

## Раздел II. Алгоритмы обработки информации

УДК 004.021

DOI 10.18522/2311-3103-2022-3-119-129

**М.С. Анферова, А.М. Белевцев**

### **АНАЛИЗ ТРЕБОВАНИЙ И РАЗРАБОТКА АЛГОРИТМОВ ИНТЕЛЛЕКТУАЛЬНЫХ СЕРВИСОВ МОНИТОРИНГА**

*Рассмотрены проблемы стратегического анализа и выбора направлений развития инновационных предприятий в условиях перехода к 6 технологическому укладу и индустрии 4.0. Определены основные уровни анализа. Обозначены цели стратегического анализа исходя из масштаба проводимого исследования. Выделены задачи анализа, решение которых позволит достичь поставленных целей. Показана сложность решения задач глобального мониторинга, которые обусловлены большим объемом разнородной и неструктурированной информации. В этих условиях тематический поиск и аналитическая обработка информации не могут быть выполнены без применения автоматизированных информационно-аналитических систем и создания поисковых сервисов на базе искусственного интеллекта. Предложена общая процедура мониторинга. Определены основные этапы мониторинга технологических трендов, показаны задачи, решаемые в рамках конкретного этапа и планируемый результат. На основе общей процедуры мониторинга определены основные приоритетные функции, которыми должны обладать разрабатываемые сервисы. А также проблемы их разработки и структуризация полученной информации в виде информационных объектов и кластеризация документов. В отличие от известных систем глобального мониторинга, в которых поиск основан на индикаторах: рост использования ключевых слов, увеличение численности новых авторов, цитирование работ из смежных областей. Предложены алгоритмы, обеспечивающие определение опорных тем, оценку ранжирования и релевантности информации. Дано описание работы алгоритмов на примере создания сводной информационной таблицы, с помощью которой происходит формирование взаимосвязей документов научно-технологического развития по заданному направлению мониторинга и поиск конкретных документов в базе данных. Построение поисковых сервисов на основе представленных алгоритмов обеспечит выделение опорных тем документов, предоставит более достоверные результаты кластеризации неструктурированной информации и формирования научно-технологических трендов, в информационно-аналитических комплексах. Для реализации алгоритма предлагается использовать язык программирования Python. Внедрение данных алгоритмов повысит качество и эффективность информационного поиска в условиях большого объема неструктурированной информации.*

*Технологические тренды; мониторинг; искусственный интеллект; Big Data; алгоритм; распознавание текста; кластеризация.*

**M.S. Anferova, A.M. Belevtsev**

### **ANALYSIS OF REQUIREMENTS AND DEVELOPMENT OF ALGORITHMS FOR INTELLIGENT MONITORING SERVICES**

*The paper considers the problems of strategic analysis and the choice of directions for the development of innovative enterprises in the conditions of transition to the 6th technological order and industry 4.0. The main levels of analysis are determined. The objectives of the strategic analysis are outlined based on the scale of the research being conducted. The analysis tasks are highlighted, the solution of which will allow achieving the set goals. The complexity of solving global monitoring tasks, which are caused by a large volume of heterogeneous and unstructured information, is shown. In these conditions, thematic search and analytical processing of information cannot be performed without the use of auto-*

*mated information and analytical systems and the creation of search services based on artificial intelligence. A general monitoring procedure is proposed. The main stages of monitoring technological trends are defined, the tasks to be solved within a specific stage and the planned result are shown. Based on the general monitoring procedure, the main priority functions that the developed services should have are determined. As well as the problems of their development and structuring of the received information in the form of information objects and clustering of documents. In contrast to the well-known global monitoring systems, in which the search is based on indicators: an increase in the use of keywords, an increase in the number of new authors, quoting works from related fields. Algorithms are proposed that provide the definition of reference topics, assessment of ranking and relevance of information. The description of the algorithms is given on the example of creating a summary information table, with the help of which the interrelationships of documents of scientific and technological development in each direction of monitoring and the search for specific documents in the database are formed. The construction of search services based on the presented algorithms will ensure the allocation of reference topics of documents, provide more reliable results of clustering of unstructured information and the formation of scientific and technological trends in information and analytical complexes. To implement the algorithm, it is proposed to use the Python programming language. The implementation of these algorithms will improve the quality and efficiency of information retrieval in conditions of a large volume of unstructured information.*

*Technological trends; monitoring; search robot; artificial intelligence; Big Data; algorithm; text recognition; clustering.*

**Введение.** Структурные, экономические и политические изменения в мировой экономике приводят к резкому обострению борьбы за технологического превосходство. В этой связи задача стратегического анализа и определения направлений и трендов научно-технологического развития становится чрезвычайно актуальной [1, 2].

Ее решение невозможно без создания проблемно-ориентированных, информационно-аналитических комплексов [3] и разработки эффективных процедур мониторинга включая поиск, извлечение и анализ информации на основе технологий искусственного интеллекта, способных работать в условиях:

- ◆ чрезвычайно быстрой динамик формирования новых направлений и (трендов) и траекторий их развития;
- ◆ взрывного характера нарастания объемов разнородной и не структурированной информации (Big Data [4]);
- ◆ резкого сокращение длительности инновационного цикла [5].

**Основная часть.** Цели и задачи стратегического анализа определяются уровнем и масштабом проводимого исследования [6]:

- ◆ Международный.
- ◆ Государственный.
- ◆ Корпоративный.

На международном уровне основными задачами являются [7]:

- ◆ создание механизмов наднационального регулирования научно-технологической сферы;
- ◆ выработка совместных программ в рамках объединений стран;
- ◆ эффективная интеграция и стандартизация деятельности в сфере науки, технологий и инноваций [8].

На Государственном уровне:

- ◆ формирование общего представления о состоянии мировой науки и технологий;

- ◆ выявление конкурентных преимуществ страны в ключевых областях и др.

На корпоративном уровне:

- ◆ определение технологических и продуктовых трендов заданных предметных областях;
- ◆ выявление потенциальных зон научно-технологического прорыва;
- ◆ анализ конкурентного окружения и др.

Для достижения поставленных целей необходимо обеспечить решение следующих задач анализа:

1. Выявление направлений фундаментальных и поисковых исследований, методов и способов, как основы для создания технологий нового поколения.
2. Выявление зарождающихся замещающих технологий приводящих к формированию новых рынков (разрушающих инноваций). В том числе, определение критических точек т.е. прогноз момента времени, когда это замещение произойдет.
3. Выявление зарождающихся новых функциональных технологий и технологических направлений, способных кардинально изменить функциональные возможности и технические характеристики технических решений, создаваемых на их основе. (полностью менять структуру рынков - разрушающие инновации).
4. Выявление зарождающихся направлений создания технических решений на базе новых функциональных технологий (инновационных продуктов), способных полностью менять структуру рынков - разрушающие инновации).
5. Выявление инновационных продуктовых, процессных, организационных и маркетинговых технологий.
6. Выявление технологических направлений для широкого междисциплинарных спектра областей применения.
7. Выявление требуемых компетенций для участия в программах перспективных направлений
8. Анализ внешних ограничений и воздействий на предметную область и проекты рассматриваемого предприятия.
9. Оценка возможности участия в выбранных проектах на базе существующих компетенций и технологического базиса рассматриваемого предприятия, в том числе и в условиях внешних ограничений.

Информационной логистической основой для решения вышеперечисленных задач является создание динамически формируемых и обновляемых на основе эффективных процедур мониторинга специализированных баз данных [9].

Общую процедуру мониторинга можно представить в виде следующих взаимосвязанных этапов (табл. 1):

Таблица 1

**Этапы мониторинга технологических трендов**

№	Этап	Процесс	Результат
1	Постановка задачи.	Выбор предметной области. Определяется цель исследования	Выбор релевантных источников данных
2	Сбор информации	Формирование запроса и осуществление сбора сведений	Формирование базы данных.
3	Обработка полученной информации	Очистка, фильтрация и автоматизированный анализ собранной информации	Формирование наиболее перспективных для рассматриваемой компании направлений (трендов)
4	Представление, интерпретация и обобщения результатов	Разработка предложений для принятия управленческих решений	Визуализация полученных результатов

Практическая реализация предлагаемой процедуры невозможна без применения автоматизированных информационно-аналитических систем и специальных поисковых сервисов с применениями методов искусственного интеллекта [10].

**Анализ требований к интеллектуальным сервисам мониторинга.** Можно выделить следующие приоритетные функции, которые должны обеспечивать разрабатываемые сервисы [11]:

Ф1. Ввод новых информационных объектов.

Цель: Обеспечение в ходе мониторинга ввода в базу данных информационных объектов различного типа и их индексация:

Ф2. Структуризация полученной информации в виде информационных объектов (тем). Определение зарождающихся, замещающих и критических технологий Автоматизированное формирование стратегических групп заданного предприятия. Формирование ключевых слов заданной тематической области

Цель: предварительное фиксирование всех значимых научно-технических направлений, выявляемых в ходе мониторинга и осуществление параллельной привязки к ним соответствующих информационных объектов (документов, предприятий, продукции), а также дальнейшее итоговое формирование направлений развития тематической области.

Ф3. Функция ранжирования соответствующих поисковому запросу документов.

Цель: Кластеризация найденных документов с целью упрощения их дальнейшего восприятия и фильтрации аналитиком.

Ф4. Полнотекстовый поиск документов в пределах базы информационных ресурсов информационно-аналитического комплекса.

Цель: Формирование взаимосвязей документов стратегического развития по заданному направлению

Ф5. Автоматизированная проверка актуальности базы информационных ресурсов с целью реализации задачи непрерывного мониторинга.

Цель: Динамическое расширение навигации мониторинга

При этом наибольшую сложность представляет собой проблема определение зарождающихся, замещающих и критических технологий и кластеризации документов (Ф2, Ф3) [12].

Проведенный анализ исследовательских программ показал, что для реализации данных функций в специализированной базе данных эффективно применять методы искусственного интеллекта (ИИ). Для решения поставленных задач были разработаны следующие алгоритмы:

- ◆ Алгоритм определения опорных тем.
- ◆ Алгоритм оценки ранжирования и релевантности информации.

#### **Алгоритм определения опорных тем**

Процедура формирования трендов может быть основана на определении опорных тем в документах, хранящихся в базе данных.

Предлагаемый алгоритм основан на определении тем по ключевым словам. Алгоритм включает в себя следующие основные этапы (рис. 1):

- ◆ Разделение данных по языкам.
- ◆ Очистка данных с помощью токенизации, лемматизации и стемминга.
- ◆ Векторизация.
- ◆ Извлечение ключевых слов.

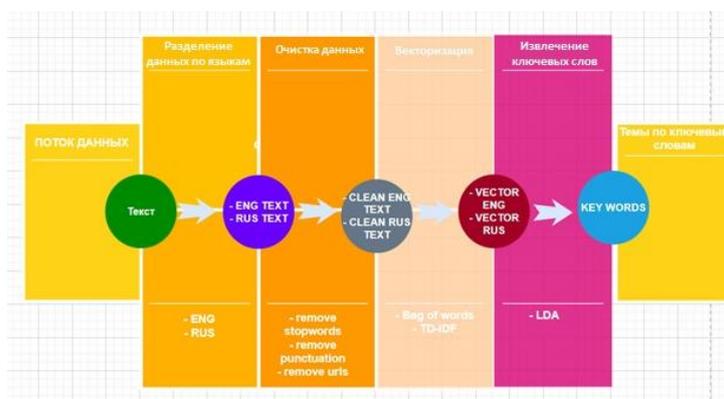


Рис. 1. Алгоритм определения опорных тем

Рассмотрим основные этапы предлагаемого алгоритма:

- ◆ Векторизация.
- ◆ Извлечение ключевых слов.

Векторизация – процесс преобразования слов в числа [13].

В машинном обучении векторизация – это этап извлечения признаков. Необходимо извлечь из текста некоторые отличительные признаки для обучения модели путем преобразования текста в числовые векторы [14].

Вес термина («важность» слова для идентификации данного текста) в документе можно определить методом TF-IDF.

TF (term frequency – частота слова) – отношение числа вхождений некоторого слова к общему числу слов документа [15]. Таким образом, оценивается важность слова в пределах отдельного документа.

$$TF(t, d) = \frac{n_t}{\sum_k n_k}, \quad (1)$$

где  $n_t$  – число вхождений слова  $t$  в документ, а в знаменателе — общее число слов в данном документе.

В качестве меры взвешивания слова можно воспользоваться F-мерой, а именно precision (точность) и recall (полнота) – это метрики, которые используются при оценке большей части алгоритмов извлечения информации. Суть точности и полноты таких мер очень проста [16].

Для решения поставленных задач удобнее будет использовать именно инвертированный индекс, дополненный мерой TF-IDF, так как он полностью удовлетворяет потребность быстрого полнотекстового поиска по коллекции документов, а за счёт дополнительного поля, содержащего частоту термов, мы сможем использовать его для организации ранжирования, а также в качестве кэша векторов документов, хранящихся в специализированной базе данных.

**Извлечение ключевых слов.** Модели анализа тем способны обнаруживать темы в тексте с помощью передовых алгоритмов машинного обучения, которые подсчитывают слова и находят и группируют похожие шаблоны слов [17]. Проведенный анализ показал эффективность использования метода LDA.

LDA (Latent Dirichlet Allocation – Латентное размещение Дирихле)

Алгоритм работы данного метода:

1. Определяется фиксированный набор тем, состоящих из представленных (неизвестных) наборов слов. Данные темы уже находятся в документах базы данных, но мы еще их не знаем.

2. LDA пытается сопоставить все (известные) документы с (неизвестными) темами таким образом, чтобы слова в каждом документе были в основном захвачены этими темами [18, 19].

Документы с похожими темами будут использовать одинаковые слова. Также предполагается, что каждый документ состоит из смеси тем, и каждое слово имеет вероятность принадлежности к определенной теме.

LDA предполагает, что документы генерируются следующим образом: выбирается смесь тем (например, 20% темы А, 80% темы В и 0% темы С), а затем выбираются слова, которые принадлежат этим темам. Слова выбираются случайным образом в соответствии с вероятностью их появления в определенном документе (рис. 2).



Рис. 2. Алгоритм генерирования документов для LDA

3. LDA видит документ и предполагает, что он был создан, как описано выше. Затем он работает в обратном направлении - от слов, составляющих документ, и пытается угадать смесь тем, которая привела к этому конкретному расположению слов (рис. 3).

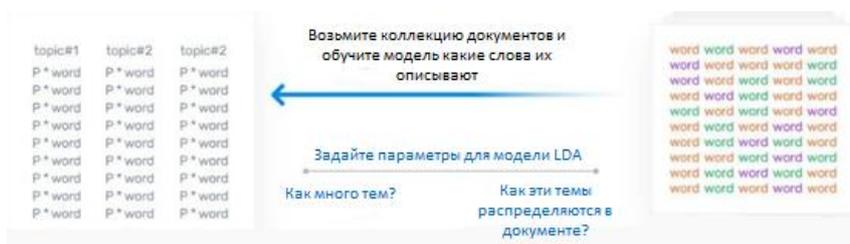


Рис. 3. Алгоритм распознавания тем LDA

В результате выполнения данного алгоритма будут выделены опорные темы документов, хранящихся в базе данных, а также процентное соотношение каждой темы в конкретном документе, что обеспечит более достоверные результаты формирования трендов, путем недопущения субъективных оценок аналитиков.

#### Алгоритм оценки ранжирования и релевантности информации

Релевантность в информационном поиске – это семантическое соответствие поискового запроса полученному документу [20].

Введем понятие формальной релевантности – соответствие, определяемое путём сравнения образа поискового запроса с поисковым образом документа.

Вес термина («важность» слова для идентификации данного текста) в документе можно определить разными способами. Если в разрабатываемом поисковом сервисе за меру веса термина берем меру TF-IDF (алгоритм которого был приведен выше), то это позволит нам использовать поисковый индекс в качестве кэша для хранения векторов документов.

С помощью алгоритма LDA мы получили не только список потенциальных тем, но и процентное соотношение (вес) каждого ключевого слова в конкретном документе. Причем, чем выше вес, тем больше вероятность того, что это подходящая тема для текста (рис. 4).

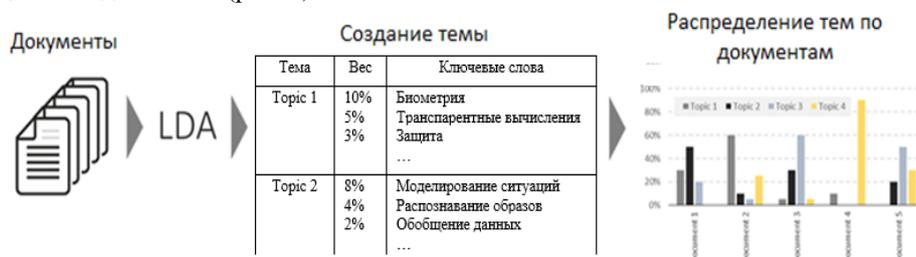


Рис. 4. Результат работы алгоритма LDA

Предположим, что мы всегда будем брать слово с наибольшим весом в качестве темы данного текста. Остальные темы необходимо сохранить в отдельный файл (мы будем рассматривать их как ключевые слова). Тогда алгоритм ранжирования и релевантности информации будет выглядеть следующим образом (рис. 5):



Рис. 5. Алгоритм оценки ранжирования и релевантности информации

На основании результатов алгоритма определения опорных тем можно создать сводную информационную таблицу для оценки ранжирования и релевантности информации (табл. 2) с колонками "файл", "тема", "ключевые слова", где в колонке:

- ◆ "файл" – хранится ссылка на файл, определенная TF-IDF;
- ◆ "тема" – тема, определенная LDA;
- ◆ "ключевые слова" – сохранение остальных ключевых слов.

Имея такую таблицу, будет легко модифицировать ее в будущем (если мы будем изучать модель с нуля), но также относительно легко подключить ее к существующей базе данных.

Используя таблицу, описанную в предыдущем разделе, облегчается поиск конкретных документов (Ф4). Достаточно выбрать определенные ключевые слова из таблицы, и мы сразу получаем доступ к заданным файлам (столбец файл).

Таблица 2

## Сводная информационная таблица

№	Файл	Тема	Ключевые слова
1.	Document 1	Технологии ИИ	Моделирование ситуаций Распознавание образов Обобщение данных ...
2.	Document 2	Кибербезопасность	Биометрия Транспарентные вычисления Защита ...
3.	Document 3	Кибероружие	Кибератака Кибервойна Самоуничтожение ...
4.	...	...	...

**Выводы.** Задача проведения непрерывного глобального мониторинга научно-технологических направлений, технологических изменений и тенденций развития актуальна и требует создания новых методов обработки и анализа информации [21].

Для решения поставленной задачи сформированы алгоритмы интеллектуального сервиса поиска и мониторинга информации в специализированной базе данных.

Внедрение данных алгоритмов в специализированную базу данных информационно-аналитического комплекса АРМ Аналитика [4] позволило сократить общее время поиска информации в 5-6 раз по отношению к запросам, формируемым в поисковых системах общего назначения.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Белевцев А.М., Балыбердин В.А., Бендерский Г.П., Белевцев А.А. Анализ направлений развития нано- и IT-технологий для построения специализированных сетевых коммуникационных систем нового поколения // Известия ЮФУ. Технические науки. – 2015. – № 3 (164). – С. 35-45.
2. Микова Н.С., Соколова А.В. Мониторинг глобальных технологических трендов: теоретические основы и лучшие практики // Форсайт. – 2014. – Т. 8, № 4.
3. Анферова М.С., Белевцев А.М. Анализ направлений создания алгоритмов эффективного поиска информации в сетях общего и специального назначения // Матер. III Всероссийской научно-технической конференции «Актуальные проблемы современной науки и производства». – Рязань: РГРТУ, 2018.
4. Белевцев А.М., Садреев Ф.Г., Белевцев А.А., Балыбердин В.А. Разработка интеллектуальных сервисов мониторинга технологических трендов в информационно-аналитических комплексах // Научно-технологические тренды. – 2019. – Т. 20, № 3. – С. 24-29.
5. Анферова М.С., Белевцев А.М. Разработка алгоритмов интеллектуального сервиса поиска и мониторинга информации // Известия ЮФУ. Технические науки. – 2021. – № 3. – С. 6-17.
6. Шваб К. Четвертая промышленная революция. – М.: Эксмо, 2018. – 285 с. – ISBN 978-5-699-98379-7.
7. Загородников А.Н. Управление общественными связями в бизнесе: учебник. – М.: Крокс, 2013. (гл.3 «SWOT-анализ: сущность, цель, содержание»).
8. Tanya Sammut-Bonnici, David Galea. PEST analysis // Wiley Encyclopedia of Management. – Chichester, UK: John Wiley & Sons, Ltd, 2015-01-22.
9. Philip Kotler, Roland Berger, Nils Bickhoff. The Quintessence of Strategic Management: What You Really Need to Know to Survive in Business.

10. *Анферова Маргарита Сергеевна, Белевцев Андрей Михайлович.* Анализ направлений развития технологий мониторинга в условиях большого объема неструктурированной информации // XXIV Всероссийская научно-техническая конференция с международным участием имени профессора О.Н. Пьявченко "Компьютерные и информационные технологии в науке, инженерии и управлении" «КомТех-2020».
11. *Анферова М.С., Белевцев А.М.* Поисковые роботы для автоматизированного мониторинга информации в сетях общего и специального назначения // 18-я Международная научно-практическая конференция «Управление качеством» 2019 г.
12. *Анферова М.С., Белевцев А.М.* Общая концепция создания технологии интеллектуального поиска информации в сетях общего и специального назначения // XXV Всероссийская научно-техническая конференция с международным участием имени профессора О.Н. Пьявченко "Компьютерные и информационные технологии в науке, инженерии и управлении" «КомТех-2021».
13. *Salton G. and Buckley C.* Term-weighting approaches in automatic text retrieval // *Information Processing & Management.* – 1988. – Vol. 24 (5). – P. 513-523.
14. *Jacob Devlin and Ming-Wei Chang.* Research Scientists, Google AI Language: Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing (англ.). – Google, Inc, 2018.
15. *Charles L.A. Clarke, Gordon V. Cormack.* Dynamic Inverted Indexes for a Distributed Full-Text Retrieval System (англ.) // *MultiText Project Technical Report MT-95-01.* – University of Waterloo, Waterloo, Ontario N2L 3G1, Canada, 1995.
16. *Павлов Ю.Н., Майструк К.А.* Сравнение методов оценки тональности текста // *Молодой ученый.* – 2016. – № 12 (116). – С. 59-64.
17. *Liu X. and Croft W.B.* Cluster-based retrieval using language models // *In Proceedings of SIGIR '04.* – 2004. – P. 186-193.
18. *Blei D.M., Ng A.Y., and Jordan M.J.* Latent Dirichlet allocation // *In Journal of Machine Learning Research.* – 2003. – No. 3. – P. 993-1022.
19. *Teh Y.W., Jordan M.I., Beal M.J., and Blei D.M.* Hierarchical Dirichlet processes // *Technical Report, Department of Statistics, UC Berkeley,* 2004.
20. *Словарь по кибернетике / под ред. академика В.С. Михалевича.* – 2-е изд. – Киев: Главная редакция Украинской Советской Энциклопедии имени М.П. Бажана, 1989. – 751 с. – (С48).
21. *Анферова М.С., Белевцев А.М.* Разработка алгоритмов интеллектуального сервиса поиска и мониторинга информации // *Известия ЮФУ. Технические науки.* – 2021. – № 3. – С. 6-17.

#### REFERENCES

1. *Belevtsev A.M., Balyberdin V.A., Benderskiy G.P., Belevtsev A.A.* Analiz napravleniy razvitiya nano- i IT-tekhnologiy dlya postroeniya spetsializirovannykh setevykh kommunikatsionnykh sistem novogo pokoleniya [Analysis of the directions of development of nano- and IT-technologies for the construction of specialized network communication systems of a new generation], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2015, No. 3 (164), pp. 35-45.
2. *Mikova N.S., Sokolova A.V.* Monitoring global'nykh tekhnologicheskikh trendov: teoreticheskie osnovy i luchshie praktiki [Monitoring global technological trends: theoretical foundations and best practices], *Forsayt* [Foresight], 2014, Vol. 8, No. 4.
3. *Anferova M.S., Belevtsev A.M.* Analiz napravleniy sozdaniya algoritmov effektivnogo poiska informatsii v setyakh obshchego i spetsial'nogo naznacheniya [Analysis of the directions of creating algorithms for effective information retrieval in general and special purpose networks], *Mater. III Vserossiyskoy nauchno-tekhnicheskoy konferentsii «Aktual'nye problemy sovremennoy nauki i proizvodstva»* [Materials of the III All-Russian Scientific and Technical Conference "Actual problems of modern science and production"]. Ryazan': RGRTU, 2018.
4. *Belevtsev A.M., Sadreev F.G., Belevtsev A.A., Balyberdin V.A.* Razrabotka intellektual'nykh servisov monitoringa tekhnologicheskikh trendov v informatsionno-analiticheskikh kompleksakh [Development of intelligent services for monitoring technological trends in information and analytical complexes], *Naukoemkie tekhnologii* [High-tech technologies], 2019, Vo.. 20, No. 3, pp. 24-29.
5. *Anferova M.S., Belevtsev A.M.* Razrabotka algoritmov intellektual'nogo servisa poiska i monitoringa informatsii [Development of algorithms for intelligent information search and monitoring service], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 3, pp. 6-17.

6. *Shvab K.* Chetvertaya promyshlennaya revolyutsiya [The Fourth Industrial Revolution]. Moscow: Eksmo, 2018, 285 p. ISBN 978-5-699-98379-7.
7. *Zagorodnikov A.N.* Upravlenie obshchestvennymi svyazyami v biznese: uchebnik [Management of public relations in business: textbook]. Moscow: Kro-kus, 2013.
8. *Tanya Sammut-Bonnici, David Galea.* PEST analysis // Wiley Encyclopedia of Management. Chichester, UK: John Wiley & Sons, Ltd, 2015-01-22.
9. *Philip Kotler, Roland Berger, Nils Bickhoff.* The Quintessence of Strategic Management: What You Really Need to Know to Survive in Business.
10. *Anferova Margarita Sergeevna, Belevtsev Andrey Mikhaylovich.* Analiz napravleniy razvitiya tekhnologiy monitoringa v usloviyakh bol'shogo ob"ema nestrukturirovannoy informatsii [Analysis of trends in the development of monitoring technologies in conditions of a large volume of unstructured information], XXIV Vserossiyskaya nauchno-tekhnicheskaya konferentsiya s mezhdunarodnym uchastiem imeni professora O.N. P'yavchenko "Komp'yuternye i informatsionnye tekhnologii v nauke, inzhenerii i upravlenii" «KomTekh-2020» [XXIV All-Russian Scientific and Technical Conference with international participation named after Professor O.N. P'yavchenko "Computer and information technologies in science, engineering and management" "Comtech-2020"].
11. *Anferova M.S., Belevtsev A.M.* Poiskovye roboty dlya avtomatizirovannogo monitoringa informatsii v setyakh obshchego i spetsial'nogo naznacheniya [Search robots for automated monitoring of information in general and special purpose networks], 18-ya Mezhdunarodnaya nauchno-prakticheskaya konferentsiya «Upravlenie kachestvom» 2019 g. [18th International Scientific and Practical Conference "Quality Management" 2019].
12. *Anferova M.S., Belevtsev A.M.* Obshchaya kontseptsiya sozdaniya tekhnologii intellektual'nogo poiska informatsii v setyakh obshchego i spetsial'nogo naznacheniya [The general concept of creating a technology for intelligent information retrieval in general and special purpose networks], XXV Vserossiyskaya nauchno-tekhnicheskaya konferentsiya s mezhdunarodnym uchastiem imeni professora O.N. P'yavchenko "Komp'yuternye i informatsionnye tekhnologii v nauke, inzhenerii i upravlenii" «KomTekh-2021» [XXV All-Russian Scientific and Technical Conference with international participation named after Professor O.N. P'yavchenko "Computer and information technologies in science, engineering and management" "Comtech-2021"].
13. *Salton G. and Buckley C.* Term-weighting approaches in automatic text retrieval, *Information Processing & Management*, 1988, Vol. 24 (5), pp. 513-523.
14. *Jacob Devlin and Ming-Wei Chang.* Research Scientists, Google AI Language: Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing (англ.). Google, Inc, 2018.
15. *Charles L.A. Clarke, Gordon V. Cormack.* Dynamic Inverted Indexes for a Distributed Full-Text Retrieval System (англ.), *MultiText Project Technical Report MT-95-01*. – University of Waterloo, Waterloo, Ontario N2L 3G1, Canada, 1995.
16. *Pavlov Yu.N., Maystruk K.A.* Sravnenie metodov otsenki tonal'nosti teksta [Comparison of methods for assessing the tonality of the text], *Molodoy uchenyy* [Young scientist], 2016, No. 12 (116), pp. 59-64.
17. *Liu X. and Croft W.B.* Cluster-based retrieval using language models, *In Proceedings of SIGIR '04*, 2004, pp. 186-193.
18. *Blei D.M., Ng A.Y., and Jordan M.J.* Latent Dirichlet allocation, *In Journal of Machine Learning Research*, 2003, No. 3, pp. 993-1022.
19. *Teh Y.W., Jordan M.I., Beal M.J., and Blei D.M.* Hierarchical Dirichlet processes, *Technical Report, Department of Statistics, UC Berkeley*, 2004.
20. Slovar' po kibernetike [Dictionary of Cybernetics], ed. by akademika V.S. Mikhalevicha. 2<sup>nd</sup> ed. Kiev: Glavnaya redaktsiya Ukrainskoy Sovetskoy Entsiklopedii imeni M.P. Bazhana, 1989, 751 p. (S48).
21. *Anferova M.S., Belevtsev A.M.* Razrabotka algoritmov intellektual'nogo servisa poiska i monitoringa informatsii [Development of algorithms for intelligent information search and monitoring service], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 3, pp. 6-17.

Статью рекомендовал к опубликованию д.т.н., профессор В.А. Балыбердин.

**Анферова Маргарита Сергеевна** – Московский авиационный институт (национальный исследовательский университет); e-mail: gludkina@yandex.ru; г. Москва, Россия; тел.: +79055220749; старший преподаватель.

**Белевцев Андрей Михайлович** – e-mail: ambelevtsev@yandex.ru; тел.: +79037691788; д.т.н.; профессор.

**Anferova Margarita Sergeevna** – Moscow Aviation Institute (National Research University); e-mail: gludkina@yandex.ru; Moscow, Russia; phone: +79055220749; senior lecturer.

**Belevtsev Andrey Michailovitch** – e-mail: ambelevtsev@yandex.ru; phone: +79037691788; dr. of eng. sc.; professor.

УДК 004.032

DOI 10.18522/2311-3103-2022-3-129-139

**С.М. Гушанский, В.И. Божич, В.С. Потапов**

### **ИССЛЕДОВАНИЕ И РАЗРАБОТКА КВАНТОВОГО КОДА ДЛЯ ИСПРАВЛЕНИЯ ОШИБОК**

*Квантовая коррекция ошибок (ККО) требуется в квантовых компьютерах для смягчения влияния ошибок на физические кубиты. Цель состоит в том, чтобы оптимизировать нейронную сеть для высокой производительности декодирования, сохраняя при этом минималистическую аппаратную реализацию. Ошибки, связанные с декогеренцией, можно уменьшить, приняв схемы ККО, которые кодируют несколько несовершенных физических кубитов в логическое квантовое состояние, аналогично классической коррекции ошибок. Актуальность данных исследований заключается в математическом и программном моделировании и реализации корректирующих кодов для исправления нескольких видов квантовых ошибок в рамках разработки и выполнения квантовых алгоритмов для решения классов задач классического характера. Научная новизна данного направления выражается в исключении одного из недостатков квантового вычислительного процесса. Разработка теории и принципов построения моделирующих систем, устойчивых к внешним помехам (зависимость искажения данных от зашумленности, зависимость ошибки квантового вычислительного процесса от меры и чистоты запутанности) для моделирования квантовых вычислений является динамичной областью, о чем свидетельствует большое количество существующих моделей отражающих те или иные квантовые вычислительные процессы и явления (квантовая телепортация, параллелизм, запутанность квантовых состояний) и научных трудов. Хотя квантовые вычисления еще не готовы к переходу от теории к практике, тем не менее, можно обоснованно догадываться какую форму, возможно, квантовый компьютер примет, или, что более важно для дизайнера языка программирования, по какому интерфейсу можно будет взаимодействовать с таким квантовым компьютером. Естественно применить уроки, извлеченные из программирования классических вычислений к квантовым вычислениям. Проведенный анализ работ данной области показал, что в настоящее время достигнут новый качественный уровень, открывающий перспективные возможности по реализации многокубитовых квантовых вычислений. Перспективы реализации и развития связаны не только с технологическими возможностями, но и с решением вопросов построения эффективных квантовых систем решения актуальных математических задач, задач криптографии и задач управления (оптимизации).*

*Моделирование; квантовый алгоритм; кубит; модель квантового вычислителя; запутывание; суперпозиция; квантовый оператор.*