Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Technology of Supporting Medical Decision-Making Using Evidence-Based Medicine and Artificial Intelligence

Georgy Lebedev\textsuperscript{a,b,c,*}, Eduard Fartushnyi\textsuperscript{a}, Igor Fartushnyi\textsuperscript{a}, Igor Shaderkin\textsuperscript{a,b}, Herman Klimenko\textsuperscript{a}, Pavel Kozhin\textsuperscript{a}, Konstantin Koshechkin\textsuperscript{a}, Ilya Ryabkov\textsuperscript{a,b}, Vadim Tarasov\textsuperscript{a}, Evgeniy Morozov\textsuperscript{a}, Irina Fomina\textsuperscript{b}

\textsuperscript{a}I.M. Sechenov First Moscow State Medical University, 2-4 Bolshaya Pirogovskaya st., Moscow, 119991, Russia
\textsuperscript{b}Federal Research Institute for Health Organization and Informatics, 11, Dobrolubova street, Moscow, 127254, Russia
\textsuperscript{c}V.I.Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. 4 Oparin st., 117997 Moscow, Russia

Abstract

Currently, Medical errors are a serious problem when examining patients. Creating information systems that use the capabilities of evidence-based medicine and artificial intelligence methods will allow the doctor to make an informed and proven decision. In this article, the authors offer a description of an information system that solves the problem of supporting medical decision making based on evidence-based medicine. This is achieved by using artificial intelligence methods. This work was supported by a grant from the Ministry of Education and Science of the Russian Federation, a unique project identifier RFMEFI60819X0278.

© 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the KES International.

Keywords: medical decision support systems, evidence-based medicine, intellectual systems, text mining

1. Introduction

According to experts from the world health organization, more than 100,000 people die each year from medical errors. This is more than the number of deaths from road accidents, breast cancer or AIDS - three causes that attract much more public attention. With the increased focus on preventing medical errors that occurred after the publication of the Institute of medicine’s landmark report in 1999, computer-based physician order entry systems

* Corresponding author. E-mail address: geramail@rambler.ru
(CPOE) in combination with CDSS were suggested as key. Element of system approaches to improving patient safety and quality of medical care [1-4].

It was in 2000 that new approaches to improving the quality of medical care were proposed as key conditions and prerequisites for reducing medical errors - the use of computer systems for entering doctor's orders (CPOE) in combination with CDSS.

Types of medical errors at different stages:

- Diagnostics (error or delay in diagnostics; rejection of using within named tests; using of outdated tests or therapies; failure to perform actions based on monitoring or testing results)
- Treatment (error in performing an operation, procedure, or test; error in prescribing treatment; error in the dose or method of use of the drug; unavoidable delay in treatment or in response to an abnormal test; inappropriate care)
- Preventative error (failure to provide preventative treatment; imperfect monitoring or follow-up treatment).

To solve the problem of reducing medical errors, we create an information system "SechenovDatamed", designed to support medical decision-making based on providing prompt access to complete and reliable information about the patient's health, implementing automated procedures for checking the compliance of the selected treatment with the standards of medical care, checking the compliance of prescribed medicines with existing contraindications and for possible drug interactions.

In the area of direct medical care, the most significant problems are problems of information support, namely:

- Prevention and early diagnosis of diseases, timely provision of medical care to patients, including those whose treatment is organized using hospital-substituting technologies
- Maximum effective using of available, including high-tech medical equipment, expensive medicines, donor materials and drugs based on them
- Support of medical decision-making, including by providing prompt access to complete and reliable information about the patient's health, implementing automated procedures for checking the compliance of the selected treatment with the standards of medical care, checking the compliance of prescribed medicines with existing contraindications and for possible drug interactions, as well as on the basis of expert systems and available case data.

Evidence-based medicine (EBM) is based on the rule that qualitative data presented in systematic reviews (SR), meta-analysis (MA), randomized controlled trials (RCT) must be used to find reliable evidence. At the same time, the laboriousness of conducting a typical systematic review [8] is estimated at least 1000 hours of highly skilled labor of the relevant specialists. Modern estimates [9] show that the time from the publication of information to the release of a systematic review is on average 450 days.

The practical implementation of evidence-based medicine methods also has a number of objective difficulties. Firstly, the processed data is presented for distribution in the form of unstructured texts in a natural language. Secondly, collections of biomedical documents are multilingual. Thirdly, not every publication contains the results of clinical trials and can be accepted for processing. Fourth, information resources containing such documents comprise tens of millions of peer-reviewed publications and their volume is growing exponentially.

Obviously, the search and extraction of evidence for the formation of sound clinical recommendations on existing volumes of biomedical data cannot be effective using traditional methods. And the reviews themselves may become obsolete and lose their relevance.

At the same time, the rapid development of artificial intelligence technologies and methods of intellectual word processing create the prerequisites for the successful solution of such problems in a mode close to on-line.

The accumulated volumes of biomedical information and the growing number of publications exponentially require the development of effective and high-quality methods for thematic categorization of documents, their categorization, extraction of facts and knowledge.

The work [10] describes the fundamental approaches that are used for various problems of medical content analysis. The practical implementation of the methods was proposed in [11]. A thorough examination of automated
methods applied to biomedical literature and their contribution to innovative biomedical research can be found in [12].

The main sources of knowledge have open access. There are over 30 million links indexed through PubMed, the largest database of biomedical literature developed and maintained by the National Center for Biotechnological Information (NCBI). Integrated with PubMed, the Entrez NCBI search engine provides access to a diverse set of 38 databases. PubMed is currently indexing publications from 5,254 journals in biology and medicine until 1948. PubMed is now the primary tool for finding biomedical literature. Every day, the system processes several million queries generated by users to keep abreast of the latest achievements and highlight priority research in their fields.

Despite PubMed providing effective search interfaces, it’s becoming increasingly difficult for users to find information that matches their individual needs. Detailed user queries generate search results containing thousands of relevant documents.

According to Clarivate Analytics, another citation database of Web of Science Core Collections contains over 1.4 billion links from over 20 thousand publication sources (https://clarivate.com/products/web-of-science/web-science-form / web-science-core-collection /).

The Microsoft Academic Search database (https://academic.microsoft.com;) in the Medicine section is more than 26 million publications and more than 17 million in the Biology section. At the same time, the most rapid growth in volumes is observed in the Biochemistry section - more than 5 million publications, of which more than a million are genetics.

Google Scholar does not report the amount of links that can be identified using their search engine, but for the query “personalized medicine” (https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=personalized+medicine&btnG =). The volume of search results amounted to 1 million 210 thousand documents.

The approximate number of daily publications only in indexed databases on medical topics today has exceeded 15 thousand publications / day. Obviously, the processing of such volumes without the use of special methods for analyzing biomedical content, searching and extracting knowledge is impossible.

2. Methods

The object of automation is the processes carried out by doctors-specialists of clinical centers of Sechenov University in the framework of their practical clinical activities, namely:

- Defining the purpose and problem of treatment
- Choosing the best Medicine
- Determination of the dose and method of injection of drugs
- Selection of methods for monitoring the effectiveness of treatment
- Getting information about the effects of drugs, including adverse ones
- Determining the "benefit-risk" ratio (treatment safety)
- Determination of criteria for stopping the using drugs.

The "SechenovDatamed" system implements the functionality of supporting medical decision - making by providing prompt access to complete and reliable information about the patient's health, implementing automated procedures for checking the compliance of the selected treatment with the standards of medical care, checking the compliance of prescribed medicines with existing contraindications and for possible drug interactions.

As part of the project, the following main functions of "SechenovDatamed" were implemented:

- Accumulation, storage and updating of data on various areas of activity in the field of medicine, obtained both from functional units (clinics) and from other sources through automatic electronic services. This functionality is based on a single corporate data warehouse, a single methodology for generating information and unifying business processes
- Providing specialists (users) diverse thematic reference and methodological information on the data contained in the repository for decision-making
• Providing multi-criteria data search in the mode of individual user queries, as well as standardizing and reducing the complexity of the process of searching and obtaining thematic information, increasing its reliability and compliance with the original search queries

• Creating a personal workspace of the user with the ability to flexibly customize the display of information. Providing easy-to-use analytical tools that allow you to implement significant responsiveness in obtaining analytical data, as well as generate a high degree of trust for end users

The following intelligent data processing technologies were implemented in the created System:

• Intelligent search tools that implement semantic search mechanisms
• Natural language processing
• Interfaces for operating the system from mobile devices and wearable gadgets
• Cross-language support and automatic translation to the main languages
• Technologies for extracting knowledge from unstructured text content (text mining).

Modern text classification and data extraction methods use machine learning (ML) methods rather than rule based methods. The methods used are divided into two groups:

- training with a teacher, which requires the marking of training data sets;
- training without a teacher and, therefore, without pre-processed training arrays.

The first group uses neural network modeling methods, and the second is actually based on probability theory and numerical optimization methods. The methods we have developed are based on thematic modeling of collections of biomedical documents.

The probabilistic thematic models [2,4,5,6] developed by us are based on the “bag of words” [7] hypothesis, when the subject and semantic content of a document is determined by the set of words included in the document and does not depend on the order of their sequence or location in the document. Thus, each topic is described by a probabilistic distribution of many terms, and each document is described by a probabilistic distribution on many topics. The hypothesis ‘bag of words’ suggests that a collection of documents is a sequence of random words and is independent of a mixture of such distributions. The solution to the problem of restoring the components of the mixture from the sample allows you to determine the subject of documents. Since a document or term can relate to many topics at the same time with different probabilities, this method allows for “soft” clustering of documents and terms by topic cluster. In medical practice, this is extremely important since one document can relate to different thematic areas - cardiology & anesthesiology, pulmonology & anesthesiology, ... and so on.

The Institute of Digital Medicine, together with the Center for Evidence-Based Medicine and Clinical Centers of the University, is working to create an automated system for the formation of clinical recommendations based on evidence-based medicine and existing international practices.

The system under development belongs to the class of databases of clinical guidelines for use in places of care:

- clinical guidelines for treatment;
- clinical diagnostic guidelines.

In addition to creating such a database of clinical guidelines, it is necessary to create specialized software for the formation of systematic reviews in arbitrary medical areas and problems. This problem is solved by creating a multilingual probabilistic thematic model for the collection of studied documents, while taking into account the n-language dictionary (by the number of types of languages represented in the selected collection of documents) and ontological relationships between documents of a comparable collection. The normalization of the terms in the collection of multilingual documents is supposed to be carried out using the thesauri of MeSH medical terms for each language of presentation of documents. Due to the fact that the current version of MeSH provides synchronization of the used terms in more than 48 languages, to cover the entire multilingual collection of documents, this choice can be considered quite reasonable.

To combine different types of classification features, additive regularization of the thematic model is used. Studies are carried out on the significance of various regularization features on the quality of the model (author, place of study, company, information source, key terms ... and so on).
Pre-processing of documents in collections is one of the key components of the developed method and involves the consistent solution of the tasks of content tokenization, filtering and cleaning of stop words, lematization of cleared samples and stemming.

Most of the documents examined are presented in English. Tokenization of English-language biomedical texts is especially difficult due to the fact that traditional English differs from biomedical texts in both its syntactic content and grammar. For effective tokenization of documents, dictionaries of clinical terms from a systematic medical nomenclature, gene symbols, names of proteins and drugs are used. Persistent combinations of terms are replaced with individual tokens.

Filtering is used to remove words that do not carry semantic content from documents. In addition to the traditional removal of stop words (prepositions, conjunctions, etc.). documents are cleared of terms not related to evidence-based medicine. To do this, on the basis of the corpus of documents available at the Sechenov University Center for Evidence-Based Medicine, a frequency-statistical analysis of the words used was performed with the use for subsequent filtering. So, as the words found in the texts quite often contain little information to separate documents, and rarely found words are not significant, both of them are deleted from the documents.

At the next stage, lemmatization of documents is carried out, namely, a morphological analysis of words is carried out, their various forms are grouped into one. Lemmatization reduces various forms of words to basic forms (verb - infinitive, nouns - singular, nominative, etc.).

Thus, the following tasks are successively solved:
- the formation of a multilingual registry of information sources for evidence-based medicine;
- parsing of information sources and the formation of datasets of the analyzed documents;
- pre-processing collections of documents, data normalization using biomedical thesauruses and ontologies, taking into account the specifics of personalized medicine;
- machine learning of the developed multilingual thematic model on a collection of documents and quality assessment according to well-known metrics
- a visual presentation of the results on the selected issues.

In addition to implementing these functions, the System provides:
- Processing of text data in Russian
- Generating xml-documents based on input documents that contain mapped out attributes
- High accuracy of detecting metainformation, as well as entities and units of measurement

The "SechenovDatamed" system is organized on the principle of a three–tier client-server architecture, namely:

- The database level is implemented by the database management system:
  - Database management system
  - Database information objects
  - program elements of the Database level
- The business logic level is implemented through the following components:
  - Application server
  - General-purpose services
  - Business logic components for application systems
- The client level includes:
  - Web server
  - Software for implementing the user interface

The created System makes it possible to deploy remote data centers on technical platforms. The "SechenovDatamed" system has a centralized database with secure access for users. A centralized database provides:

- Automatic data export and import (working with a single database)
• Possibility of permanent (online) access to the system and all current information (without linking to the user's stationary workplace)
• Keep unified reference lists, classifiers, and other formalized documents, and provide centralized control over the system's document content

The "SechenovDatamed" system provides the ability to work in a thin client mode (the user works through a web browser), operating in various operating environments – Microsoft Windows, Unix (Linux), as well as the web browsers Microsoft Internet Explorer, Mozilla Firefox, Google Chrome, Apple Safari.

3. Results

The "SechenovDatamed" System consists of the following subsystems:

• database and knowledge storage subsystem
• subsystem of processing search queries and issuing clinical recommendations for the rational use of pharmacotherapeutic drugs
• machine learning subsystem for creating evidence-based medicine knowledge bases
• subsystem for providing access from mobile devices

In addition to these subsystems, the "SechenovDatamed" System includes a set of software and hardware tools.

3.1. Subsystem for storing databases and knowledge

The subsystem for storing databases and knowledge is designed to accumulate, store and update data on various areas of activity in the field of medicine, from various sources through automatic electronic services. This functionality is based on a single corporate data warehouse, a single methodology for generating information and unifying business processes.

3.2. Subsystem for processing search queries and issuing clinical recommendations for the rational use of pharmacotherapeutic drugs

The subsystem for processing search queries and issuing clinical recommendations for the rational use of pharmacotherapeutic drugs is designed to implement the ability of a specialist (doctor) when prescribing any pharmacotherapy to get acquainted not only with the standard characteristics of the drug obtained from the classifiers, the Unified State Register of Medicines, standards of medical care, information about drug interactions in the human body, drug interactions outside the human body, drug interactions and anthropometric parameters of a person, etc., but also with the expert opinions of other specialized specialists.

3.3. Machine learning subsystem for creating evidence-based medicine knowledge bases

The machine learning subsystem for the formation of evidence-based medicine knowledge bases is designed to provide specialists (users) with diverse thematic reference and methodological information on the data contained in the storage for decision-making. As well as to ensure multi-criteria data search in the mode of individual user requests, standardization and reducing the complexity of the search process and obtaining thematic information, increasing its reliability and compliance with the original search queries.

3.4. Subsystem for providing access from mobile devices

The subsystem for providing access from mobile devices ensures optimal display of the listed sub-items of the System functions on mobile devices (phones and tablets) running Android and iOS operating systems.
3.5. Implementation of requirements for methods and means of communication for information exchange between components of subsystems

Information exchange between the server components of the "SechenovDatamed" System is carried out over the local computer network of the software and technical complex on which they are located – software and technical complex of Sechenov University.

Data transmission between all components of the “SechenovDatamed” System in the normal mode of operation is carried out on-line.

4. Discussion

The subsystem for storing databases and knowledge is designed to accumulate, store and update data on various areas of activity in the field of medicine, from various sources through automatic electronic services. This functionality is based on a single corporate data warehouse, a single methodology for generating information and unifying business processes.

The database and knowledge Storage subsystem provides solutions to the following tasks:

- Preparation of materials for placement in storage
- Storage of source documents
- Extracting and storing various types of entities from document texts
- Creating a search index

To solve these problems, the subsystem implements the following functions:

- Transferring data in the files from various open sources
- Uploading information to the storage
- Storing the accumulated data
- Extracting metainformation from a document and performing automatic identification of entities and units of measurement in the text
- Semantic analysis of data based on semantic and linguistic components of information extraction;
- Decoding medical abbreviations and abbreviations
- Identification of concepts for diseases, pathologies, interventions, medications, and examinations
- Identifying semantic relationships between these concepts, as well as defining the attributes of these concepts
- Extracting document attributes
- Extracting typed entities from texts (specific types of entities are provided by the Customer during system maintenance)
- Extract all units of measurement from the text (based on the all-Russian classifier of units of measurement)
- Normalization of the selected entity
- Automatic classification of documents by category
- Information search
- Search for similar data extracting and storing various types of entities from document texts

In addition to these functions, the System generates xml documents based on input documents that contain marked attributes, and also provides high accuracy in detecting metainformation, as well as entities and units of measurement.

The developed and applied methods and algorithms for intellectual processing of biomedical texts meet the requirements of completeness, adequacy, and provide the possibility of using both domestic and foreign databases and knowledge:

- Evidence-based medicine clinical recomendations of Sechenov University
- Digital archive of biomedical literature PubMed (process at least > 30 million documents)
• Digital archive of full-text biomedical and journal literature, including PMC clinical medicine (process at least >7 million documents)
• Register of medicines and drugs registered on the territory of the Russian Federation (State register of medicines)
• Pharmacological, peer-reviewed knowledge base DrugBank.ca
• Pharmacogenomic knowledge base on the effect of genetic variations on drug reactions PharmGKB
• Database of standardized clinical drug names RxNORM

The subsystem is developed using open source software.

The basic mechanism for the System to work with incoming thematic content is clusterization of descriptions.

4.1. Designing

As part of the project, the functionality of clustering bibliographic descriptions of the PubMed library was implemented using bigartm thematic modeling methods, namely:

• The technological process has been developed (a technological process that operates in automatic mode, without the participation of an operator, 24 hours a day, 7 days a week), which includes
  • Collecting information from the PubMed library website
  • The control and elimination of re-downloading data
  • Primary processing of the received information
  • Filtering the input data stream for each of the specified subject areas
  • Building and updating a clustering model for each thematic area based on BIGARTM
  • Processing clustering results and automatically configuring Web services for user access
  • Activation of the process according to the specified schedule

4.2. Realisation

The following applications and scripts were developed while working on the project.

Application t17_java_ftp_http—provides collection of new arrivals of bibliographic descriptions from the FTP server of the PubMed library. t17_java_ftp_http downloads only previously unprocessed archives, it is implemented in Java. The application is called using the command line with two parameters: the first is the file describing access to the FTP server, and the second is the path to the archive location.

ICM-02 application—provides processing of XML file of bibliographic descriptions. Processing includes selecting the data fields required for clusterization (“article titles”, "unique article ID-link to the original bibliographic description on the PubMed server", "abstract text"), filtering the message flow in accordance with the thematic filter (currently the filter is configured for two thematic areas: erectile dysfunction "and" urolithiasis”), clearing the text of punctuation marks and words included in the"stop words" dictionary. As a result of the application, files are generated for top-up in thematic areas in the format "word bag" ("vowpal_wabbit") and a file corresponding to the serial number of the article to its ID and name (used for the Web application user interface). The app is developed in Python. The application is called using the command line with two parameters: the first is the path to the folder with XML files with new bibliographic descriptions, and the second is the file with descriptions of filters for thematic areas.

App icm_arm provides the training of primary data (batch file) to generate model topic modeling BIGARTM, forming and updating the model, the implementation of a clustering model data preparation and preprocessing of the data Phi matrices — the distribution of words by topics and Theta the distribution of articles by topics, as well as data preparation for Web application user interface. The app is developed in Python.

Application index.php -presenting clustering results to users via the Web interface. Application index.php it is implemented in PHP and operates under the Apache Web server. index.php the source data for a topic area (for example, "erectile dysfunction") uses the file distribution of articles by topic and the file matching the article number to its title and ID.
4.3. Evaluating the quality

Perception was used to assess the quality of building thematic models - a measure of how well the model predicts the details of the test collection (the lower the perplexity, the better the model).

Sparsity of the Phi matrix: 0.987408578396. Sparsity of the theta matrix: 0.678321182728. List of the most significant words for each topic.

The compliance of the content of the documents found to the search query was evaluated by experts. For example, the model was tested to find evidence on the treatment of newborns with SARS CoV-2 infections and COVID-19 diseases. As a result of the simulation, a set of documents related to the treatment of newborns was discovered - 394 documents.

The clinician is interested in evidence on the reduction (avoidance) of mortality in newborns or women in childbirth. Using the constructed thematic model, publications related to the results of a meta-analysis, systematic reviews, or randomized clinical trials were selected. Using additional MeSH filters (publication type) and a mortality request, the collection of selected documents from the initial set of 394 documents was reduced to 5 documents Table 1.

| Publication | Source |
|-------------|--------|
| Outcome of Coronavirus spectrum infections (SARS, MERS, COVID 1-19) during pregnancy: a systematic review and meta-analysis. Di Mascio, D; Khalil, A; Saccone, G; Rizzo, G; Buca, D; Liberati, M; Vecchiet, J; Nappi, L; Scambia, G; Berghella, V; D'Antonio, F | American journal of obstetrics & gynecology MFM |
| Maternal and perinatal outcomes with COVID-19: A systematic review of 108 pregnancies. Zaigham, M; Andersson, O | Acta obstetricia et gynecologica Scandinavica |
| Coronavirus disease 2019 (COVID-19) and pregnancy: a systematic review. Yang, Z; Wang, M; Zhu, Z; Liu, Y | The journal of maternal-fetal & neonatal medicine |
| A systematic scoping review of COVID-19 during pregnancy and childbirth. Elshafeey, F; Magdi, R; Hindi, N; Elshebiny, M; Farrag, N; Mahdy, S; Sabbour, M; Gebril, S; Nassar, M; Kamel, M; Amir, A; Emara, MM; Nabhan, A | International journal of gynaecology |
| Severe maternal morbidity and mortality associated with COVID-19: The risk should not be down-played. Westgren, M; Pettersson, K; Hagberg, H; Acharya, G | Acta obstetricia et gynecologica Scandinavica |

After processing them, the following facts were obtained:

- pneumonia was diagnosed in 91.8% of cases, and the most common symptoms were fever (82.6%), cough (57.1%), shortness of breath (27.0%), then general fatigue (22.5%) diarrhea (8.8%), shortness of breath (11.3%), sore throat (7.5%) and myalgia (16.3%);

- preterm birth rate <37 weeks - 24.3% ... 41.1% depending on the review from 84% to 91% of births are allowed by caesarean section;

- 57.2% of newborns were admitted to the neonatal intensive care unit;

- premature birth is the most common adverse pregnancy outcome;

- miscarriages, cesarean section and perinatal death (7-11%) were also more common;

- no clinical evidence of vertical transmission of infection has been published;

- stillbirths (1.2%), neonatal death (1.2%), low birth weight (<2500 g, 5.3%), fetal distress (10.7%) and newborn asphyxia (1.2%) were recorded);

- severity distribution of women in labor with Covid-19 (95.6%) lungs, (3.6%) severe and (0.8%) critical.
4. Conclusions

The methodology proposed by the authors in this paper for creating a system for supporting medical decision-making can certainly be useful for practitioners throughout the Russian Federation. The implementation of the access to knowledge bases through mobile app will make the process more efficient.

The developed methodology for searching and extracting knowledge can be used by specialist doctors to automatically monitor highly specialized sections of interest to them and formulate clinical recommendations and guidelines on the principles of evidence-based medicine. Automated intelligent search for documents similar in subject matter to the query of interest allows doctors not to waste time monitoring new publications on topics of interest, but to obtain semantically close documents that can be selected by an intelligent robot that continuously monitors the flow of documents published on PubMed, WebOfScience, Scopus, Google Scholar, et al.

References

[1] Kohn LT, Corrigan JM, Donaldson MS, editors: To Err is Human: Building a Safer Health System. Institute of Medicine (US) Committee on Quality of Health Care in America. Washington (DC): National Academies Press (US); 2000. DOI: 10.17226/9728
[2] Blei D. M., Ng A. Y., Jordan M. I. Latent Dirichlet allocation // Journal of Machine Learning Research. — 2003. — Vol. 3. — Pp. 993–1022.
[3] Distributed Representations of Words and Phrases and their Compositionality. / Tomas Mikolov, Ilya Sutskever, Kai Chen et al. // NIPS / Ed. by Christopher J. C. Burges, L’eon Bottou, Zoubin Ghahramani, Kilian Q. Weinberger. — 2013. — Pp. 3111–3119.
[4] Knowledge discovery through directed probabilistic topic models: a survey / Ali Daud, Juanzi Li, Lizhu Zhou, Faqir Muhammad // Frontiers of Computer Science in China. — 2010. — Vol. 4, no. 2. — Pp. 280–301.
[5] Vorontsov K. V., Potapenko A. A. Additive Regularization of Topic Models // Machine Learning Journal. — 2015. — Vol. 101. — Pp. 303–323.
[6] Statistical topic models for multi-label document classification / Timothy N. Rubin, America Chambers, Padhraic Smyth, Mark Steyvers // Machine Learning. — 2012. — Vol. 88, no. 1-2. — Pp. 157–208.
[7] Vulić Ivan, Smet Wim, Moens Marie-Francine. Cross-language information retrieval models based on latent topic models trained with document-aligned 140 comparable corpora // Information Retrieval. — 2012. — Pp. 1–38.
[8] Allen IE, Olkin I. Estimating time to conduct a meta-analysis from number of citations retrieved. JAMA. 1999;282:634–5
[9] Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open. 2017;7:e012545.
[10] Tsafnat G, Dunn A, Glasziou P, Coiera E. The automation of systematic reviews. BMJ. 2013;346:f139.
[11] O’Connor AM, Tsafnat G, Gilbert SB, Thayer KA, Wolfe MS. Moving toward the automation of the systematic review process: a summary of discussions at the second meeting of International Collaboration for the Automation of Systematic Reviews (ICASR). Syst Rev. 2018;7:3.
[12] Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic reviews: 2. Combining human and machine effort. J Clin Epidemiol. 2017;91:31–7.