

**Разработка интеллектуальной вопросно-ответной системы основанной на библиотеке стандартов научно-исследовательского центра**

**Волгин Л.А. (ИТМО).**

**Научный руководитель: доцент Фёдоров Д.А.  
(ИТМО)**

**Введение.** Крыловский государственный научный центр является исследовательским институтом в области судостроения и морской техники, где хранится обширный массив специализированной информации, включая проектную документацию и библиотеку стандартов. В настоящее время поиск информации с целью получения регламентирующих документов, необходимых для принятия решений сотрудниками в различных ситуациях, осуществляется вручную. Такой подход требует значительных временных затрат и может приводить к ошибкам. Внедрение поисковой системы на основе технологии RAG позволит значительно упростить этот процесс и повысить эффективность доступа к данным.

**Основная часть.** В рамках исследования был проведен анализ существующих коммерческих решений для интеллектуального поиска. Хотя подобные системы предлагают определенные преимущества, они также имеют значительные недостатки: высокая стоимость внедрения и поддержки, зависимость от облачных сервисов и высокие требования к техническим ресурсам. Эти недостатки делают их нецелесообразными для использования в условиях КГНЦ.

Было также рассмотрено несколько open-source RAG-решений для интеллектуального поиска, а также облачные AI платформы. Эти системы предоставляют возможности для создания кастомных решений, однако они требуют значительных усилий для адаптации, в то время как их универсальный функционал может оказаться избыточным для специфических задач, которые стоят перед организацией.

Разработка собственной RAG-системы представляет собой значительные преимущества, так как она адаптирована к конкретным требованиям и ограничениям КГНЦ, что делает её более целенаправленным и эффективным решением. Данная система позволяет быстро изменять функционал под новые задачи и требования без необходимости дожидаться обновлений от внешних поставщиков. Кроме того, она поддерживает локальное хранение данных, что позволяет ей работать в условиях закрытой сети и обеспечивает безопасность и доступность информации без зависимости от внешних облачных хранилищ.

Предлагаемая RAG-система осуществляет поиск релевантного фрагмента документации на основе запроса пользователя и предоставляет его в удобном для восприятия формате. При этом система использует запрос для генерации сжатого и четкого ответа, что позволяет пользователю быстро получать необходимую информацию. Для оценки эффективности работы системы пользователям предлагается оценивать качество генерируемых ответов.

Данная система упрощает доступ к информации и способствует повышению качества научной работы, сокращая время на поиск данных.

**Выводы.** В результате исследования была разработана RAG-система. Проведенные эксперименты подтвердили возможность эффективного применения

данной системы и в дальнейшем она может быть внедрена в рабочие процессы КГНЦ для повышения эффективности доступа к информации.

**Список использованных источников:**

1. Yuxiang Zheng, Shichao Sun, Lin Qiu, Dongyu Ru, Cheng Jiayang, Xuefeng Li, Jifan Lin, Binjie Wang, Yun Luo, Renjie Pan, Yang Xu, Qingkai Min, Zizhao Zhang, Yiwen Wang, Wenjie Li, Pengfei Li. *OpenResearcher: Unleashing AI for Accelerated Scientific Research* <https://arxiv.org/pdf/2408.06941v2>

2. Nir Diamant *Advanced RAG Techniques: Elevating Your Retrieval-Augmented Generation Systems* [https://github.com/NirDiamant/RAG\\_Techniques/blob/main/all\\_rag\\_techniques/reliable\\_rag.ipynb](https://github.com/NirDiamant/RAG_Techniques/blob/main/all_rag_techniques/reliable_rag.ipynb)

3. Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, Haofen Wang *Retrieval-Augmented Generation for Large Language Models: A Survey* <https://paperswithcode.com/paper/retrieval-augmented-generation-for-large>

4. Mohammad Reza Rezaei, Maziar Hafezi, Amit Satpathy, Lovell Hodge, Ebrahim Pourjafari *AT-RAG: An Adaptive RAG Model Enhancing Query Efficiency with Topic Filtering and Iterative Reasoning* <https://paperswithcode.com/paper/at-rag-an-adaptive-rag-model-enhancing-query>

Автор \_\_\_\_\_ Волгин Л.А.

Научный руководитель \_\_\_\_\_ Фёдоров Д.А.