

# Подсистема пополнения базы знаний интеллектуальной системы информационной поддержки создания и развития перспективных бионических технологий

В.В. Баранюк, О.С. Смирнова, А.М. Володина, М.Ю. Блинов

**Аннотация** – В статье рассмотрены вопросы классификации информационных ресурсов при пополнении базы знаний. Представлены результаты моделирования процесса работы с модулем добавления информационных ресурсов и модулем автоматизированного пополнения базы знаний интеллектуальной системы информационной поддержки процессов создания и развития перспективных бионических технологий. Моделирование процессов функционирования модулей отражено с помощью диаграмм деятельности и диаграмм последовательности языка UML.

**Ключевые слова** – классификация информационных ресурсов, автоматическая классификация, пополнение базы знаний, мониторинг открытых источников, модельное представление, бионические технологии.

В последнее время среди направлений исследований, охватывающих вопросы повышения качества поиска и работы с информационными ресурсами, особо выделяется направление, связанное с автоматизацией интеллектуальной деятельности пользователя, которое обычно реализуется при создании интеллектуальных систем.

Важным компонентом интеллектуальной системы является база знаний.

База знаний (БЗ) – это формализованная система сведений о некоторой предметной области, содержащая данные о свойствах объектов, закономерностях процессов и правила использования в задаваемых ситуациях этих данных для принятия новых решений [1].

Статья получена 14.09.2016 г.

Исследование выполнено федеральным государственным бюджетным образовательным учреждением высшего образования «Московский технологический университет» (МИРЭА) за счет гранта Российского научного фонда (проект №14-11-00854).

К.т.н., с.н.с. В.В. Баранюк, МИРЭА (e-mail: valentina\_bar@mail.ru).

О.С. Смирнова, МИРЭА (e-mail: mail.olga.smirnova@yandex.ru).

А.М. Володина, МИРЭА (e-mail: a.kholopova@mirea.ru).

М.Ю. Блинов, МИРЭА (e-mail: xxmaxxx2009@gmail.com).

В интеллектуальной системе информационной поддержки процессов создания и развития перспективных бионических технологий сосредоточены знания в области бионики и бионических технологий. Наличие в базе этих знаний способствует решению различных задач и появлению новых идей по созданию перспективных технологий, аналогичных реализованным в природе.

В публикациях [2, 3, 4] рассматривались различные вопросы, связанные с построением базы знаний в составе интеллектуальной системы информационной поддержки процессов создания и развития перспективных бионических технологий.

При разработке интеллектуальной системы [5, 6] планируется в рамках подсистемы пополнения базы знаний реализация возможности пополнения базы знаний информационными ресурсами, а также возможности её автоматизированного пополнения из открытых источников.

В вопросе пополнения БЗ новыми информационными ресурсами важным аспектом является их классификация (категоризация, рубрикация) – определение ресурса в одну или несколько тематик (категорий, рубрик) на основании его содержания. Для автоматизированного пополнения БЗ [5] соответственно необходимо использование автоматической классификации информационных ресурсов на основе заранее заданной схемы классификации и уже имеющегося множества классифицированных документов. При решении задач автоматической классификации текстовых документов используются методы информационного поиска (Information Retrieval, IR) [7, 8, 9] и машинного обучения (Machine Learning, ML) [10, 11].

Кроме этого, прорабатывается подход, основанный на построении онтологической модели в сфере бионики [12, 13, 14]. Его применение позволяет получить сравнительно простую для восприятия, но в тоже время гибкую модель классификации информационных ресурсов. Задача классификации

решается с применением специализированной базы знаний, основанной на онтологической модели.

Помимо классификации информационных ресурсов при пополнении базы знаний важным этапом также является формирование перечня открытых источников информации. В статье [15] предложен список источников с открытым доступом на русском и английском языках, планируемый к использованию при функционировании интеллектуальной системы информационной поддержки создания и развития перспективных бионических технологий.

С целью более детальной проработки процесса пополнения базы знаний проведено моделирование основного процесса. Для моделирования процесса работы с подсистемой пополнения БЗ выбран унифицированный язык моделирования UML (Unified Modeling Language), предназначенный для визуализации и документирования объектно-ориентированных систем и бизнес-процессов [16]. В процессе моделирования разработаны диаграммы деятельности и диаграммы последовательности, описывающие работу модуля для самостоятельного добавления информационных ресурсов, а также модуля автоматизированного пополнения базы знаний указанной подсистемы.

Диаграмма деятельности детализирует особенности реализации операций, выполняемых подсистемой, отражая динамику подсистемы и показывая схемы потоков управления в подсистеме от действия к действию, а также параллельные действия и альтернативные потоки.

Диаграмма последовательности отображает взаимодействие объектов, упорядоченное по времени. Основными компонентами диаграммы последовательности являются объекты, линии их жизни – вертикальные линии, отображающие течение времени, фокусы управления на линиях жизни, формируемые для демонстрации деятельности определенного объекта, а также сообщения между объектами, показываемые с помощью стрелок.

На рисунках 1 – 4 представлены диаграммы деятельности, а также диаграммы последовательности, отражающие процесс взаимодействия администратора с базой знаний через подсистему пополнения базы знаний, включающую в себя модуль добавления информационных ресурсов и модуль автоматизированного пополнения базы знаний.

Диаграмма последовательности модуля добавления ИР отражает взаимодействие между объектами во времени. Так после запуска системы администратором возможен цикл ввода и добавления информационного ресурса в базу знаний, отображенный с помощью операнда «loop». Операнд «alt» позволяет увидеть альтернативный фрагмент взаимодействия объектов, условие которого истинно.

Диаграмма последовательности модуля автоматизированного пополнения базы знаний отражает взаимодействие между объектами во времени. Осуществляя запуск проверки, администратор стартует автоматическую работу модуля, включающую в себя проверку наличия новых ИР во внешних ресурсах, обработку и сохранение данных в базе знаний. Операнд «alt» позволяет увидеть альтернативный фрагмент взаимодействия объектов, условие которого истинно, операнд «rag» указывает на фрагмент, где действия выполняются параллельно.

С помощью модельного представления наглядно продемонстрирована возможность и последовательность действий при мониторинге и загрузке из открытых источников информационных ресурсов, а также их последующей автоматизированной классификации.

Использование подсистемы пополнения базы знаний интеллектуальной системы информационной поддержки процессов создания и развития перспективных бионических технологий обеспечит накопление на различных языках актуальной информации о биологических объектах – бионических прототипах и об уже существующих бионических технологиях и их аналогах для информационной поддержки исследовательской деятельности

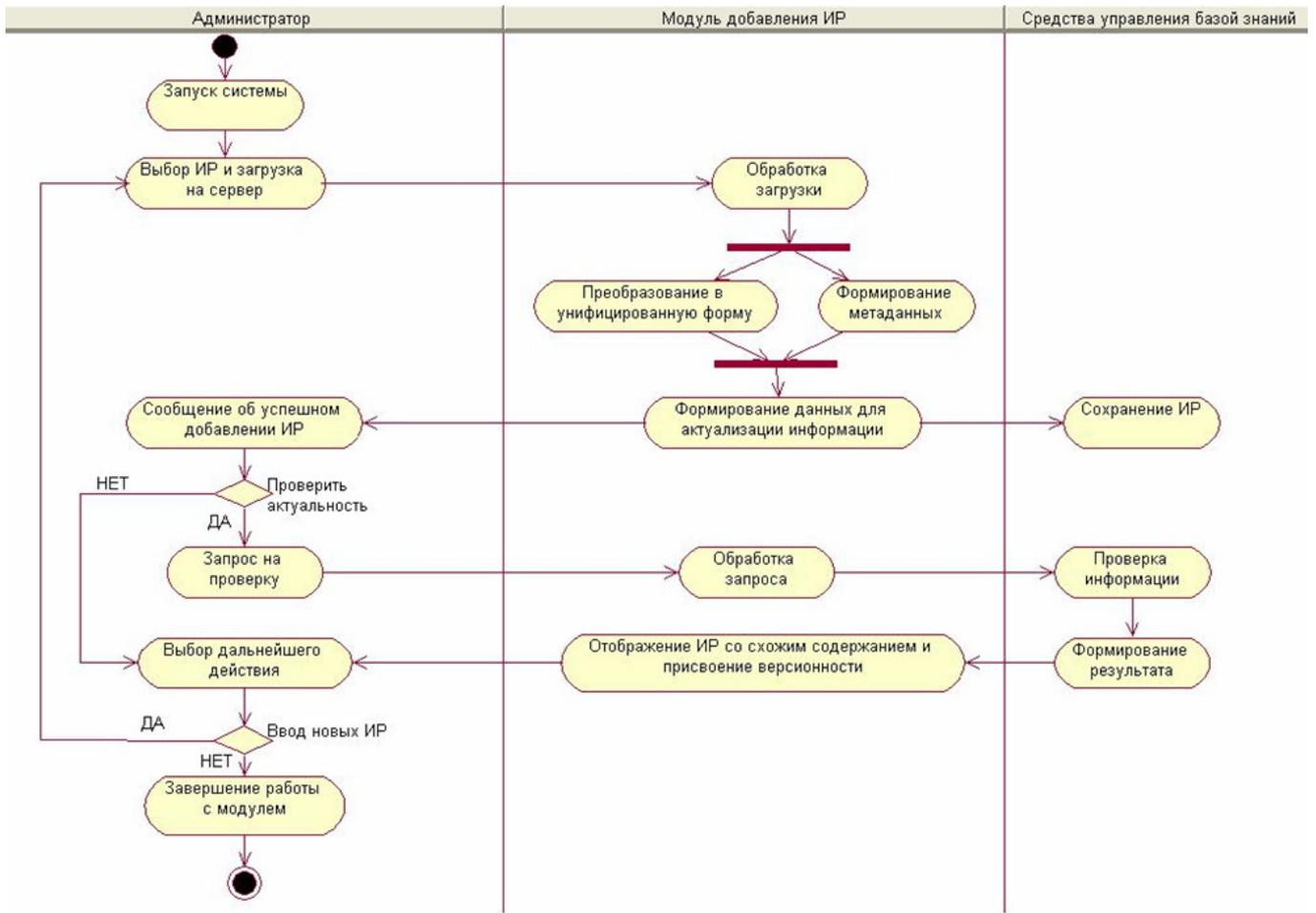


Рисунок 1 – Диаграмма деятельности, отображающая работу с модулем добавления IP

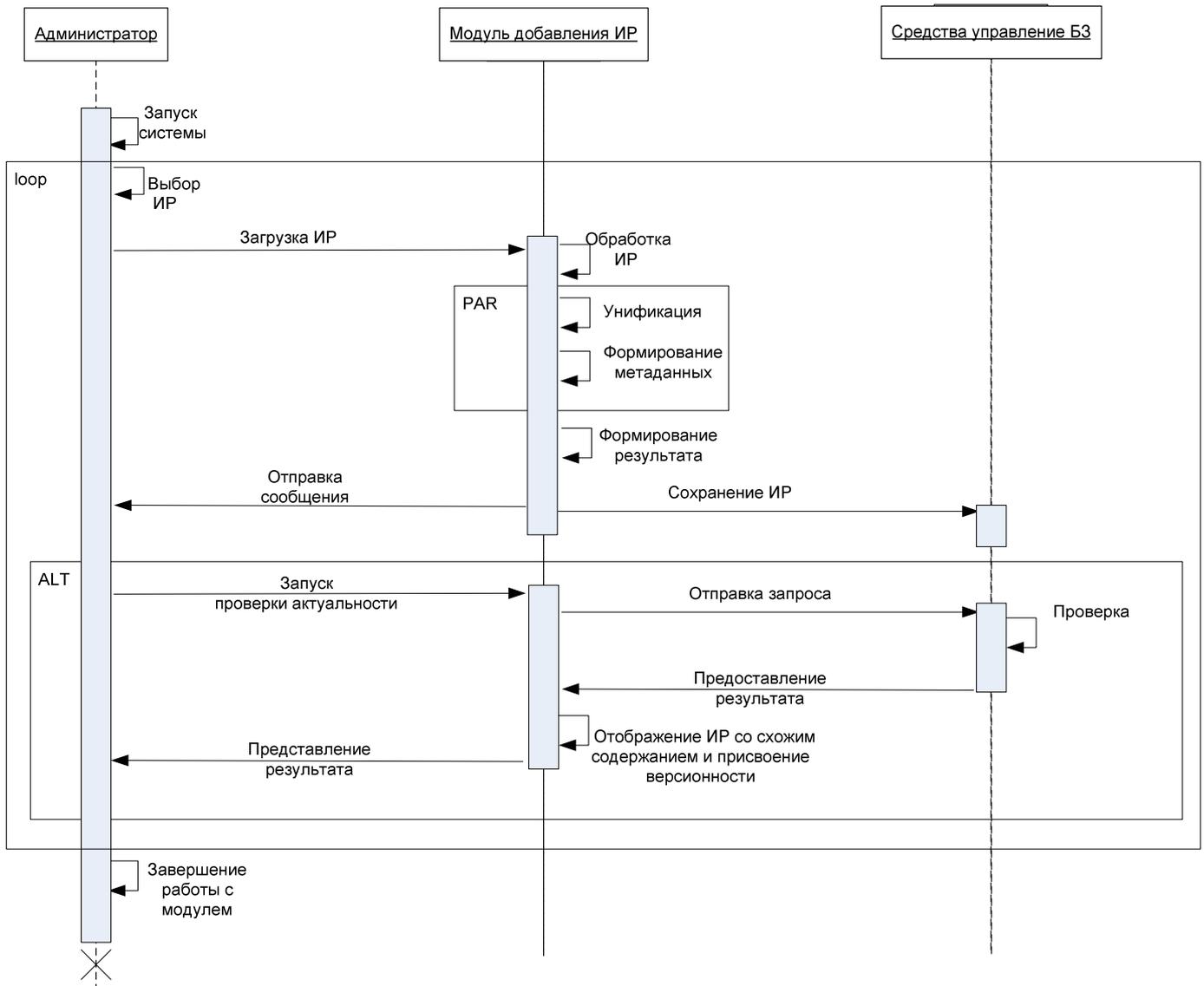


Рисунок 2 – Диаграмма последовательности, отображающая работу с модулем добавления IP

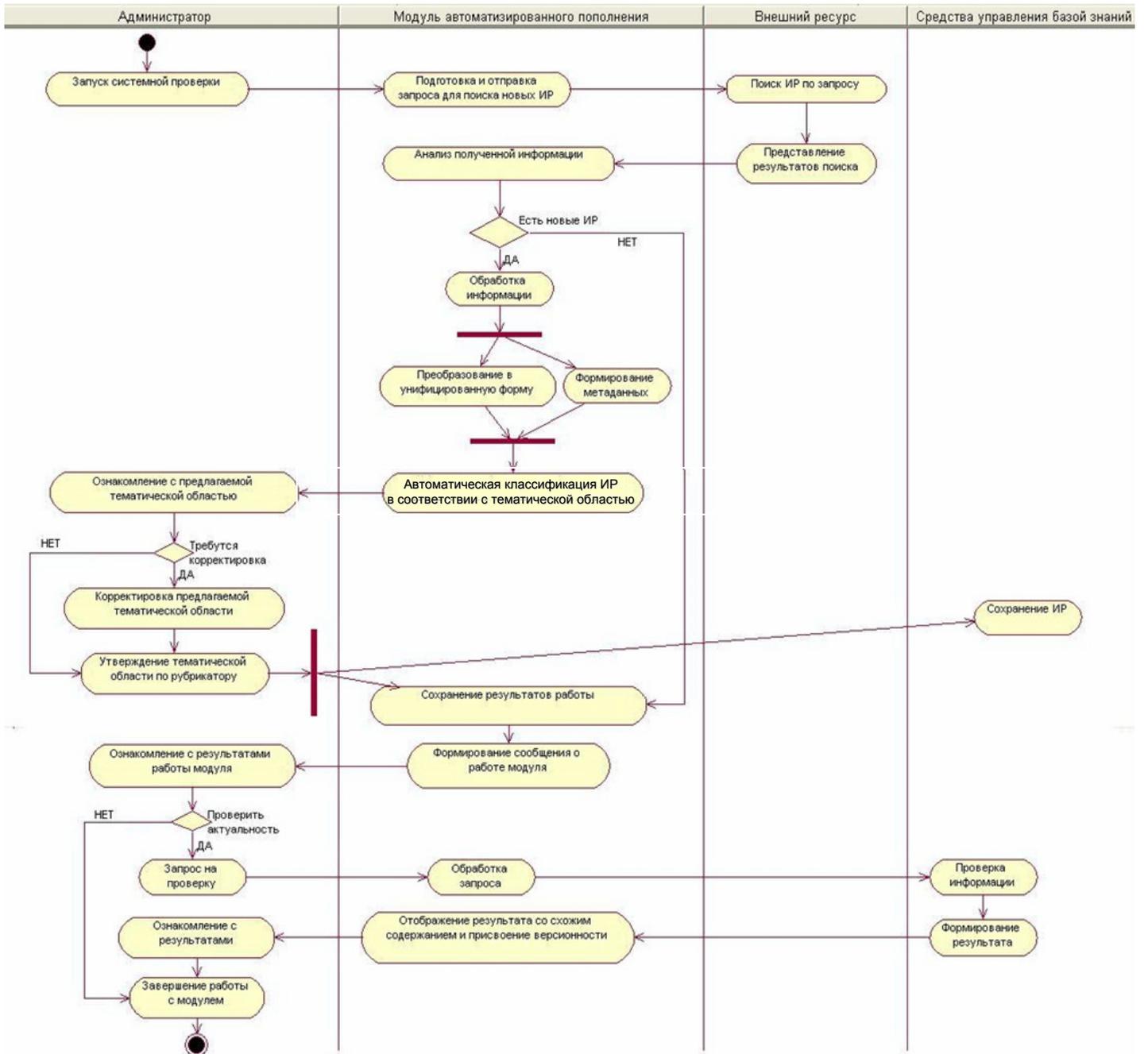


Рисунок 3 – Диаграмма деятельности, отображающая работу с модулем автоматизированного пополнения БЗ

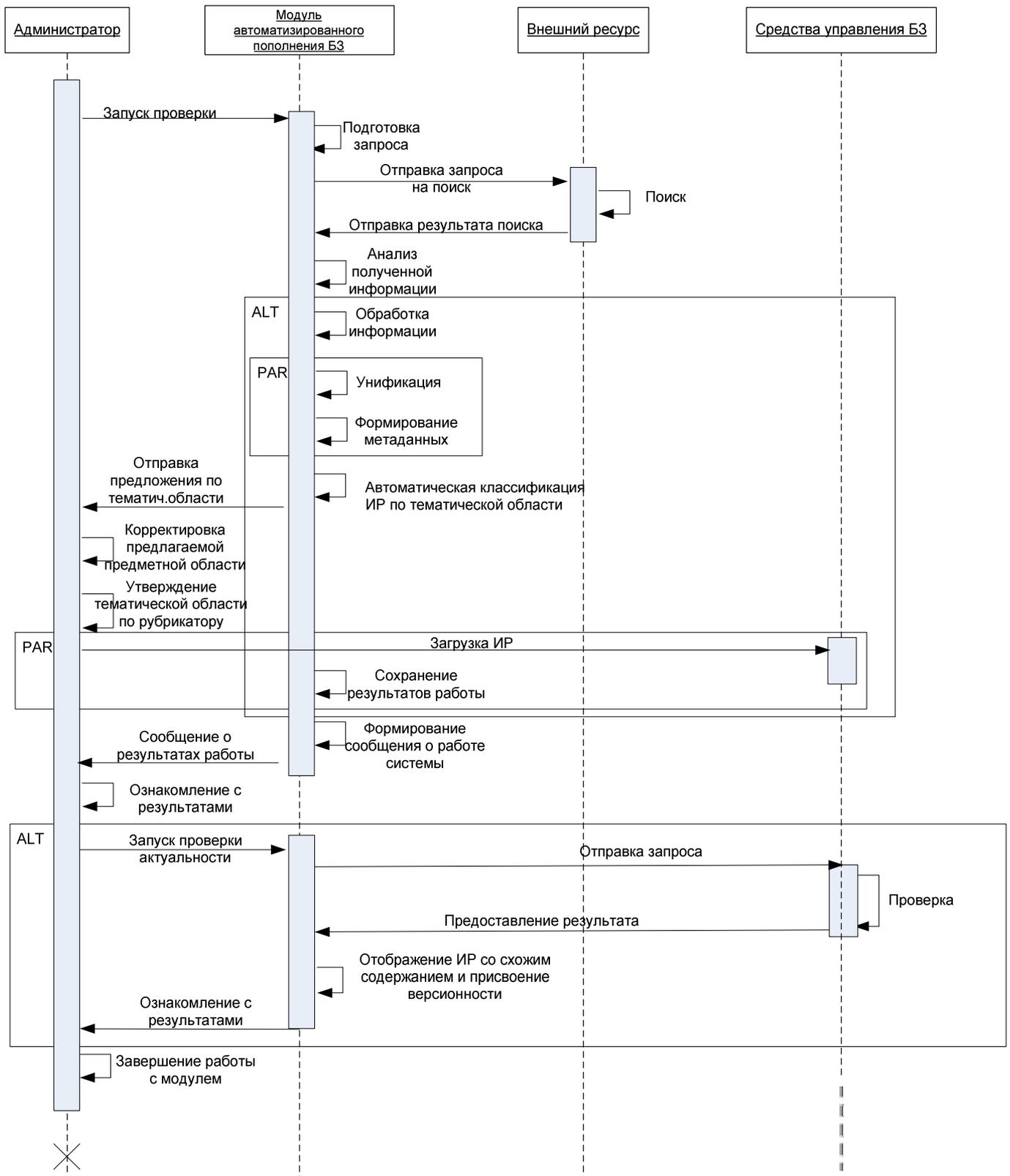


Рисунок 4 – Диаграмма последовательности, отображающая работу с модулем автоматизированного пополнения БЗ

## БИБЛИОГРАФИЯ

- [1] Захарова И.Г. Информационные технологии в образовании. – М.: Изд. центр «Академия», 2003. – 192 с.
- [2] Сигов А.С., Нечаев В.В., Кошкарёв М.И. Архитектура предметно-ориентированной базы знаний интеллектуальной системы. *International Journal of Open Information Technologies*. Том 2, №12 (2014), с.1 – 6.
- [3] A. Sigov, V. Nechaev, V. Baranyuk, M. Koshkarev, A. Melikhov, O. Smirnova, A. Bogoradnikova. Architecture of domain-specific data warehouse for bionic information resources. *Ecology, Environment and Conservation Paper*. Vol. 21, Nov. 2015 Suppl. Issue, pp 181 – 186.
- [4] Нечаев В.В., Баранюк В.В., Смирнова О.С., Кошкарёв М.И., Володина А.М., Богорадникова А.В., Маркелов К.С. Учебное пособие «Информационные ресурсы и технологии» по курсам «Базы данных», «Хранилища данных и OLAP-технологии», «Системный анализ», «Информационные технологии» для студентов, обучающихся по направлению 09.03.04 «Программная инженерия». Изд.: ФГБОУ ВО «МИРЭА». 2015. – 92 с.
- [5] A. Sigov, V. Nechaev, V. Baranyuk, O. Smirnova, A. Melikhov, M. Koshkarev, A. Bogoradnikova. Bionic-oriented information system for innovation activities. *Indian Journal of Science and Technology*, Vol 9 (30), 2016. URL: <http://www.indjst.org/index.php/indjst/article/view/98743>
- [6] Баранюк В.В., Смирнова О.С., Богорадникова А.В. Интеллектуальная система информационной поддержки развития перспективных бионических технологий: основные направления работ по созданию. *International Journal of Open Information Technologies*. Том 2, №12 (2014), с.17 – 19.
- [7] Manning C., Raghavan P., Schütze H. *Introduction to Information Retrieval*. Cambridge University Press, 2008. 544 p.
- [8] Adam Berger. *Statistical Machine Learning for Information Retrieval*. Carnegie Mellon University, 2001. 143 p.
- [9] Hand D., Mannila H. «Principles of Data Mining», The MIT Press 2001.
- [10] Witten I. H., Frank E. *Data Mining: Practical Machine Learning Tools and Techniques (Second Edition)*. Morgan Kaufmann, 2005. 525 p.
- [11] Paliouras G., Karkaletsis V., Spyropoulos C. D. *Machine Learning and Its Applications: Advanced Lectures (Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence)*. Springer, 2001. 325 p.
- [12] Баранюк В.В., Смирнова О.С. Роевой интеллект как одна из частей онтологической модели бионических технологий. *International Journal of Open Information Technologies*. Том 3, № 12 (2015), с. 13 – 17.
- [13] Баранюк В.В., Смирнова О.С. Детализация онтологической модели по роевым алгоритмам, основанным на поведении насекомых и животных. *International Journal of Open Information Technologies*. Том 3, № 12 (2015), с. 18 – 27.
- [14] Смирнова О.С., Богорадникова А.В., Блинов М.Ю. Описание роевых алгоритмов, инспирированных неживой природой и бактериями, для использования в онтологической модели. *International Journal of Open Information Technologies*. Том 3, № 12 (2015), с. 28 – 37.
- [15] Мелихов А.А., Нечаев В.В. Пополнение базы знаний интеллектуальной системы информационной поддержки развития перспективных бионических технологий: формирование перечня источников. *Научный и общественно-информационный журнал «Информационные и телекоммуникационные технологии» №28, 2015. с. 16 – 20.*
- [16] Буч Г., Рамбо Д., Якобсон А. *Язык UML. Руководство пользователя*. 2-е издание. ДМК, 2006, 496 с.

# Knowledge base population subsystem of the intellectual informational support system for creation and development of the bionic technologies

V.V. Baranjuk, O.S. Smirnova, A.M. Volodina, M.U. Blinov

**Abstract** – In this article, we discuss primary issues of information resources classification during knowledge base population. Two operation models are rendered for information resources supplementation and automated knowledge base population module respectively. Modelling results are represented by activity and sequence UML diagrams embodying the corresponding subsystems of a greater information support system for bionic research.

**Keywords** – information resources classification, automated classification, knowledge base population, open information resources monitoring, model representation, bionic technologies.