompozit	20	molekulyar şaperon	31	mikron
10	Agent	21	Yanacaq	32	plastik
11	Dəyər	22	inyeksiya	33	su

Table 7. Azerbaijani adjective keyword database.

Term №	Adj. keyword	Term №	Adj. keyword	Term №	Adj. keyword
1	PE ilə modifikasiya edilmiş	7	izolyasiyaedici	13	molar
2	düzbucaqlı	8	üzvi	14	poliuretan
3	maqnit	9	silanol	15	keçirici
4	submikron	10	reaktiv	16	dağıdıcı olmaya
5	görünməz	11	aprotik	17	nanoölçü
6	elektrik	12	HLB		

Table 8. Azerbaijani configuration keyword database.

Term №	Configuration keyword	Term №	Configuration keyword	Term №	Configuration keyword
1.	Quraşdırılımış	2.	Qoşulub	3.	Mərkəzindən keçir
2.	Oturdulmuş	4.	Əks tərəf	6.	Sabitlənmiş
3.	Bərkidilmiş	6.	Düzülüb	9.	Daxil edilib
4.	Yerləşən	8.	Bitişik	12.	Birləşib

Table 9. Russian noun keyword database.

Term №	Noun keyword	Term №	Noun keyword	Term №	Noun keyword
1	оборудование	12	пМГК	22	D-тартрат
2	мультимер	13	модификация	23	DBCO
3	ДНК	14	SpyTag	24	нитрид
4	Углерод	15	эмиссия	25	миллиэквивалента
5	Состав	16	SpyCatcher	26	полистирол
6	N-сукцинимидил-S-ацетилтиопропионат	17	глюкоцереброзидаза	27	оксид
7	полюс	18	поликарбонатов	28	карбид
8	pH	19	борид	29	полиакрилатов
9	Самолет	20	полимер	30	Изофагомин
10	полиэтилен	21	сплав	31	полипропилен
11	N-сукцинимидил-S-ацетилтиопропионат				

Table 10. Russian noun adjuncts keyword database (nouns used as adjectives).

Term №	Noun ad. keyword	Term №	Noun ad. keyword	Term №	Noun ad. keyword
1	соотношение	12	Сопротивление	23	формование
2	соль	13	Полет	24	сетка
3	лекарство	14	Наночастица	25	С-связывание
4	слияние	15	глина	26	аминожидкости
5	цепь	16	покрытие	27	металл
6	захват	17	растворитель	28	эмульгирующий
7	Поверхность	18	белок	29	Крем
8	Порошок	19	нить	30	Жидкость
9	Нанокомпозит	20	молекулярный шаперон	31	микрон
10	Агент	21	Топливо	32	пластик
11	Значение	22	инъекция	33	вода

Table 11. Russian adjective keyword database.

Term №	Adj. keyword	Term №	Adj. keyword	Term №	Adj. keyword
1	ПЭ-модифицированный	7	изоляционных	13	Молярный
2	полиуретан	8	органический	14	полиуретан
3	магнитный	9	силанол	15	проводящий
4	субмикронный	10	реактивный	16	неразрушающих
5	Невидимый	11	апротонный	17	наноразмер
6	электропроводящий	12	HLB		

Table 12. Russian configuration keyword database.

Term №	Configuration keywords	Term №	Configuration keywords	Term №	Configuration keywords
1.	Установлено	2.	Подключенный	3.	Проходит через центр
2.	Поставленный	4.	Напротив	6.	Фиксированный
3.	Закреплено	6.	Выстроены	9.	Включающий
4.	Размещено	8.	Прикрепленный	12.	Связанный

4. Limitations

Limitations associated with the use of the technical terms database include the manual development and management of technical terminology and synonyms for it. At present, the technical terms database contains only selected terminology related to nanotechnology, and other variations of technical terms that may be present in patent literature have not been included yet. Translation issues across multiple languages may additionally reduce effectiveness due to different semantic meanings of technical terms across various languages. Moreover, the technical terms database functions mainly as a database of keywords.

5. Conclusion

In order to demonstrate the importance of terminology structure for automated patentability search engines, the following suggested technical term database was designed. The multilingual technical keywords contribute to the consistency between search queries and found patents in various patent databases.

Within the structure of the database, it is possible to perform: grouping of noun-based technical terminology; multilingual terminology mapping from Azerbaijani, English, and Russian; inclusion of a synonym-based query expansion mechanism; development of a search query generator.

The suggested technical term database is capable of improving the quality of prior art searches within specialized technical fields such as nanotechnologies.

The constructed multilingual technical database revealed that terminology standardization is critical to automatic patent retrieval systems. While creating the database, it was found that there are a lot of different technical terms in patents related to polymer nanotechnology in Azerbaijani, English, and Russian languages. Creating structured noun-based categories helped to solve the problem of translating technical expressions literally and having many terminological variants of one expression.

This study suggests that the construction of a technical database could also help to build further semantic retrieval systems by providing structured terminology databases that can be used in machine learning or expert systems. Although now, the database is focused on organizing keywords and synonyms, in the future, the database can become a part of more complex systems such as similar patents detection systems and others.

Author Contributions

The author was responsible for the conceptualization of the study, development of the research methodology, investigation, analysis and interpretation of the findings, preparation and revision of the manuscript, and approval of the final version for publication.

Conflict of Interest

The author declares no conflicts of interest.

Funding

This research received no external funding.

Acknowledgment

The author acknowledges Western Caspian University for providing institutional support and academic resources that contributed to the completion of this research.

Abbreviations

Natural Language Processing (NLP), International Patent Classification (IPC).

References

[1] Intellectual Property Agency of the Republic of Azerbaijan. Azerbaijan Patent Database [database]. Baku (Azerbaijan): Intellectual Property Agency of the Republic of Azerbaijan; [cited 2026 April 20]. Closed-access database.

[2] Federal Institute of Industrial Property (FIPS). Information Retrieval System (IPS) [Internet]. Moscow (Russia): Federal Institute of Industrial Property; [cited 2026 April 20]. Available from: https://www.fips.ru/iiss/db.xhtml

[3] Eurasian Patent Organization (EAPO). Eurasian Patent Information System (EAPATIS) [Internet]. Moscow (Russia): Eurasian Patent Organization; [cited 2026 April 20]. Available from: https://www.eapatis.com/

[4] World Intellectual Property Organization (WIPO). PATENTSCOPE [Internet]. Geneva (Switzerland): World Intellectual Property Organization; [cited 2026 April 20]. Available from: https://patentscope.wipo.int/search/en/search.jsf
[5] European Patent Office (EPO). Espacenet [Internet]. Munich (Germany): European Patent Office; [cited 2026 April 20]. Available from: https://worldwide.espacenet.com/patent/search
[6] United States Patent and Trademark Office (USPTO). Patent Public Search [Internet]. Alexandria (VA): United States Patent and Trademark Office; [cited 2026 April 20]. Available from: https://ppubs.uspto.gov/pubwebapp/
[7] Google Patents. Patent database [Internet]. [cited 2026 April 20]. Available from: https://patents.google.com/
[8] Patent haqqında: Azərbaycan Respublikasının Qanunu. Maddə 8 [Internet]. Bakı: Azərbaycan Respublikası; 1997 [cited 2026 April 20]. Available from: https://patent.copat.gov.az/files//26290292722835030361312-IQ%20Patent%20haqq%C4%B1nda%20(2).pdf
[9] Helmers L, Horn F, Biegler F, Oppermann T, Müller KR. Automating the search for a patent's prior art with a full text similarity search. PloS one. 2019 Mar 4;14(3):e0212103. https://doi.org/10.1371/journal.pone.0212103
[10] Risch J, Alder N, Hewel C, Krestel R. Patentmatch: a dataset for matching patent claims & prior art. arXiv preprint arXiv:2012.13919. 2020 Dec 27. https://doi.org/10.48550/arXiv.2012.13919
[11] Simian D, Șerban ME. Improving search query accuracy for specialized websites through intelligent text correction and reconstruction models. Information. 2024 Nov 1;15(11):683. https://doi.org/10.3390/info15110683
[12] Rzaguliyev A. Automation of patentability for inventions using multi-step automated search model. In: Proceedings of the International Conference on Recent Advances in Computer Science and Information Technology; 2026.
[13] PATENTSCOPE. Patent database [Internet]. Geneva: World Intellectual Property Organization (WIPO); [cited 2026 April 20]. Available from: https://patentscope.wipo.int/search/en/search.jsf
[14] Azərbaycan Respublikası Əqli Mülkiyyət Agentliyi, Patent və Əmtəə Nişanlarının Ekspertizası Mərkəzi. Patent, faydalı model və sənaye nümunələri [Internet]. 2026 [cited 2026 Apr 20]. Available from: https://patent.copat.gov.az/publish-Inventions-utility-models-and-industrial-examples