

АВТОМАТИЗАЦИЯ ЗАКУПОК: ИННОВАЦИОННЫЕ ТЕХНОЛОГИИ В УПРАВЛЕНИИ ЗНАНИЯМИ

PROCUREMENT AUTOMATION: INNOVATIVE TECHNOLOGIES IN KNOWLEDGE MANAGEMENT

**A. Gorbunov
M. Kuznetsova**

Summary. This article examines an urgent problem related to the development and use of an automatic knowledge base in the field of modern procurement. First of all, the importance of creating such a knowledge base and its application in the context of procurement processes is considered. The scope of the knowledge base is highlighted, including both customers and suppliers, which is a key aspect for ensuring the efficiency and transparency of procurement processes. Special attention is paid to the study of the basic legislation governing procurement processes, including Federal Laws No. 44-FZ and 223-FZ, as well as other legal acts essential for this area. This is important to understand the context in which the automatic knowledge base will be used, as well as to ensure that it meets all legal requirements. In addition, the article provides a detailed overview of the use of Large Language Model (LLM) technology with information vectorization based on semantic analysis. This technique allows you to analyze and vectorize information to create a knowledge base, which ensures effective search and processing of data in the context of procurement.

Keywords: automatic knowledge base, procurement, legislation, Federal Laws No. 44-FZ and 223-FZ, document management, machine learning, natural language processing, Large Language Model (LLM), information vectorization, semantic analysis, knowledge management, embedding vectors, semantic distance, cosine distance, information systematization.

Горбунов Александр Николаевич

Аспирант, Аккредитованное образовательное
частное учреждение высшего образования «Московский
финансово-юридический университет МФЮА»;
старший разработчик, ООО «Клик Групп»
a.gorbunov@click-group.ru

Кузнецова Марина Николаевна

доктор экономических наук, профессор, «Московский
финансово-юридический университет МФЮА»
marina_kuzn82@mail.ru

Аннотация. В данной статье исследуется актуальная проблема, связанная с разработкой и использованием автоматической базы знаний в сфере современных закупок. В первую очередь, рассматривается важность создания такой базы знаний и ее применение в контексте процессов закупок. Освещается область применения базы знаний, включая как заказчиков, так и поставщиков, что является ключевым аспектом для обеспечения эффективности и прозрачности процессов закупок. Особое внимание уделяется изучению основного законодательства, регулирующего процессы закупок, включая Федеральные законы №44-ФЗ и 223-ФЗ, а также другие правовые акты, существенные для данной сферы. Это важно для понимания контекста, в котором будет использоваться автоматическая база знаний, а также для обеспечения ее соответствия всем требованиям законодательства. Кроме того, в статье представлен подробный обзор использования технологии Large Language Model (LLM) с векторизацией информации на основе семантического анализа. Эта методика позволяет анализировать и векторизовать информацию для создания базы знаний, что обеспечивает эффективный поиск и обработку данных в контексте закупок.

Ключевые слова: автоматическая база знаний, закупки, законодательство, Федеральные законы №44-ФЗ и 223-ФЗ, документооборот, машинное обучение, обработка естественного языка, Large Language Model (LLM), векторизация информации, семантический анализ, управление знаниями, эмбединг векторов, семантическое расстояние, косинусное расстояние, систематизация информации.

В современном мире документооборот в организациях является неотъемлемой частью ее функционирования, представляя собой непрерывный поток документов [1], проходящих определенные маршруты. Особенно важным аспектом является соответствие этого потока требованиям законодательства, таким как Федеральные законы № 44-ФЗ и № 223-ФЗ, а также другим правовым актам. Однако, в условиях постоянных изменений и обновлений законодательства, а также большого объема информации, актуализация знаний в этой области может замедляться или быть неполной. Возникает причинно-следственные проблемы: противоречивость ряда статей законодательства, расхожесть правоприме-

нительной практики [2]. В связи с этим возникает необходимость в использовании автоматической базы знаний закупок, которая не только упростит процесс обработки документов, но и обеспечит надежное соблюдение всех требований законодательства. В данной статье мы обсудим ключевые аспекты автоматической базы знаний закупок и ее важность для современных организаций.

В настоящее время регламентация представления знаний в сфере закупок в России осуществляется через законодательные акты, нормативные документы и методические рекомендации, а также внутренние документы организации и документацию о закупочных процедурах. Эти документы обладают ссылочной структурой, обе-

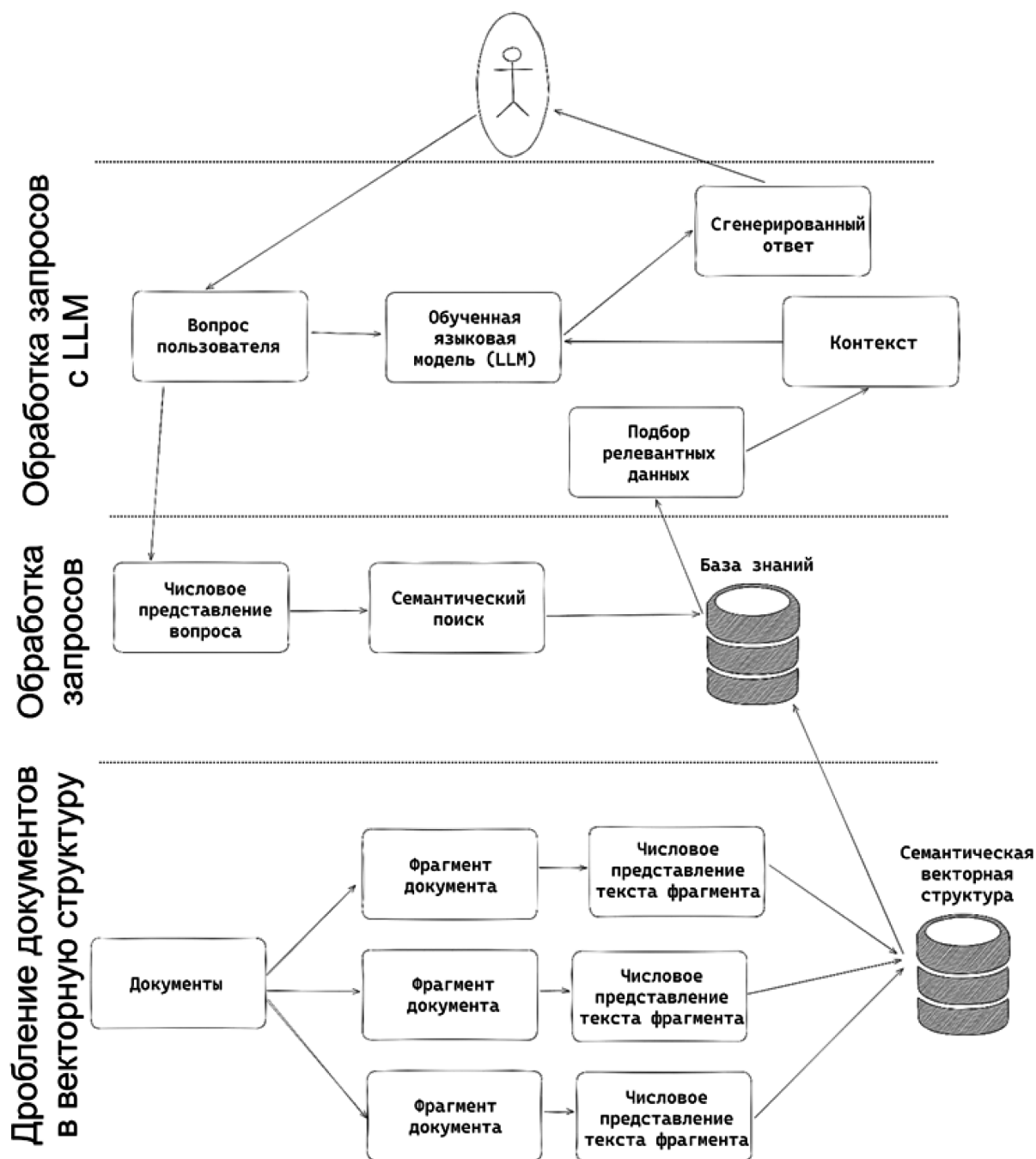


Рис. 1. Графическое представление базы знаний по закупкам

спечивая пользователям доступ к необходимой информации при необходимости. Однако, из-за множества потоков этой информации, создание и поддержание целостной базы знаний с таким подходом может оказаться непрактичным. Разнообразие и динамичность законодательства, а также специфика требований организаций, делают сложным обеспечение актуальности и полноты информации в такой базе данных. Вместо этого, эффективное управление знаниями требует систематизации и организации информации в соответствии с потребностями и стратегией конкретной организации.

Для реализации автоматической базы знаний в области закупок был выбран подход, основанный на комбинации методов машинного обучения и обработки естественного языка (Рисунок 1). В частности, использовалась технология LLM, которая позволяет анализировать и векторизировать информацию на основе семантического анализа.

Все документы законов, нормативных актов, а также внутренних правил и положений организации представлены в виде векторизированной базы, разделенной

на фрагменты, и приведены к числовому представлению. Это позволяет создать векторную структуру, которая формирует базу знаний, где каждый фрагмент соответствует определенному правилу, норме или процедуре. Поступающий запрос на получение информации преобразуется в числовое представление с использованием LLM для возможности поиска по базе знаний. Этот процесс включает в себя анализ векторных представлений запроса и их сопоставление с векторными представлениями фрагментов базы знаний.

Запросы обрабатываются через LLM для поиска по базе знаний. Для вычисления семантического расстояния между запросом и фрагментом базы знаний используется формула[3]:

$$\text{SemDist}(e_{ref}, e_{hyp}) = 1 - \frac{(e_{ref})^T \cdot e_{hyp}}{\|e_{ref}\| \cdot \|e_{hyp}\|},$$

где e_{ref} — эмбединг векторного представления эталонного текста e_{hyp} — эмбединг векторного представления гипотетического текста, $\|e_{ref}\|$ и $\|e_{hyp}\|$ — обозначают их нормы.

Косинусоидальное расстояние учитывает только угол между двумя вставками в предложении, а не их величины. Значение SemDist ограничено значением от 0 до 1, где более низкие значения указывают на более высокое семантическое сходство, и наоборот.

На основе введенных коэффициентов отбора ответов из базы мы получаем контекст, который уже обработанный через языковую модель, выдает ответ пользователю. Релевантность используемых источников определяется по следующей формуле:

$$\text{relevance}(q, d) = \sum_{i=1}^n w_i \cdot \text{feature}_i(q, d)$$

где w_i — весовые коэффициенты, а $\text{feature}_i(q, d)$ — характеристика i -го признака между запросом q и фрагментом d , которая оценивает косинусное сходство. Чем выше это полученное значение, тем более релевантен фрагмент. Весовыми коэффициентами в данном подходе возможны признаки:

- Тематическое соответствие: оценка, насколько содержание документа соответствует тематике запроса пользователя, например, законодательство о закупках.
- Актуальность: оценка, насколько документ актуален на момент запроса, учитывая последние изменения в законодательстве.
- Полнота информации: оценка, насколько документ содержит всю необходимую информацию по запросу пользователя.
- Принадлежность: признак, который указывает, написан ли документ со стороны поставщика или заказчика.

Этот признак может принимать два значения: «поставщик» или «заказчик».

Для наглядности сравнение двух документов может быть представлено следующим образом: Федеральный закон №44-ФЗ имеет высокое тематическое соответствие и актуальность, имеет большую упоминаемость действий со стороны заказчика и среднюю полноту информации, Постановление Правительства РФ, в свою очередь, имеет высокое тематическое соответствие и полноту информации, среднюю актуальность и больше упоминает действия со стороны заказчика относительно поискового запроса пользователя. Тогда релевантность может быть оценена по следующей формуле:

$$\text{relevance}(q, d) = w_1 \cdot \text{feature}_1(q, d) + w_2 \cdot \text{feature}_2(q, d) + w_3 \cdot \text{feature}_3(q, d) + w_4 \cdot \text{feature}_4(q, d)$$

Для того, чтобы отразить приоритеты пользователя, предположим, что веса $w_1 = 0.5$, $w_2 = 0.3$, $w_3 = 0.2$ и $w_4 = 0.4$. Теперь оценка релевантности каждого документа с учетом этих признаков равна:

$$\begin{aligned} \text{relevance}(q, \text{«Федеральный закон №44 - ФЗ»}) \\ = 0.5 \cdot 1 + 0.3 \cdot 1 + 0.2 \cdot 0.5 + 0.4 \cdot 0 = 0.7 \end{aligned}$$

$$\begin{aligned} \text{relevance}(q, \text{«Постановление Правительства РФ»}) \\ = 0.5 \cdot 1 + 0.3 \cdot 0.5 + 0.2 \cdot 1 + 0.4 \cdot 0 = 0.8 \end{aligned}$$

Таким образом, документ Постановления Правительства РФ будет оценен как более релевантный для запроса пользователя, так как он соответствует всем критериям запроса, поэтому его содержание попадет в контекст с более высоким приоритетом для выдачи ответа пользователю.

Описанный подход имеет свои преимущества, включая возможность учета различных признаков для оценки релевантности документов. Однако у него есть несколько минусов. Один из основных недостатков состоит в неоднозначности весовых коэффициентов. Определение оптимальных весов для каждого признака может быть сложной задачей, требующей экспертного опыта или многочисленных экспериментов[4]. Кроме того, оценка релевантности документов может быть субъективной и зависеть от предпочтений или предвзятости пользователя. Это может привести к разным оценкам релевантности.

Вывод

Использование LLM с векторизацией информации на основе семантического анализа открывает новые перспективы в области информационного обеспечения закупочных процессов. Это позволяет создавать интеллектуальные системы, способные эффективно анализировать и обрабатывать огромные объемы данных, что в свою очередь повышает производительность и качество принимаемых решений в сфере закупок.

ЛИТЕРАТУРА

1. Моряков Д.А. Моделирование и автоматизация процессов закупочной деятельности организации. Вестник науки. 2019; № 6; 395 с.
2. Бирюкова Е. А. Система госзакупок в Российской Федерации: проблемы, нарушения, изменения в законодательстве. Научные записки молодых исследователей. 2022;10(1):44 с.
3. Kim S., Arora A., Le D., Yeh C.-F., Fuegen C., Kalinli O., Seltzer M.L. Semantic Distance: A New Metric for ASR Performance Analysis Towards Spoken Language. Posts and Telecommunications Institute of Technology. 2021,2 p.
4. В.С. Тарасян. Математическое моделирование в логистике: учебно-методическое пособие. Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Уральский Государственный Университет Путей Сообщения». 2013, 4 с.

© Горбунов Александр Николаевич (a.gorbunov@click-group.ru); Кузнецова Марина Николаевна (marina_kuzn82@mail.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»