

О.В. Золотарев, А.Х. Хакимова, И.М. Зацман, Донахью Гу

ПРОЕКТ ПОСТРОЕНИЯ МУЛЬТИМОДАЛЬНОЙ МЕДИЦИНСКОЙ
ИНФОРМАЦИОННОЙ СИСТЕМЫ

Рассматривается подход к построению мультимодальной медицинской информационной системы, которая предполагает построение трехязычного словаря медицинских терминов, включая английский, китайский и русский языки. Идея системы основывается на анализе эволюции терминологии в медицине. Данные системы выгружаются автоматически из медицинской электронной библиотеки PubMed, формируется корпус текстов на основе поискового запроса, проводится анализ текстов документов, автоматическое выделение потенциально новых медицинских терминов, формирование словаря терминов мультимодальной медицинской информационной системы. Структура словаря включает сам термин, его дефиниции, категории, контексты со ссылками на оригиналы публикаций и др. Система будет размещаться на сервере с возможностью организации доступа к ней на основе полномочий пользователей. В функционал системы будет входить добавление новых терминов, коррекция дефиниций, контекстов, проведение статистических расчетов. Система будет полезна медицинским работникам для поиска информации, выделения тенденций развития заболеваний, постановки диагноза.

Ключевые слова: мультимодальная медицинская информационная система, медицинские термины, анализ текстов, формирование словарей терминов.

O.V. Zolotarev, A.Kh. Khakimova, I.M. Zatsman, Dongxiao Gu

PROJECT FOR BUILDING A MULTIMODAL MEDICAL
INFORMATION SYSTEM

An approach to the construction of a multimodal medical information system is considered, which involves the construction of a trilingual dictionary of medical terms, including English, Chinese and Russian. The idea of the system is based on the analysis of the evolution of terminology in medicine. Data are automatically uploaded from the medical electronic library PubMed, a text corpus is formed based on a search query, the text of documents is analyzed, potential new medical terms are automatically identified, and a dictionary of terms of a multimodal medical information system is formed. The structure of the dictionary includes the term itself, its definitions, categories, contexts with links to the originals of publications, etc. The system will be hosted on a server, with the ability to organize access to it based on user rights. The functionality of the system will include the addition of new terms, correction of definitions, contexts, and statistical calculations. The system will be useful for medical professionals to search for information, highlight trends in the development of diseases, and make a diagnosis.

Keywords: multimodal medical information system, medical terms, text analysis, formation of dictionaries

Введение

В данной работе описывается подход к построению автоматизированной мультимодальной медицинской информационной системы (ММИС) на основе информации из англоязычной базы медицинских электронных публикаций PubMed [12], созданной Национальным центром биотехнологической информации (NCBI, США). Система предназначена для поддержки медицинских словарей, поиска новых медицинских терминов, определения тенденций динамики популярности терминов, проведения статистических расчетов.

Золотарев Олег Васильевич

кандидат технических наук, доцент, заведующий кафедрой информационных систем в экономике и управлении Института информационных систем и инженерно-компьютерных технологий Российского нового университета, Москва. Сфера научных интересов: компьютерная лингвистика. Автор 79 опубликованных научных работ. Московский инженерно-физический институт.

Email: ol-zolot@yandex.ru

Хакимова Аида Хатифовна

кандидат биологических наук, ведущий научный сотрудник Института информационных систем и инженерно-компьютерных технологий Российского нового университета, Москва. Сфера научных интересов: компьютерная лингвистика. Автор 40 опубликованных научных работ. Московский инженерно-физический институт.

Email: aida_khatif@mail.ru

Зацман Игорь Моисеевич

доктор технических наук, заведующий отделом «Представление знаний в информационных и управляющих системах» федерального исследовательского центра «Информатика и управление» РАН, Москва. Сфера научных интересов: информатика, семиотика. Автор 151 опубликованной научной работы. Московский авиационный институт.

Email: izatsman@yandex.ru

Донахью Гу

Ph.D., профессор, профессор Школы менеджмента Технологического университета Хэфэй, Китай. Сфера научных интересов: медицинские информационные системы. Автор 74 опубликованных научных работ.

Email: dongxiaogu@yeah.net

Структура системы включает три словаря: а) исходный словарь терминов (ИСТ), который составляется вручную экспертом; б) словарь потенциально новых терминов (СПТН), использующийся для выделений новых терминов; в) словарь терминов (СТ), представляющий собой пополняемый словарь медицинских терминов ММИС, который является ядром системы.

Функционал ММИС состоит из нижеперечисленных подсистем.

1. Подсистема формирования корпуса текстов (ПФК), строится автоматически путем загрузки документов из электронной библиотеки PubMed на основе поискового запроса [15]. В проекте используется free full-text archive of biomedical and life sciences journal literature PMC (PubMed Central).

2. Подсистема формирования словаря потенциально новых терминов (СПНТ). В процессе обработки научных документов выделяются (извлекаются из неструктурированных текстов) медицинские термины, включающие ключевые слова и словосочетания из поискового запроса (не обязательно совпадающие полностью и терминами, представленными в запросе).

3. Рабочее место эксперта (РМЭ) для исследования и коррекции представленной в словаре терминов информации, внесения изменений в словарные статьи.

Проект построения мультимодальной медицинской информационной системы

4. Рабочее место оператора (РМО) для ввода информации в систему.
5. Рабочее место пользователя (РМП) для работы с системой.
6. Подсистема статистики (ПС).
7. Подсистема внешних словарей (ВС).

Система ММИС является многоязычной. Основным языком системы – английский (в силу того, что наибольшее количество статей в международных электронных библиотеках представлены на английском языке). Планируется сделать словарь терминов (СТ) трехязычным, с включением английского, русского и китайского языков. Также планируется построить трехязычный интерфейс. Для этого будут построены таблицы межъязыкового соответствия. Перевод статей с английского на другие языки не предусматривается.

Методы

Выгрузка данных из библиотеки PMC ведется с использованием Entrez утилит (E-utilities). Это интерфейс, позволяющий осуществлять доступ к 38 базам биомедицинских данных. Запрос к NCBI направляется в виде URL-адреса из программы (единственное условие – программная система должна поддерживать отправку URL-адреса). В результате на сервере формируется корпус текстов, соответствующий поисковому запросу (работает подсистема формирования корпуса текстов).

Далее запускается вторая подсистема – обработка научных текстов. Цель подсистемы – сформировать словарь потенциально новых терминов. На основе поискового запроса и с учетом шаблонов для выделения терминов из текстов выполняется анализ текстовых документов. В результате обработки каждого документа корпуса выделяются потенциально новые термины в виде *n*-грамм, выполняется лемматизация (приведение слов текста к нормальной форме), сохраняются адреса вхождения терминов, оригинальная и нормальная формы терминов, устанавливается связь между термином в нормальной и оригинальной формах для каждого документа, сохраняются контексты вхождения терминов [9; 16].

Структура ММИС

В системе предполагается ввести следующие роли для доступа:

- *администратор* (полный доступ к функциям системы, включая изменение прав доступа);
- *эксперт* (полный доступ к функциям системы кроме изменения прав доступа, включая:
 - доступ к просмотру и редактированию словарей СТ и СПНТ, поиск терминов, их категоризация;
 - поиск документов, расширение или сужение их контекстов;
 - поиск, формирование и редактирование аннотаций терминов, дефиниций, контекстов, списков синонимов и их аббревиатур;
 - изменение статуса термина – утвержден, не утвержден;
- *пользователь* (просмотр информации БД без ее редактирования).

Сначала группа экспертов формирует исходный словарь терминов (ИСТ), который в результате согласования утверждается всеми экспертами. Первоначально ИСТ заполняется в виде таблицы Excel, затем загружается в БД ММИС.

Далее запускается подсистема формирования корпуса текстов (ПФК) на основе поискового запроса, утвержденного группой экспертов. Запрос включает поисковую строку, период, дату создания корпуса и выполняется автоматически. В результате запроса в автоматическом режиме из библиотеки PMC выгружаются искомые документы, и создается текстовый корпус, которому соответствует массив публикаций, содержащий название

статьи, информацию об авторах, тип и вид документа, дату публикации, аннотацию, полные тексты документов, доступные в РМС.

После завершения формирования корпуса текстов включается в работу подсистема формирования словаря СПНТ. Выделяемый потенциально новый термин добавляется в список терминов СПНТ, сохраняется его уникальный адрес (выполняются морфологическая и метатекстовая разметки текста документов корпуса), включающий ссылку на статью и положение термина в статье. После этого производится нормализация термина, создается отдельный список нормализованных терминов корпуса, включающий адрес термина в оригинальной форме (список терминов таблицы СПНТ). В данной подсистеме организуется поиск по терминам и навигация по вхождениям термина в статью и в корпус. Далее осуществляется поиск термина в нормализованной форме в словаре терминов (СТ). В случае нахождения термина в словаре СТ в списке нормализованных терминов устанавливается статус (новый, не новый, обработан). Вся работа по формированию СПНТ и списка нормализованных терминов выполняется автоматически (эксперт имеет доступ только к изменению статуса термина).

Рабочее место эксперта используется для анализа информации из словаря СПНТ и коррекции словаря медицинских терминов (СТ), каждый новый элемент которого формируется экспертом и утверждается группой экспертов. Словарь СТ в базе данных представлен набором таблиц и включает следующие элементы: термин, его категория, группа, синонимы, сокращения, аннотация, дефиниции, контексты, поясняющие термин со ссылками на статьи из библиотеки РМС, из которых они были выделены. Эксперт имеет доступ к словарям СПНТ, СТ, публикациям корпуса. Он может корректировать только словарь СТ (исключение – изменение статуса термина в СПНТ).

Словарь терминов имеет иерархическую структуру (включая гипо-гиперонимические отношения). Верхние уровни иерархии служат для классификации. Лепестковый уровень иерархии содержит информацию о термине.

Выбранные экспертом термины в словаре СПНТ могут быть либо добавлены в основной словарь терминов (СТ), либо, если анализируемый термин входит в оба словаря (СПНТ и СТ), могут быть скорректированы свойства термина в словаре СТ (аннотация, дефиниции, контексты). Если эксперт добавляет новый контекст, то автоматически сохраняется ссылка на адрес контекста в документе (эксперт в статье выделяет нужный контекст и выполняет команду привязки контекста к термину). Если в контексте определятся новое значение термина, то добавляется новая дефиниция и новый контекст. Дефиниции формулируются экспертом, контекст берется из публикаций корпуса. При просмотре контекста эксперту выводится часть текста документа. Контекст может быть расширен или сужен за счет окружающего исходный контекст текста.

Предполагается в БД делить термины на несколько категорий с изменяемым числом категорий и их названий. Первоначально термины в БД делятся на следующие 9 категорий: *general scientific terms, general medical terms, cancer, breast cancer, molecular biology, genetics, bioorganic chemistry, drugs for treatment and diagnostics, methods and technologies of diagnostics and treatment.*

Структура аннотации термина включает 6 компонентов:

- Meanings & Definitions & Abbreviations;
- Synonyms;
- Relationships;

Проект построения мультимодальной медицинской информационной системы

- Mapping;
- Term & Synonyms contexts;
- Term associations with disease stages.

Подсистема статистики (ПСР) позволяет вычислять абсолютные и относительные частотности встречаемости термина в корпусе текстов и в каждом документе. Для определения тенденций развития терминологии рассчитываются абсолютные и относительные частотности встречаемости терминов за определенные периоды времени [1; 4].

Подсистема внешних словарей (ВС) позволяет подключать общедоступные словари медицинских терминов для выполнения анализа новизны термина, уточнения его значений в процессе работы эксперта.

Ниже приведен пример поиска терминов для категории «Рак груди». В поиск включены следующие паттерны поиска:

- триграммы с одним термином из списка состава: ТХХ, ХТХ, ХХТ;
- биграммы: ХТ, ТХ.

Здесь *T* означает поисковый термин, *X*... – возможные кандидаты на включение в многословный термин в результате анализа текста публикации.

Поиск квадграммов был сочтен нецелесообразным из-за небольшого числа четырехсловных медицинских англоязычных терминов. Приведем примеры квадграммов: triple negative breast cancer, triple-negative breast cancer, Node-Negative Breast Cancer. Трехсловные термины встречаются чаще: Breast Cancer Cells, breast cancer patients, Human Breast Cancer, Breast Cancer Survivors, Advanced Breast Cancer, Metastatic breast cancer и др. Наиболее распространены биграммы: Breast Cancer, triple-negative, Breast Carcinoma, Triple Negative, breast reconstruction, brain metastases, breast density и др.

Был осуществлен поиск статей за 2020 г. в РМС по поисковому запросу: early [All Fields] AND ("breast neoplasms" [MeSH Terms] OR ("breast"[All Fields] AND "neoplasms" [All Fields]) OR "breast neoplasms" [All Fields] OR ("breast"[All Fields] AND "cancer" [All Fields]) OR "breast cancer" [All Fields]). Из найденных 49675 статей было отобрано 10 наиболее релевантных запросу публикаций [2; 3; 5–8; 10; 11; 13; 14]. Для последующего экспертного анализа был создан корпус полных текстов этих статей.

Ниже приведены примеры *n*-грамм, выделенные из корпуса, содержащие термин Microcalcification из списка СТ.

Паттерн	Выдержки из текста	<i>n</i> -граммы после нормализации
TXX	microcalcifications are composed of either calcium Microcalcifications appearance at mammography microcalcifications in the breast microenvironment microcalcifications and the breast microenvironment stratified for the presence of microcalcifications microcalcifications formation processes microcalcification resorption process microcalcifications in breast cells microcalcifications development and carcinogenesis microcalcifications compared with tumor microcalcifications and reported that hydroxyapatite microcalcifications in the primary tumor	microcalcification compose calcium Microcalcification appearance mammography microcalcification breast microenvironment stratify presence microcalcification microcalcification formation process microcalcification resorption process microcalcification breast cell microcalcification development carcinogenesis microcalcification compare tumor microcalcification report hydroxyapatite

TXX	microcalcifications development and hyperplasia microcalcifications production and resorption microcalcifications should also be evaluated microcalcifications could help identifying microcalcifications preventive therapeutics	microcalcification primary tumor microcalcification development hyperplasia microcalcification production resorption microcalcification also evaluate microcalcification help identify microcalcification prevent therapeutics
XTX	Breast microcalcifications are composed presence of microcalcifications in the breast presence of microcalcifications compared associated with microcalcifications, breast existence of microcalcifications formation Breast microcalcifications formation, impact of microcalcifications in breast hydroxyapatite microcalcifications were actively breast microcalcifications development developing microcalcifications preventive	Breast microcalcification compose presence microcalcification breast presence microcalcification compare associate microcalcification breast exist microcalcification formation Breast microcalcification formation impact microcalcification breast hydroxyapatite microcalcification active breast microcalcification development develop microcalcification prevent
XXT	cancer lesions with microcalcifications cancer lesions without microcalcifications carcinogenesis associated with microcalcifications dysregulation is associated with microcalcifications, essential to consider microcalcifications existence of multiple microcalcifications involution and in presence of microcalcifications presence of abnormal microcalcifications stratified for the presence of microcalcifications suggests the existence of microcalcifications tumor tissue without microcalcifications.	cancer lesion microcalcification carcinogenesis associate microcalcification dysregulation associate microcalcification essential consider microcalcification exist multiple microcalcification involution presence microcalcification presence abnormal microcalcification stratify presence microcalcification suggest existence microcalcification tumor tissue microcalcification
XT	abnormal microcalcifications avoid microcalcifications Breast microcalcifications 5 developing microcalcifications hydroxyapatite microcalcifications 2 lesions with/without microcalcifications multiple microcalcifications presence of microcalcification 9	abnormal microcalcification avoid microcalcification Breast microcalcification 5 develop microcalcification hydroxyapatite microcalcification 2 lesion Microcalcification 4 multiple microcalcification presence microcalcification 9
TX	microcalcification resorption microcalcifications and the breast microcalcifications in breast	microcalcification resorption microcalcification breast 4

Проект построения мультимодальной медицинской информационной системы

TX	microcalcifications in the breast microcalcifications, breast microcalcifications development 4 microcalcifications formation 3	microcalcification development 4 microcalcification formation 3
----	--	--

Затем эксперт провел удаление повторяющихся биграмм, словосочетаний сущ. + гл., гл. + сущ. Потенциально новыми терминами могут быть следующие: а) состава ХТ – abnormal microcalcification, cell microcalcification, invasive Microcalcification, multiple microcalcification; б) состава TX – microcalcification primary, microcalcification production, microcalcification resorption.

Заключение

Представлен проект разработки мультимодальной медицинской системы с возможностью автоматизированного выделения новых медицинских терминов из полнотекстовых документов (научных статей, патентов), опубликованных в международных электронных базах. Предлагаемое решение позволяет отслеживать динамику изменения популярности медицинской терминологии. В данном проекте для каждого элемента словаря терминов предложено не только сохранение контекстов термина для пояснения его значений, но и загрузка публикаций – источников контекста, доступных по ссылке.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ и ГФЕН в рамках научного проекта № 21-57-53018

Acknowledgments

The reported study was funded by RFBR and NSFC, project number 21-57-53018.

Литература

1. Asooja K, Bordea G, Vulcu G, Buitelaar P. (2016) Forecasting Emerging Trends from Scientific Literature. Proceedings of the 10th International Conference on Language Resources and Evaluation, Portorož, Slovenia, 23–28 May, pp. 417–420.
2. Biganzoli L, Cardoso F, Beishon M, Cameron D, Cataliotti L, Coles C.E., Delgado Bolton R.C., Trill M.D., Erdem S, Fjell M., Geiss R., Goossens M., Kuhl C., Marotti L, Naredi P, Oberst S, Palussière J, Ponti A., Rosselli Del Turco M., Rubio I.T., Poortmans P. (2020). The requirements of a specialist breast centre. Breast (Edinburgh, Scotland), no. 51, pp. 65–84. <https://doi.org/10.1016/j.breast.2020.02.003>
3. Clemenceau A., Michou L., Diorio C., Durocher F. (2020). Breast Cancer and Microcalcifications: An Osteoimmunological Disorder? International journal of molecular sciences, no. 21 (22), p. 8613. <https://doi.org/10.3390/ijms21228613>
4. Dancy-Scott N, Dutcher G.A., Keselman A., Hochstein C., Coptly, C., Ben-Senia D., Rajan S., Asencio M.G., Choi J.J. (2018) Trends in HIV Terminology: Text Mining and Data Visualization Assessment of International AIDS Conference Abstracts Over 25 Years. JMIR Public Health Surveill, 4, e50. [CrossRef] [PubMed].
5. Duggan C., Dvaladze A., Rositch A.F., Ginsburg O., Yip C.H., Horton S., Camacho Rodriguez R., Eniu A., Mutebi M., Bourque J.M., Masood S., Unger-Saldaña K., Cabanes A., Carlson R.W,

- Gralow J.R., Anderson B.O.* (2020). The Breast Health Global Initiative 2018 Global Summit on Improving Breast Healthcare Through Resource-Stratified Phased Implementation: Methods and overview. *Cancer*, 126 Suppl 10(Suppl 10), pp. 2339–2352. <https://doi.org/10.1002/cncr.32891>
6. *Escala-Garcia M, Abraham J, Andrulis IL, Anton-Culver H, Arndt V. et al.* (2020) A network analysis to identify mediators of germline-driven differences in breast cancer prognosis. *Nat Commun*, vol. 16, no. 11 (1), p. 312. doi: 10.1038/s41467-019-14100-6. PMID: 31949161; PMCID: PMC6965101.
7. *Ginsburg O, Yip C.H., Brooks A., Cabanes A., Caleffi M. et al.* (2020). Breast cancer early detection: A phased approach to implementation. *Cancer*, 126 Suppl 10(Suppl 10), pp. 2379–2393. <https://doi.org/10.1002/cncr.32887>
8. *Helian Feng Alexander Gusev, Bogdan Pasaniuc Lang Wu, Jirong Long et al.* (2020) Transcriptome-wide association study of breast cancer risk by estrogen-receptor status. *Genet Epidemiol*, no. 44 (5), pp. 442–468. doi: 10.1002/gepi.22288. Epub 2020 Mar 1. PMID: 32115800; PMCID: PMC7987299
9. *Khakimova A.Kh., Zolotarev O.V., Berberova M.A.* (2021) Coronavirus Infection Study: Bibliometric Analysis of Publications on Covid-19 using PubMed and Dimensions Databases. *Scientific Visualization*, no. 12 (5), pp. 112–129. DOI: 10.26583/sv.12.5.10
10. *Lofterod T, Frydenberg H, Flote V, Eggen A.E., McTiernan A., Mortensen E.S., Akslen L.A., Reitan J.B., Wilsgaard T, Thune I.* (2020). Exploring the effects of lifestyle on breast cancer risk, age at diagnosis, and survival: the EBBA-Life study. *Breast cancer research and treatment*, no. 182 (1), pp. 215–227. <https://doi.org/10.1007/s10549-020-05679-2>
11. *Minouk J. Schoemaker, Hazel B. Nichols, Lauren B. Wright, Mark N. Brook, Michael E. Jones et al.* (2020) Adult weight change and premenopausal breast cancer risk: A prospective pooled analysis of data from 628,463 women. *Int J. Cancer*, Sep 1, no. 147 (5), pp. 1306–1314. doi: 10.1002/ijc.32892. PMCID: PMC7365745. NIHMSID: NIHMS1568040. PMID: 32012248
12. National Library of Medicine. PubMed.gov. Available at: <https://pubmed.ncbi.nlm.nih.gov>
13. Ontario Health (Quality) (2020). Gene Expression Profiling Tests for Early-Stage Invasive Breast Cancer: A Health Technology Assessment. Ontario health technology assessment series, no. 20 (10), pp. 1–234.
14. *Pashayan N., Antoniou A.C., Ivanus U., Esserman L.J., Easton D.F., French D., Sroczynski G., Hall P., Cuzick J., Evans D.G., Simard J., Garcia-Closas M., Schmutzler R., Wegwarth O., Pharoah P., Moorthie S., De Montgolfier S., Baron C., Herceg Z., Turnbull C., ..., Widschwendter M.* (2020). Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nature reviews. Clinical oncology*, no. 17 (11), pp. 687–705. <https://doi.org/10.1038/s41571-020-0388-9>
15. *Zolotarev O.V.* (2020) Research and development of linguo-statistical methods for forming a portrait of a subject area. *CEUR Workshop Proceedings*, no. 2763, pp. 167–175.
16. *Zolotarev O., Solomentsev Y., Khakimova A., Charnine M.* (2019) Identification of semantic patterns in full-text documents using neural network methods. *CEUR Workshop Proceedings*, no. 2485, pp. 276–279.

References

1. Asooja K, Bordea G, Vulcu G, Buitelaar P. (2016) Forecasting Emerging Trends from Scientific Literature. Proceedings of the 10th International Conference on Language Resources and Evaluation, Portorož, Slovenia, 23–28 May, pp. 417–420.
2. Biganzoli L., Cardoso F., Beishon M., Cameron D., Cataliotti L., Coles C.E., Delgado Bolton R.C., Trill M.D., Erdem S., Fjell M., Geiss R., Goossens M., Kuhl C., Marotti L., Naredi P., Oberst S., Palussière J., Ponti A., Rosselli Del Turco M., Rubio I.T., Poortmans P. (2020). *The requirements of a specialist breast centre*, no. 51, pp. 65–84. <https://doi.org/10.1016/j.breast.2020.02.003>
3. Clemenceau A., Michou L., Diorio C., Durocher F. (2020). Breast Cancer and Microcalcifications: An Osteoimmunological Disorder? *International Journal of molecular sciences*, no. 21 (22), p. 8613. <https://doi.org/10.3390/ijms21228613>
4. Dancy-Scott N., Dutcher G.A., Keselman A., Hochstein C., Coptly, C., Ben-Senia D., Rajan S.; Asencio M.G., Choi J.J. (2018) Trends in HIV Terminology: Text Mining and Data Visualization Assessment of International AIDS Conference Abstracts Over 25 Years. *JMIR Public Health Surveill*, 4, e50. [CrossRef] [PubMed].
5. Duggan C., Dvaladze A., Rositch A.F., Ginsburg O., Yip C.H., Horton S., Camacho Rodriguez R., Eniu A., Mutebi M., Bourque J.M., Masood S., Unger-Saldaña K., Cabanes A., Carlson R.W, Gralow J.R., Anderson B.O. (2020). The Breast Health Global Initiative 2018 Global Summit on Improving Breast Healthcare Through Resource-Stratified Phased Implementation: Methods and overview. *Cancer*, 126 Suppl 10(Suppl 10), pp. 2339–2352. <https://doi.org/10.1002/cncr.32891>
6. Escala-Garcia M, Abraham J, Andrulis IL, Anton-Culver H, Arndt V. et al. (2020) A network analysis to identify mediators of germline-driven differences in breast cancer prognosis. *Nat Commun*, vol. 16, no. 11 (1), p. 312. doi: 10.1038/s41467-019-14100-6. PMID: 31949161; PMCID: PMC6965101.
7. Ginsburg O., Yip C.H., Brooks A., Cabanes A., Caleffi M. et al. (2020). Breast cancer early detection: A phased approach to implementation. *Cancer*, 126 Suppl 10(Suppl 10), pp. 2379–2393. <https://doi.org/10.1002/cncr.32887>
8. Helian Feng Alexander Gusev, Bogdan Pasaniuc Lang Wu, Jirong Long et al. (2020) Transcriptome-wide association study of breast cancer risk by estrogen-receptor status. *Genet Epidemiol*, no. 44 (5), pp. 442–468. doi: 10.1002/gepi.22288. Epub 2020 Mar 1. PMID: 32115800; PMCID: PMC7987299
9. Khakimova A.Kh., Zolotarev O.V., Berberova M.A. (2021) Coronavirus Infection Study: Bibliometric Analysis of Publications on Covid-19 using PubMed and Dimensions Databases. *Scientific Visualization*, no. 12 (5), pp. 112–129. DOI: 10.26583/sv.12.5.10
10. Lofterød T., Frydenberg H., Flote V., Eggen A.E., McTiernan A., Mortensen E.S., Akslén L.A., Reitan J.B., Wilsgaard T., Thune I. (2020). Exploring the effects of lifestyle on breast cancer risk, age at diagnosis, and survival: the EBBA-Life study. *Breast cancer research and treatment*, no. 182 (1), pp. 215–227. <https://doi.org/10.1007/s10549-020-05679-2>

11. Minouk J. Schoemaker, Hazel B. Nichols, Lauren B. Wright, Mark N. Brook, Michael E. Jones et al. (2020) Adult weight change and premenopausal breast cancer risk: A prospective pooled analysis of data from 628,463 women. *Int. J. Cancer*, Sep 1, no. 147 (5), pp. 1306–1314. doi: 10.1002/ijc.32892. PMID: 32012248
12. National Library of Medicine. PubMed.gov. Available at: <https://pubmed.ncbi.nlm.nih.gov>
13. Ontario Health (Quality) (2020). Gene Expression Profiling Tests for Early-Stage Invasive Breast Cancer: A Health Technology Assessment. *Ontario health technology assessment series*, no. 20 (10), pp. 1–234.
14. Pashayan N., Antoniou A.C., Ivanus U., Esserman L.J., Easton D.F., French D., Sroczynski G., Hall P., Cuzick J., Evans D.G., Simard J., Garcia-Closas M., Schmutzler R., Wegwarth O., Pharoah P., Moorthie S., De Montgolfier S., Baron C., Hecceg Z., Turnbull C., ..., Widschwendter M. (2020). Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nature reviews. Clinical oncology*, no. 17 (11), pp. 687–705. <https://doi.org/10.1038/s41571-020-0388-9>
15. Zolotarev O.V. (2020) Research and development of linguo-statistical methods for forming a portrait of a subject area. *CEUR Workshop Proceedings*, no. 2763, pp. 167–175.
16. Zolotarev O., Solomentsev Y., Khakimova A., Charnine M. (2019) Identification of semantic patterns in full-text documents using neural network methods. *CEUR Workshop Proceedings*, no. 2485, pp. 276–279.