



ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ
«Begemotik»

Система поиска и установления информационных
связей между данными

ОПИСАНИЕ РЕАЛИЗАЦИИ

СОДЕРЖАНИЕ

1. Аннотация	3
2. Назначение ПО	3
3. Программно-аппаратные среды функционирования ПО	3
4. Общие принципы функционирования ПО	3
5. Реализация ПО	4
5.1.Модуль интерфейса пользователя.....	4
5.2.Модуль интерфейса авторизации	4
5.3.Модуль коммуникации и контроля	5
5.4.Модуль анализа и обработки данных	5
5.5.Модуль визуализации данных	5
5.6.Модули получения данных.....	6
6. Взаимодействие ПО с автоматизированными системами	6



1. Аннотация

Настоящий документ содержит описание реализации программного обеспечения «Begemotik» (система поиска и установления информационных связей между данными, далее – ПО).

2. Назначение ПО

Система поиска и установления информационных связей между данными «Begemotik» предназначена для поиска и анализа данных из подключенных источников.

Основными целями создания системы являются:

- Предоставление удобного и понятного Оператору интерфейса по отображению и анализу данных
- Предоставление единого интерфейса для поиска данных во всех подключенных к системе источниках данных
- Консолидация разрозненной информации из источников данных и её отображение в стандартизированном виде

3. Программно-аппаратные среды функционирования ПО

ПО поставляется в виде инструкции по установке, содержащей все необходимое для установки системы, конфигурационные файлы, логины и пароли для доступа к серверам обновления.

Установка возможна на IBM-PC-совместимом компьютере, удовлетворяющем следующим минимальным характеристикам:

- Процессор: Intel/AMD с 4 и более ядрами 2400+MHz с поддержкой технологии Hyper Threading
- Оперативная память: 32GB и более (рекомендуемое количество — 64GB)
- Дискретная видеокарта с GPU и 4GB собственной памяти (рекомендуется — серия GeForce от Nvidia)
- Монитор, клавиатура, мышь.

4. Общие принципы функционирования ПО

Система поиска и установления информационных связей между данными «Begemotik» подключается к источникам данных с помощью программного интерфейса (API) соответствующего ресурса (при наличии) или с помощью соответствующего модуля Системы с последующей обработкой ответа и приведения его в вид, понятный Системе.

ПО не хранит никаких данных на своей стороне (кроме настроек подключения к источникам данных, а также алгоритмов обработки и интерпретации данных) и обрабатывает данные непосредственно в момент запроса Оператора с последующим отображением результатов запроса в виде интерактивного графа (т.е. набора узлов — найденных сущностей и набора рёбер — связей между ними)

5. Реализация ПО

ПО состоит из следующих модулей:

- Модуль интерфейса пользователя
- Модуль интерфейса авторизации
- Модуль коммуникации и контроля
- Модуль анализа и обработки данных
- Модуль визуализации данных
- Модули получения данных (по количеству типов подключенных источников данных)

5.1. Модуль интерфейса пользователя

Данный модуль имеет следующий функционал:

- Обеспечение коммуникации Пользователя с Системой
- Работа с ресурсными файлами, необходимыми для работы Системы (изображения, шрифты, скрипты, таблицы стилей и т.д.)
- Работа с кэшем (для ускорения работы интерфейса)
- Визуализация данных

5.2. Модуль интерфейса авторизации

Данный модуль имеет следующий функционал:

- Работа с ресурсными файлами, необходимыми для работы системы авторизации (изображения, шрифты, скрипты, таблицы стилей и т.д.)
- Коммуникация с модулем безопасности

5.3. Модуль коммуникации и контроля

Данный модуль необходим для обеспечения безопасности Системы. Здесь происходит авторизация пользователей, валидация всех данных и проводятся все проверки безопасности.

Модуль имеет следующий функционал:

- Работа с базой пользователей Системы, аутентификация и авторизация пользователей
- Валидация запросов (как внешних — запросы пользователей, так и внутренних — необходимых для функционирования Системы)
- Мониторинг работоспособности других компонентов Системы

5.4. Модуль анализа и обработки данных

В данном модуле идёт непосредственный анализ полученных данных и их интерпретация. Модуль обладает следующим функционалом:

- Получение данных от Модулей получения данных
- Обработка полученных данных и их конвертация в вид, понятный Системе
- Проведение анализа полученных данных, применение разработанных алгоритмов, поиск связей, генерация эмпирических данных (т.е. данных, не содержащихся непосредственно в данных, полученных от источников, но добавленных Системой на основе проведённого анализа)
- Передача полученных данных в Модуль визуализации данных

5.5. Модуль визуализации данных

Поскольку объём данных, с которыми работает Система достаточно велик для обработки в модуле интерфейса пользователя, все расчеты, необходимые для отображения данных в виде графа происходят в отдельном модуле, специально предназначенном для визуализации данных и их подготовке для отображения на экране пользователя.

Основной функционал данного модуля:

- Хранение оперативных данных по текущим запросам (в виде набора вершин и ребёр)
- Расчёт координат узлов и их актуализация по мере поступления данных (для выбранной раскладки графа, по-умолчанию — ForceAtlas2)
- Расчёт весов узлов и связей на основе аналитических данных и внутренних алгоритмах

- Принятие решения об отображении или скрытии данных на графе
- Синхронизация данных графа с клиентским модулем визуализации графа

5.6. Модули получения данных

Система построена таким образом, что для получения данных со всех внешних источников используются отдельные модули, разработанные для получения данных из определённого типа источника (например, конкретной СУБД), либо определённого способа получения данных (например, с помощью HTTP-запросов) — в данном случае такой модуль может обеспечивать получение данных со всех ресурсов, поддерживающих соответствующий тип запросов и имеющих схожую структуру данных.

Все модули получения данных построены по общей схеме и имеют аналогичный функционал:

- Валидация и нормализация входных данных, необходимых для осуществления запроса, например, приведение телефона или адреса электронной почты в стандартный вид — удаление лишних символов из телефона (пробелы, дефисы, скобки и т.д.), приведение адреса электронной почты в нижний регистр и т.д.
- Осуществление запроса к внешней системе (источнику данных)
- Получение данных и их валидация (принятие решения об их соответствии ожиданию Системы)
- Контроль работоспособности источника данных (доступность, возможность отправки запросы и получения адекватного результата)
- Контроль за количеством и частотой запросов к источникам данных
- Ведение статистики о количестве и типах отправленных запросов

6. Взаимодействие ПО с автоматизированными системами

Взаимодействие ПО с автоматизированными системами не предусмотрено.