
NOV 21, 2022

WORKS FOR ME 1

# Text mining approaches applied to patents: A scoping review protocol

COMMENTS 0

DOI
dx.doi.org/10.17504/protocols.io.6qpvr43kzgmk/v1

Homa Arshadi[1], Maryam Okhovati[2], Zohre Zahedi[3], Maryam Asharfi[4]

[1]Ph.D. Candidate in Medical Library and Information Science, Student Research Committee, School of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran.;
[2]Associate Prof., Medical Library and Information Sciences Department, School of Management and Medical Information Science, Kerman University of Medical Sciences, Kerman, Iran.;
[3]Assistant Professor of Information Science, Department of Information Science, Faculty of Humanities Persian Gulf University, Bushehr, Iran. Research, Centre for Science & Technology Studies (CWTS), Leiden University, The Netherlands;
[4]Assistant Professor, Department of Industrial Engineering and Management Systems, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran

Homa Arshadi

## ABSTRACT

This study will map the current state of different text mining approaches in patents. This will be of value to policy makers and researchers by allowing them to find the latest efforts to patent text mining. The further outcomes will be identifying the different text mining approaches, the most used data sources, metadata and subject areas in different application areas of the patents.
Web of Science Core Collection (Clarivate Analytics), Scopus, IEEE Xplore Digital Library, ACM Guide to Computing Literature digital will be searched to identify published studies. We will also search the reference lists of included papers. Keywords were founded by reviewing IEEE thesaurus, free text method, expert opinions, and the review of some relevant systematic review studies
After running search syntax in each database, the results of all the search will be exported to EndNote 20. Duplicate papers will be removed. The remaining papers will be imported to the Rayyan, for inclusion in review. Two independent research members will screen titles and abstracts of all papers against the inclusion and exclusion criteria. Subsequently the full-text of all potentially relevant papers will be assessed independently by two reviewers. In the case of conflicts, they discussed and then consulted the third author to reach a consensus.
Key findings relevant to the review will be charted from the included studies using a data extraction tool developed in Excel software by the members of the review team. Two review team members will extract data independently and discrepancies will be solved by consulting with a third expert.

## ATTACHMENTS

Patent text mining. Scoping review Protocol.pdf

DOI
dx.doi.org/10.17504/protocols.io.6qpvr43kzgmk/v1

PROTOCOL CITATION

Homa Arshadi, Maryam Okhovati, Zohre Zahedi, Maryam Asharfi 2022. Text mining approaches applied to patents: A scoping review protocol. **protocols.io** https://dx.doi.org/10.17504/protocols.io.6qpvr43kzgmk/v1

FUNDERS ACKNOWLEDGEMENT

Kerman University of Medical Sciences
Grant ID: 40100010

MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Arshadi, H., Okhovati, M., Zahedi, Z., & Ashrafi, M. (2022, November 16). Text mining approaches applied to patents: A scoping review protocol. https://doi.org/10.17605/OSF.IO/YNGHW

KEYWORDS

Patent, Text mining, Text mining Techniques, Patent mining, Scoping review



CREATED

Nov 16, 2022

LAST MODIFIED

Nov 21, 2022

PROTOCOL INTEGER ID

72810

ATTACHMENTS

Patent text mining. Scoping review Protocol.pdf

## Title

1 **Text mining approaches applied to patents: A scoping review protocol**

## Original language title



**Citation:** Homa Arshadi, Maryam Okhovati, Zohre Zahedi, Maryam Asharfi Text mining approaches applied to patents: A scoping review protocol https://dx.doi.org/10.17504/protocols.io.6qpvr43kzgmk/v1

2 English

## Stage of review at time of this submission

3

| A | B | C |
|---|---|---|
| **Review stage** | **Started** | **Completed** |
| Preliminary searches | Yes | Yes |
| Piloting of the study selection process | Yes | Yes |
| Formal screening of search results against elig | Yes | Yes |
| Data extraction | No | No |
| Data analysis | No | No |

4 This review is a part of a Ph.D. research project approved by the ethical committee of Kerman University of Medical Sciences, No. 40100010, which will be carried out with the financial support of the Vice Chancellor for Research and Technology of this university. The funding source had no involvement in the study process.

5 **Review question (s)**

RQ1. Which text mining techniques are frequently used by researchers in mining patent?
RQ2. Which data sources are the most often used for text mining in patents?
RQ3. Which metadata (Claims, Abstract, Title or description) are frequently used for text mining?
RQ4. In which subject areas is text mining used more in patents?
RQ5. What is the most preferred sample size selected by text mining researchers when applying text mining techniques to patents?

6 **search keywords/terms:**

(text* AND (analy* OR mining OR categorization OR classification OR cluster* OR extract* OR preprocessing OR processing OR transformation)), (data AND mining), "document classification", "document cluster*", "document summarization", "machine learning", "keyword extraction", "keyword discovery", "keyword retrieval", (information OR knowledge) AND (extract* OR discovery OR retrieval), "Latent Dirichlet Allocation", LDA, "Latent Semantic Analysis", LSA, "Natural Language Processing", NLP, "content analysis", "topic extraction", "topic model*", "unstructured text", "unsupervised learning", "Vector Space Model", "VSM", "support vector machines", "naive bayes classifier", "association rules", "k-nearest neighbor", "neural networks" OR "decision trees" AND (patent OR patents)
**"patent analy*", "patent mining", "patent cluster*", "patent map*", "patent roadmap*", "patent network", "patent visualization", "patent** visualisation**", patentometric*, "patent classification*", "patent retrieval"**

The literature search will not be limited by year of publication or geographic area and the language will be limited to English.



**Citation:** Homa Arshadi, Maryam Okhovati, Zohre Zahedi, Maryam Asharfi Text mining approaches applied to patents: A scoping review protocol https://dx.doi.org/10.17504/protocols.io.6qpvr43kzgmk/v1

7 **URL to search strategy**

We will upload this protocol on Open Sciences Framework (OSF)
We give permission for this file to be made publicly available:
Yes

8 **Condition or domain being studied**

Studies that include text mining on unstructured data on patent (such as title, abstract, description, claim).

9 **Participants/ Population**

The English Original/Conference papers published that meet the eligibility criteria of this review will be included.

10 **Intervention (s), Exposure(s)**

This scoping review does not have Intervention (s), Exposure(s) group.

11 **Comparator (s)/ control**

This scoping review does not have Comparator (s)/ control group.

12 **Types of study to be included**

The text mining techniques can be utilized to extract the information from structured or unstructured data. In this study our focus will be on unstructured text of patent. The following inclusion and exclusion criteria will apply:

**Inclusion criteria:**

-Studies that include text mining on unstructured data on patent (such as title, abstract, description, claim);
-peer-reviewed papers published in selected databases and in English;
-publication outlets (Original articles and conference proceedings);
-Full access to the document.

**exclusion criteria:**

-all publications that do not meet the inclusion criteria;
-Papers which did mining on structured data on patent;
-non-English results will be removed during the review process;
-Articles that focus on just citation analysis of patents;
-Secondary and tertiary studies, such as reviews, meta-analyses and surveys will be drawn;
-Editorial, meeting abstract, reviews, book reviews, books, book chapters and cover letters, and commentaries;
-duplicate publications and retracted publications.

13 **Context**

Identification of studies that uses text mining methods on unstructured data on patent (such as title, abstract, description, claim (



**Citation:** Homa Arshadi, Maryam Okhovati, Zohre Zahedi, Maryam Asharfi Text mining approaches applied to patents: A scoping review protocol https://dx.doi.org/10.17504/protocols.io.6qpvr43kzgmk/v1

14 **Primary outcome(s)**

This study will map the current state of different text mining approaches in patents.This study will be of value to policy makers and researchers by allowing them to find the latest efforts to patent text mining

15 **Secondary outcomes**

The secondary outcomes will be identifying the different text mining approaches, the most used data sources, metadata and subject areas in different application areas of the patents.

16 **Data extraction (selection and coding)**

After running search syntax in each database, the results of all the search will be exported to EndNote 20. Duplicate papers will be removed. The remaining papers will be imported to the Rayyan, for inclusion in review. Two independent research members will screen titles and abstracts of all papers against the inclusion and exclusion criteria. Subsequently the full-text of all potentially relevant papers will be assessed independently by two reviewers. In the case of conflicts, they discussed and then consulted the third author to reach a consensus.

Key findings relevant to the review will be charted from the included studies using a data extraction tool developed in Excel software by the members of the review team. The following data will be extracted: title, author, publication year, Journal, techniques, tools or applications of text mining, data sources, patent metadata (Claims, Abstract, Title or description), subject area, sample size and the purpose of the study. Two review team members will extract data independently and discrepancies will be solved by consulting with a third expert.

The results of the search and the study inclusion process will be reported following the principles of the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analysis extension for Scoping Reviews)

17 **Strategy for data synthesis**

The extracted data will be categorized based on types of different techniques of patent text mining, patent data sources, and key features like comparison of the frequency of use of different patent metadata, identifying the most used subject areas in this topicand sample size.

18 **Analysis of subgroups or subsets**

There will be no analysis of subgroups or subsets.

19 **type and method of review**

Scoping review

20 **Country**

Iran

21 **Current review status**



**Citation:** Homa Arshadi, Maryam Okhovati, Zohre Zahedi, Maryam Asharfi Text mining approaches applied to patents: A scoping review protocol https://dx.doi.org/10.17504/protocols.io.6qpvr43kzgmk/v1

On-going

22 **any additional information**

Appendix 1: Search syntax for this Scoping review in WoS:

| A | B |
|---|---|
| | **Search syntax** |
| #1 | (TS=(( ( text* AND ( analy* OR mining OR categorization OR classification OR c |
| #2 | **TS=(( "patent analy*" OR "patent mining" OR "patent cluster*" OR "pa** |
| #1 OR #2 | Limited to: Article, Conference paper |

23 **Details of final report/publication(s).**

It will be published in a peer-reviewed journal.



**Citation:** Homa Arshadi, Maryam Okhovati, Zohre Zahedi, Maryam Asharfi Text mining approaches applied to patents: A scoping review protocol https://dx.doi.org/10.17504/protocols.io.6qpvr43kzgmk/v1