

CONTENTS

TECHNOLOGY

Vedeneev V. A., Ershov E. V., Kovyrshin R. G. Metallurgical enterprise knowledge management model based on embedding models	7
Egorova L. G., Kalitaev A. N., Logunova O. S., Panov D. A. Simulation model of the Kuka youBot robotic manipulator for stacking and unloading parcels	17
Kochkin D. V., Gorbunov V. A. Analysis of the current state of multi-agent systems and promising directions of development	28
Nuzhin D. V., Andreev S. M., Mukhina E. Yu. An optimal control system for heating workpieces in a through furnace using a digital twin	41
Rychago M. E. Predicting the security level of a penitentiary institution based on a machine learning model	59

PHILOLOGY

Arinova B. N. Irrealis in judicial discourse: points and counterpoints of semantics, pragmatics and rhetoric	72
Divakova E. A. Duality of the Belovodye image in the story "Scarlet Snowdrifts" by V. Ya. Shishkov	83
Dudurich O. V. Umnyi vs smart: linguistic evidence of the digital epoch	95
Ilyina E. N. Morphemic analysis of a dialect word: to the problem of compiling a dictionary of dialect root nests and affixal paradigms	107
Erokhina E. A., Lavrova S. Yu. On the role of concepts in the verbal self-portrait of the artist (based on Vincent Van Gogh's epistolary discourse).....	121
Mukhina E. A. The <i>corporeal man</i> cluster in Russian spiritual verses	132
Chernyak M. A., Somov D. A. Gamification as a strategy of modern Russian literature (based on V. Pelevin's novels)	143

PEDAGOGY

Ivanova N. V., Borisova M. V. Personalized approach in the system of teachers' professional growth: prospects for the development of methodological support in preschool education	160
Lapin A. S., Smolonskaia A. N. Conceptual aspects of teacher's personal growth as a condition of readiness for sustainable development of moral values in primary schoolchildren	172
Pryadekhina M. S. Project activity as a means of promoting ecological thinking of schoolchildren aged 12–14	181

REVIEWS

Fedulenkova T. N. Variability vs stylistic differentiation of language. Monograph review: Isakova L. D. The diversity of languages in their variability. Moscow: Prospekt, 2022. 176 p.	191
Fedulenkova T. N., Shekhovskaya Yu. A. Linguocreativity as a promising trend of diverse discourses. Book review: Linguistic creativity in discourses of different types: Limits and possibilities: a collective monograph; ed. by I. V. Zyкова. Moscow: R. Valent, 2021. 564 p.	197
For the authors attention	204

Вестник Череповецкого государственного университета. 2025. № 3 (126). С. 7–16.
Cherepovets State University Bulletin, 2025, no. 3 (126), pp. 7–16.

Научная статья
УДК 669.1+519.677
<https://doi.org/10.23859/1994-0637-2025-3-126-1>
<https://elibrary.ru/vphqsj>

Модель управления знаниями металлургического предприятия на основе эмбединговых моделей

Виктор Алексеевич Веденеев^{1✉}, Евгений Валентинович Ершов²,
Роман Георгиевич Ковыркин³

^{1, 2}Череповецкий государственный университет,

^{1, 3}АО «Северсталь менеджмент»,
Череповец, Россия,

^{1✉}Super.victor228@yandex.ru

²evershov@chsu.ru, <https://orcid.org/0000-0003-2888-4242>

³rgkovyrshin@severstal.com

Аннотация. В статье предложена модель интеграции и управления корпоративными данными на металлургическом предприятии, основанная на использовании современных технологий обработки естественного языка (NLP) и эмбединговых моделей. Модель позволяет эффективно работать с разнородными данными, поступающими из различных систем хранения корпоративной информации, таких как ERP, CRM, MES и др. Для обеспечения гибкости и масштабируемости применяется векторная база данных, которая обеспечивает семантический поиск и анализ данных с учетом контекста. Использование Large Language Models (LLM) и эмбединговых моделей позволяет преобразовывать текстовые и структурированные данные в векторные представления, что упрощает интеграцию и обработку информации. Разработанная модель поддерживает распределение прав доступа к данным, обеспечивая безопасность и соответствие корпоративным требованиям. Предложенное решение может быть использовано для повышения операционной эффективности, улучшения процессов принятия решений и усиления конкурентоспособности предприятия.

Ключевые слова: управление данными, интеграция корпоративных систем, металлургическое предприятие, векторная база данных, эмбединговые модели, Large Language Models (LLM), семантический поиск, распределение прав доступа

Для цитирования: Веденеев В. А., Ершов Е. В., Ковыркин Р. Г. Модель управления знаниями металлургического предприятия на основе эмбединговых моделей // Вестник Череповецкого государственного университета. 2025. № 3 (126). С. 7–16. <https://doi.org/10.23859/1994-0637-2025-3-126-1>; EDN: VPHQSJ

Metallurgical enterprise knowledge management model based on embedding models

Viktor A. Vedeneev^{1✉}, Evgeniy V. Ershov², Roman G. Kovyrshin³

^{1, 2}Cherepovets State University,

^{1, 3} JSC Severstal Management,

Cherepovets, Russia,

^{1✉}Super.victor228@yandex.ru

²ershov@chsu.ru, <https://orcid.org/0000-0003-2888-4242>

³rgkovyrshin@severstal.com

Abstract. The article presents a model for integrating and managing corporate data at a metallurgical enterprise based on modern natural language processing (NLP) technologies and embedding models. The model enables efficient handling of heterogeneous data from various corporate information storage systems, such as ERP, CRM, MES, and others. To ensure flexibility and scalability, a vector database is employed, facilitating semantic search and contextual data analysis. The use of Large Language Models (LLM) and embedding models allows textual and structured data to be transformed into vector representations, which simplifies data integration and processing. The developed model supports access control mechanisms, ensuring data security and compliance with corporate requirements. The proposed solution can be used to enhance operational efficiency, improve decision-making processes, and strengthen the competitiveness of the enterprise.

Keywords: data management, corporate system integration, metallurgical enterprise, vector database, embedding models, Large Language Models (LLM), semantic search, access control

For citation: Vedeneev V. A., Ershov E. V., Kovyrshin R. G. Metallurgical enterprise knowledge management model based on embedding models. *Cherepovets State University Bulletin*, 2025, no. 3 (126), pp. 7–16. (In Russ.) <https://doi.org/10.23859/1994-0637-2025-3-126-1>; EDN: VPHQSJ

Введение

В современной металлургической промышленности на различных уровнях формирования данных эффективное управление знаниями является ключевым фактором успеха при реализации инициатив и проектов. С точки зрения исследования и разработки новых видов продуктов и технологий упорядоченный, структурированный метод управления данными позволяет оптимизировать подход к принятию решений, высвободить время на решение нетиповых задач. С появлением LLM (Large Language Model) разработка, проектирование и внедрение информационных систем, позволяющих управлять уровнями знаний различных вариантов представления и форматов, становятся все более важными для развития отраслевых предприятий¹.

Задачами управления знаниями на металлургическом предприятии являются выявление, оценка, структурирование, хранение и распространение знаний и опыта сотрудников (явных и неявных знаний), а также обеспечение доступа к важной информации для принятия обоснованных решений. Помимо вышеперечисленных факторов, управление знаниями включает в себя разработку и реализацию механизмов,

¹ Габова Ю. А., Замбрицкая Е. С. Оптимизационное планирование на предприятии черной металлургии с использованием автоматизированных информационных систем // Корпоративная экономика. 2022. № 2 (30). С. 4–13. EDN: YPSSW

позволяющих сотрудникам делиться знаниями и опытом, а также использовать их в практической деятельности при реализации инициатив.

Стандартные подходы к формированию оптимальных процессов управления знаниями на металлургических предприятиях включают создание корпоративных баз знаний, реализацию систем управления знаниями и организацию обучения сотрудников¹. Эти подходы позволяют обеспечить доступ к необходимой информации, развивать навыки и компетенции сотрудников, а также повысить эффективность производственного процесса.

Однако эти подходы не всегда эффективны в условиях металлургической отрасли, особенно, если источники формирования данных постоянно меняются и насыщаются неструктурированным набором данных различных форматов. Из этого следует, что проектирование и создание новых подходов к обработке информации в системе управления знаниями, которые будут учитывать специфику металлургического производства и обеспечивать доступ к необходимой информации в реальном времени, является актуальной задачей.

Разработка модели управления информацией в корпоративной среде с применением методов и алгоритмов обработки неструктурированных наборов данных, позволяющей управлять различным уровнем знаний на металлургическом предприятии, является одним из возможных решений этой задачи. Такая модель должна обеспечивать доступ к необходимой информации, переводить и актуализировать знания и опыт сотрудников, а также обеспечивать эффективное использование знаний в практической деятельности. В данной статье рассмотрен вариант реализации указанной модели и формирующих ее алгоритмов.

Основная часть

Модель управления информацией в корпоративной среде используется на всех этапах работы с данными: начиная с момента их занесения в одну из множества информационно-систем и заканчивая поиском информации и получением сводного отчета по всем подсистемам. Данные могут храниться в реляционных, документо-ориентированных или объектно-ориентированных базах данных с учетом распределения прав доступа для каждой записи. Таким образом, в качестве входных данных модели управления информацией будут служить все существующие метаданные систем стандартных типов, обработанных с помощью эмбединговой модели ruBERT:

- целые числа;
- вещественные числа;
- символы;
- строки;
- логические значения;
- дата и время;
- массивы;
- коллекции.

ruBERT – модель, которая преобразует текст в векторы, чтобы впоследствии их можно было обрабатывать и анализировать. Единицами работы с данными в эмбединговых моделях являются чанки – фрагменты текста, которые преобразуются в

векторное представление. Чанки могут быть использованы для улучшения эффективности и качества эмбединговых моделей, поскольку они позволяют модели уделить больше внимания конкретным частям текста¹.

Задача модели заключается в формировании единого индекса источника данных, распределенных в векторном формате с распределением прав доступа к каждой единице данных всех подсистем неструктурированного формата (ролевая модель).

Связь систем с точки зрения обмена информацией в корпоративной среде может быть представлена в виде графа, где вершины представляют собой различные подсистемы (информационные системы), а ребра – взаимодействия между ними с единицей записи. Модель может быть описана следующим образом:

$$G = (V, E),$$

где V – множество вершин (систем управления знаниями); E – множество ребер (взаимодействий между системами).

Каждая вершина в графе может быть описана вектором, представляющим собой особенности системы управления знаниями. Например, подсистема интеллектуальной собственности может быть описана вектором, включающим информацию о патентах, авторских правах и других правах интеллектуальной собственности, а подсистема управления сообществами может включать список экспертов в сообществе, уровень компетенций и реестр решенных проблем. Пример представления такого графа, построенного на нескольких подсистемах, приведен на рис. 1.

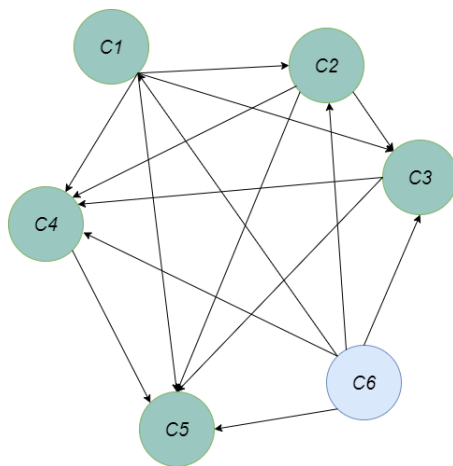


Рис. 1. Пример построения графа связей между подсистемами обмена информацией в корпоративной среде металлургической компании

¹ Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space // Journal of Artificial Intelligence Research. 2013. No. 48. P. 315.

Характеристика параметров подсистем графа указана в табл. 1.

Таблица 1

Характеристика графа связей между подсистемами обмена информацией
в корпоративной среде металлургической компании

Количество вершин				6		
Количество ребер (связей)				15		
Прямая связь сформированных вершин						
№ вер- шины	1	2	3	4	5	6
Номера связей вершин	E1: V1 – V2 E2: V1 – V3 E3: V1 – V4 E4: V1 – V5	E5: V2 – V3 E6: V2 – V4 E7: V2 – V5	E8: V3 – V4 E9: V3 – V5	E10: V4 – V5	–	E11: V6 – V1 E12: V6 – V2 E13: V6 – V3 E14: V6 – V4 E15: V6 – V5

Для объединения данных из различных систем управления знаниями использует-ся эмбединговая модель, которая преобразует векторы подсистем хранения инфор-мации в общую векторную базу данных на основе чанков. Эмбединговая модель может быть представлена в виде следующего уравнения:

$$V_i = f(W_i, X_i),$$

где V_i – вектор, представляющий систему управления знаниями i ; W_i – матрица ве-сов; X_i – вектор особенностей системы управления знаниями i ; f – функция актива-ции.

Перед использованием эмбединговая модель была обучена на корпусе научно-технических данных из реестра ЕГИСУ НИОКТР¹ для обогащения знаниями и уве-личения шанса на то, что терминология в заданном пользователем запросе уже будет знакома дообученной модели.

Дообучение модели состояло из следующих этапов:

1. Определение оптимального количества кластеров с помощью метода локтя.
2. Подтверждение результата с помощью метода силуэтного коэффициента.
3. Выбор оптимального количества кластеров.
4. Группировка данных по темам с помощью алгоритма кластеризации K-means.
5. Выбор самой популярной темы (кластера) и выборка текстов из этого класте-ра.
6. Перемешивание текстов кластера и выбор первой тысячи текстов для предоб-работки и дообучения модели.

В первую очередь было определено оптимальное количество кластеров, на кото-рое необходимо группировать данные. Для этого был использован метод локтя (Elbow Method), в рамках которого была подсчитана сумма квадратов расстояний

¹ Домен «Наука и инновации». URL: <https://gisnauka.ru/> (дата обращения: 08.11.2024).

внутри кластера WSS (Within-Cluster Sum of Squares). На рис. 2 можно увидеть, что при уменьшении WSS точки более плотно собираются внутри кластера, т. е. более качественно распределяются по кластерам, количество самих кластеров при этом увеличивается. Также по рис. 2 можно определить, что примерное оптимальное количество кластеров будет равно 10.

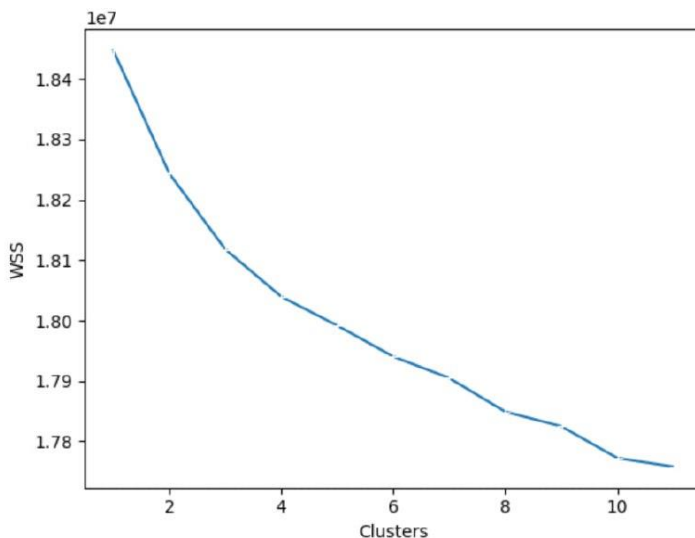


Рис. 2. Распределение данных по кластерам с использованием метода WSS (Within-Cluster Sum of Squares)

Для группировки данных по темам и выявления самой популярной темы используется алгоритм кластеризации K-means. Данный алгоритм работает путем инициализации центров (или центроидов) каждого кластера, затем каждой точке данных присваивается ближайший центр, пока не достигается заданное нами количество кластеров. На рис. 3 показано распределение данных (100 000 аннотаций к диссертациям) на 8 кластеров.

Далее было подсчитано, какая метка label чаще всего встречается в данных, и также мы определили, что label = 5. Таким образом, был получен самый популярный кластер, собравший в себе 24730 аннотаций. Тексты кластера были дополнительно перемешаны. Помимо этого были взяты первые десять тысяч текстов в качестве выборки для последующей предобработки и дообучения модели.

После преобразования данных и их обработки с помощью эмбединговой модели в векторной базе данных можно проводить полнотекстовый поиск информации по любой из систем управления знаниями с учетом распределения прав доступа для каждой записи. Для обобщения результатов поиска можно использовать любую подключаемую LLM-модель, которая может быть развернута локально, либо взаимодействовать по API с помощью облачных решений.

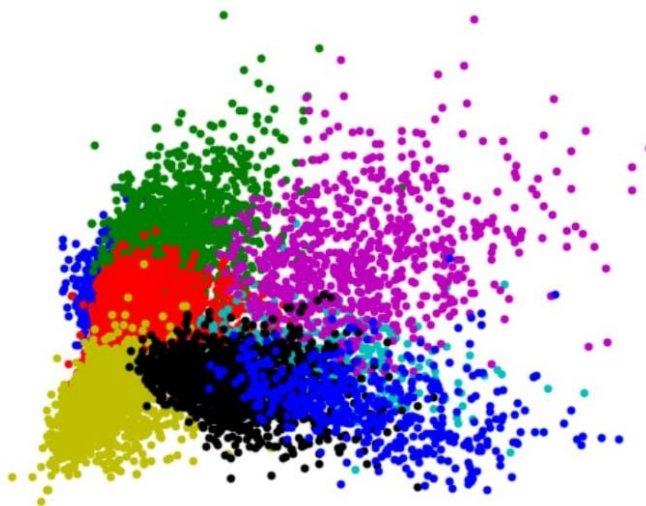


Рис. 3. Группировка данных по темам с помощью K-means

Алгоритм работы модели управления знаниями на металлургическом предприятии:

1. Получение данных из существующих подсистем управления знаниями.
2. Построение графа G-взаимодействий метаданных подсистем с учетом распределения прав доступа к каждой записи и их связей из подсистем.
3. Преобразование данных в векторную базу данных с помощью эмбединговой модели.
4. Объединение данных из различных подсистем управления знаниями в одной векторной базе данных с созданием и сохранением отдельных полей метаданных, отвечающих за ролевую модель доступа к записям.
5. Проведение полнотекстового поиска информации по данным из любой подсистемы управления знаниями с помощью подключаемой LLM-модели для суммаризации настраиваемого количества результатов.
6. Вывод результатов поиска в виде списка документов, содержащих необходимую информацию, а также суммаризация посредством LLM-модели.

Для проверки работоспособности предложенной модели был проведен качественный и количественный эксперимент. За основу входных данных для проверки использовались запросы пользователей, основанные на методе экспертной оценки. Это означает, что эксперты сравнивают результаты поиска с идеальными или ожидаемыми результатами и оценивают их соответствие. Также для оценки работы модели были рассчитаны процент правильности предоставляемого объема данных пользователю относительно роли в системах и среднее время выполнения поискового запроса, результаты которого приведены в табл. 2.

Таблица 2

**Результаты работы модели обработки информации
из подсистем управления знаниями на металлургическом предприятии**

Запрос	Количество элементов графа <i>G</i> подсистем	Релевантность (score)	% правильности данных	Среднее время вы- полнения запроса, с
Что такое гальвани- ческая коррозия?	8	'score': 0.804077041 'score': 0.778830540	95 %	13,1 с
Влияние pH на корро- зию металлических поверхностей	8	'score': 0.984344291 'score': 0.878566336	100 %	9,3 с
Что такое гальвани- ческая коррозия?	6	'score': 0.837996459 'score': 0.805738723	95 %	7,8 с
Влияние pH на корро- зию металлических поверхностей	6	'score': 0.703968524 'score': 0.699626922	90 %	7,2 с
Какие свойства инерт- ного газа были изуче- ны в ходе охлаждения?	4	'score': 0.960688006 'score': 0.841825640	100 %	6,6 с
Какие свойства компо- зиционных материалов были исследованы?	4	'score': 0.94017686 'score': 0.82636265	98 %	6,4 с
Какие свойства инерт- ного газа были изуче- ны в ходе охлаждения?	2	'score': 0.81944013 'score': 0.70942751	95 %	5,1 с
Какие свойства компо- зиционных материалов были исследованы?	2	'score': 0.819768726 'score': 0.636752722	90 %	6,4 с

Выводы

Разработанная модель управления знаниями на металлургическом предприятии позволяет оперировать различным уровнем знаний, включая интеллектуальную собственность, знания сообществ, партнерские отношения, экспертную сеть и систему менторинга и коучинга. Использование векторной базы данных и эмбединговой модели позволяет объединить данные из различных систем управления знаниями и проводить полнотекстовый поиск информации по любой из систем. Разработанная модель может быть использована для повышения эффективности и конкурентоспособности металлургического предприятия.

Список литературы / References

Виноградова Е. Ю., Галимова А. И. Принципы формирования корпоративной информационной системы для внедрения на российских предприятиях. *Известия Уральского государ-*

ственного экономического университета, 2017, № 2 (70), с. 111–123.
<https://doi.org/10.29141/2073-1019-2017-14-2-10>; EDN: YSSFMR

Vinogradova E. Iu., Galimova A. I. The principles of building a corporate information system for implementation at Russian enterprises. *Journal of New Economy*, 2017, no. 2 (70), pp. 111–123. (In Russ.). <https://doi.org/10.29141/2073-1019-2017-14-2-10>; EDN: YSSFMR

Габова Ю. А., Замб르지цкая Е. С. Оптимизационное планирование на предприятии черной металлургии с использованием автоматизированных информационных систем. *Корпоративная экономика*, 2022, № 2 (30), с. 4–13. EDN: YPLSSW

Gabova Iu. A., Zambrzhitskaia E. S. Optimization planning at a ferrous metallurgy enterprise using automated information systems. *Corporate economics*, 2022, no. 2 (30), pp. 4–13. (In Russ.) EDN: YPLSSW

Домен «Наука и инновации». URL: <https://gisnauka.ru/> (дата обращения: 08.11.2024).

Domain “Science and Innovations”. (In Russ.) Available at: <https://gisnauka.ru/> (accessed: 08.11.2024).

Дорожков Н. Д., Купчинская Ю. А., Юдалевич Н. В. Корпоративные информационные системы: проблемы, тенденции и перспективы развития. *Бизнес-образование в экономике знаний*, 2016, № 1 (3), с. 26–30. EDN: TIEDEG

Dorozhkov N. D., Kupchinskaja Iu. A., Iudalevich N. V. Enterprise resource planning systems: problems, tendencies and perspectives for development. *Business education in the knowledge-based economy*, 2016, no. 1 (3), pp. 26–30. (In Russ.) EDN: TIEDEG

Заложнев А. Ю., Заложнева Л. Л., Чистов Д. В., Шуремов Е. Л. Эволюция и принципы построения информационных систем управления предприятием. *Программные продукты и системы*, 2014, № 2 (106), с. 34–38. EDN: TPOUUV

Zalozhnev A. Iu., Zalozhneva L. L., Chistov D. V., Shuremov E. L. Evolution and information systems design principles for enterprise management. *Software & Systems*, 2014, no. 2 (106), pp. 34–38. (In Russ.) EDN: TPOUUV

Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. *Journal of Artificial Intelligence Research*, 2013, no. 48, pp. 311–323.

Сведения об авторах

Виктор Алексеевич Веденеев – аспирант; Super.victor228@yandex.ru, Череповецкий государственный университет (д. 5, пр-т Луначарского, 162600 Череповец, Россия); **Viktor A. Vedeneev** – Postgraduate Student, Super.victor228@yandex.ru, Cherepovets State University (5, pr. Lunacharskogo, 162600 Cherepovets, Russia).

Евгений Валентинович Ершов – доктор технических наук, профессор; <https://orcid.org/0000-0003-2888-4242>, evershov@chsu.ru, Череповецкий государственный университет (д. 5, пр-т Луначарского, 162600 Череповец, Россия); **Evgeniy V. Ershov** – Doctor of Technical Sciences, Professor, <https://orcid.org/0000-0003-2888-4242>, evershov@chsu.ru, Cherepovets State University (5, pr. Lunacharskogo, 162600 Cherepovets, Russia).

Роман Георгиевич Ковыршин – ведущий эксперт (куратор); rgkovyrshin@severstal.com, АО «Северсталь менеджмент»; **Roman G. Kovyrrshin** – Leading Expert (advisor), rgkovyrshin@severstal.com, JSC Severstal Management.