

УДК 004.4+681.518

ПРОГРАММНОЕ СРЕДСТВО ДЛЯ ПЕРСОНАЛЬНОЙ БАЗЫ ЗНАНИЙ С АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИЕЙ ТЕКСТОВОЙ ИНФОРМАЦИИ

Веревкин А.А., студент гр.181074

*Белорусский государственный университет информатики и радиоэлектроники,
Институт информационных технологий,
г. Минск, Республика Беларусь*

Кунцевич О.Ю. – канд. пед. наук, доцент, доцент кафедры ИСиТ

Аннотация. Рассмотрены вопросы разработки программного средства для ведения и управления персональной базой знаний, представлен концепт пользовательского интерфейса приложения, обоснован выбор средств разработки, перечислены достоинства и недостатки применения разработки в виде настольного приложения. В качестве языка программирования выбран язык JavaScript. Отличительной чертой программного средства от аналогов являются: визуализация данных, автоматическая классификация данных.

Ключевые слова. Программное обеспечение, разработка приложений, визуализация данных.

Введение. Предметная область данной работы связана с классификацией и категоризацией информации, что позволяет систематизировать знания и упорядочивать восприятие мира. Категоризация рассматривается как процесс формирования понятий, обобщающих результаты познавательной деятельности, и основывается на различных подходах: классическом, прототипическом, когнитивном, логическом, полевым и оппозитивным.

Например, прототипическая теория предполагает, что у каждой категории есть центральный пример (прототип), вокруг которого группируются остальные её элементы в зависимости от степени сходства. Когнитивный подход делает акцент на гибкости, градиентности и зависимости категорий от контекста и опыта [1]. Современные методы управления знаниями, такие как Zettelkasten, помогают создавать взаимосвязанные сети заметок, структурировать информацию и стимулировать развитие идей [2]. Для автоматизации классификации и категоризации текстовой информации используются методы обработки естественного языка и тематического моделирования. Эти технологии позволяют эффективно анализировать большие объёмы текстов, создавая структурированные базы знаний и улучшая поиск и обработку данных.

На сегодняшний день существует достаточно большое количество программных средств, реализующих ведение и группировку заметок для структуризации знаний. К таким программным средствам можно отнести TiddlyWiki, Zoho Notebook, Obsidian. В свою очередь, TiddlyWiki представляет собой веб-приложение, основанное на HTML/JavaScript, которое позволяет создавать персональные веб-страницы. Недостатками данного приложения являются сложность его первоначальной настройки и проблемы с сохранением данных на устройстве. Поскольку приложение после каждого сохранения генерирует новый HTML файл и сохраняет его на устройстве, то это вызывает проблемы с контролем версий персональных веб-страниц.

Zoho Notebook является облачным программным средством для ведения заметок [3]. Представляет собой визуальную систему управления автономными заметками. Недостатками этого средства являются: отсутствие локального хранения данных, нет поддерживаемых связей между заметками, поскольку программное средство фокусируется на простых независимых заметках.

Obsidian представляет собой программное средство для управления заметками и базами знаний, которое фокусируется на создании взаимосвязанных заметок с помощью внутренней сети связей [4]. К недостаткам можно отнести сложность использования, особенно для пользователей, которым нужны простые текстовые заметки и отсутствие сохранения по умолчанию данных в удаленном хранилище.

Практическая значимость результатов разработки заключается в возможности применения программного средства в области самообразования, исследования новых процессов, структуризации объемных и комплексных областей знаний. Оно помогает автоматически распределять заметки, документы или статьи по категориям, что ускоряет поиск информации. Система полезна для учёных, студентов и аналитиков, позволяя эффективно управлять знаниями, выявлять связи между идеями и находить еще не описанные области. Также инструмент подходит для самообразования, позволяя создавать тематические подборки материалов и облегчая работу с новыми данными. Такое решение значительно повышает продуктивность в изучении новой области знаний.

Основная часть. После анализа предметной области и обзора существующих аналогов, в рамках данного исследования были определены назначение и возможности программного средства:

— главной задачей разрабатываемого приложения является облегчение ведения и управления пользовательской информацией о различных областях исследований, реализуя современные методы управления знаниями.

— программное средство должно быть кроссплатформенным и предоставлять возможность сохранения данных как локально на устройстве, так и в удаленном хранилище.

— программное средство должно осуществлять возможность автоматизации работы категоризации информации среди определенных пользователем категорий.

— программное средство предоставит возможность визуализации данных пользователя в виде графа, который также возможно использовать для навигации по заметкам.

Уникальность и отличия разработки от аналогичных программных средств заключается, в частности, в следующих пунктах:

- связывание заметок через ссылки;
- поддержка локального сохранения и в удаленном хранилище;
- визуализация связей и заметок;
- автоматическая категоризация пользовательских заметок.

Реализация программного средства была выполнена с помощью языка программирования JavaScript, используя фреймворк NW.js [5]. Кроме того, была использована система управления реляционными базами данных MySQL [6], а также редактор кода Visual Studio Code.

Одним из ключевых предпроектных этапов разработки программного обеспечения является системный анализ и моделирование предметной области. В рамках этого этапа была создана контекстная диаграмма IDEF0, отображенная на рисунке 1.



Рисунок 1 – Контекстная диаграмма A-0 в нотации IDEF0

При запуске программного средства отображается окно авторизации пользователя. Здесь также доступна регистрация нового пользователя по имени и адресу электронной почты. После успешной авторизации в системе, пользователю будут доступны основные модули программного средства.

На рисунке 2 отображено главное окно программного средства.

Левая панель программного средства отвечает за навигацию и отображение категорий и заметок пользователя. Данная панель поддерживает функции создания, редактирования и удаления категорий и заметок пользователя.

Редактирование заметок и категорий на левой панели предполагает изменение их названия, а для изменения содержания заметки, используется центральная панель программного средства. Кроме того, для каждой категории и заметки здесь отображены кнопки выполнения функции отправки заметок в удаленное хранилище и обновления заметки. Также здесь присутствует переключатель отображения графа заметок.

Центральная панель представляет собой модуль форматирования активной заметки, которую выбрал пользователь на левой панели. Здесь присутствуют кнопки инструментов редактирования текста в разметке Markdown, а также переключатель отображения разметки. Для связывания двух заметок, можно воспользоваться тегом Markdown разметки, отвечающим за отображения ссылок. В качестве ссылки можно использовать название заметки. После установки ссылки связь между заметками будет отображена на графе.

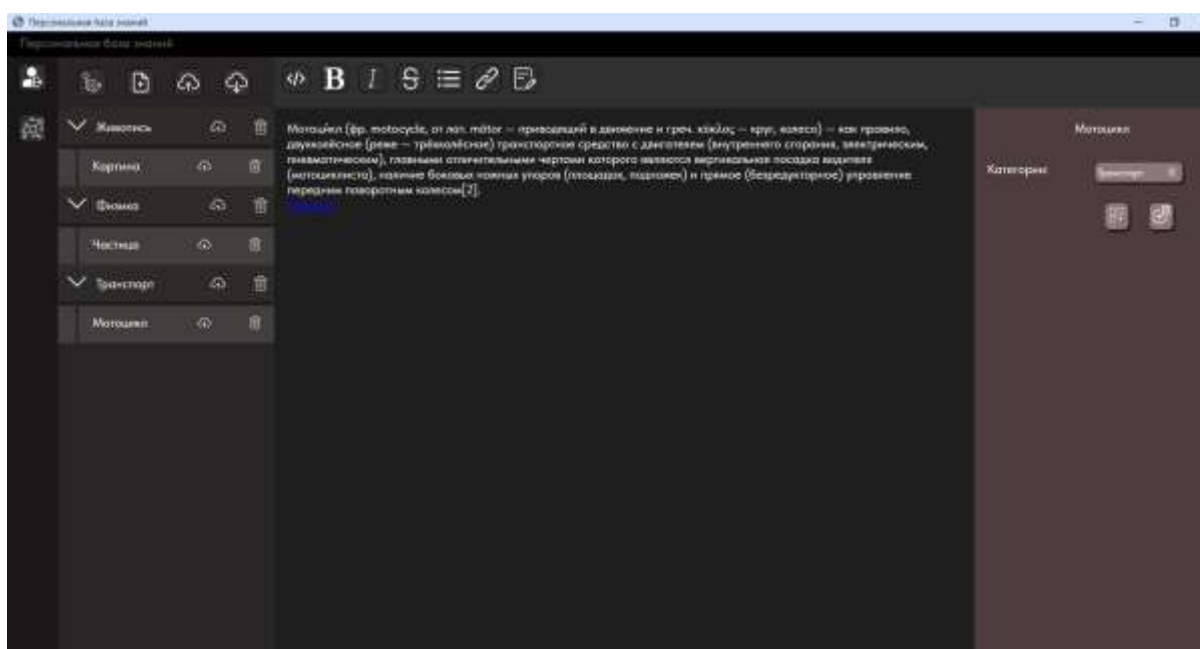


Рисунок 2 – Основные панели программного средства

Кроме того, здесь есть переключатель отображения правой панели программного средства, где находится информация о категориях, к которым относится активная заметка. Также на правой панели есть инструменты управления заметкой с точки зрения категорий. Здесь отображен лист категорий, к которым относится активная заметка.

В разработанном приложении существует два режима добавления категории: ручная и автоматическая. Ручной способ предполагает добавление активной заметки к одной из уже определенных пользователем категорий. Другими словами, ручной способ состоит из выбора пользователем категорий, к которым данная заметка еще не была отнесена. Автоматический способ добавляет наиболее подходящую по тексту категорию к заметке. Программное средство, используя текстовую модель данных, анализирует текст и названия категорий пользователя. Когда наиболее подходящий вариант категории найден, она добавляется к заметке автоматически. Этот способ наиболее простой, и хорошо подойдет для заметок, которые наиболее полным образом описывают ту или иную сущность. Также пользователю предоставлена возможность удаления заметки из категории.

На рисунке 3 отображен пример графа заметок.

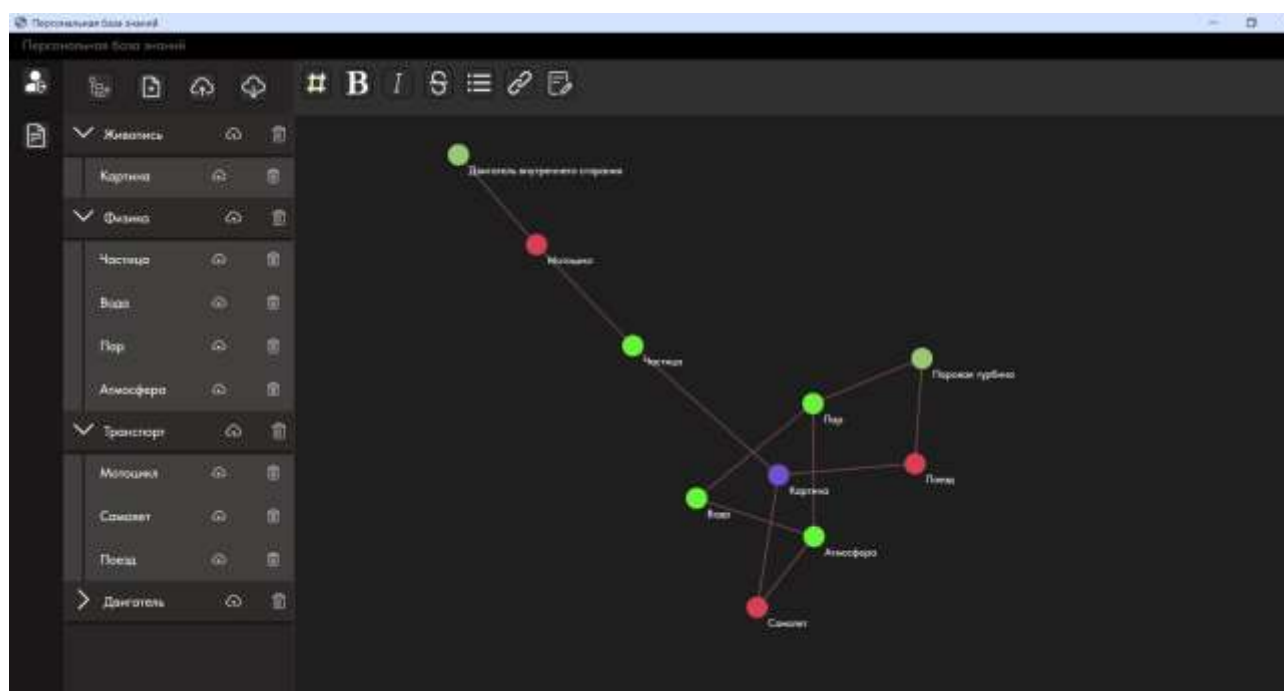


Рисунок 3 – Пример графа заметок

Каждый узел графа отображает заметку пользователя, при этом цвет узла одинаковый для каждой категории. Этот граф возможно двигать и изменять положения узлов для удобства пользователя или на случай попадания названий узлов друг на друга. Также этот граф может

использоваться для навигации, например, по двойному нажатию на узел будет произведен переход к заметке.

В ходе проверки работоспособности программного средства были протестированы функции управления категориями и заметками, локального сохранения, сохранения в удаленном хранилище, установки ссылок между заметками, ручного и автоматического присваивания категорий, отображения графа заметок.

Так, например, проведено тестирование автоматической классификации пользовательской заметки. В программном средстве были определены пять новых категорий заметок. Затем была создана текстовая заметка небольшого размера на одну из тем. Задача данной проверки состояла в удостоверении, что автоматический классификатор способен установить тему текста заметки из предложенных категорий. Ожидаемый результат совпал с фактическим, таким образом классификатор справился со своей задачей.

Для проведения функционального тестирования программного средства рассмотрены тест-кейсы по проверке создания новой заметки, проверке поиска по заметкам, проверке синхронизации заметок, проверке создания и использования ссылок на другие заметки. После тестирования выявленные ошибки в работе программного средства были успешно исправлены.

Заключение. Разрабатываемое программное средство ведения персональной базы знаний отображает пользовательские категории, заметки и связи между ними, включает модули авторизации пользователя, управления категориями и заметками, форматирования заметок, визуализации данных и навигации. Выполняет функции создания, управления и ведения пользовательских категорий и заметок. Позволяет форматировать заметки и создавать ссылки между заметками. Программное средство также позволяет отобразить связанные заметки в виде графа, который также выполняет функцию навигации по заметкам. Кроме того, программное средство облегчает задачу классификации текстовых заметок, выполняя ее автоматически по пользовательским категориям. Главное преимущество программного средства перед аналогами заключается в визуализации пользовательских заметок и их автоматической категоризации.

В рамках разработки приложения были смоделированы контекстная диаграмма и ее декомпозиции первого и второго уровней, диаграмма вариантов использования, диаграмма классов, а также алгоритмы работы клиентской и серверной частей системы.

Расчет экономических показателей подтвердил, что разработанный проект успешен и окупается быстрее ожидаемого. Основными рисками, связанными с разработкой и продажей программного средства широкой аудитории являются: высокий уровень конкуренции и недостаточный спрос на продукцию такого вида.

Список использованных источников:

- 1 Боярская, Е. Л. Категоризация как базовая когнитивная процедура / Е. Л. Боярская – Вестник Балтийского федерального университета им. И. Канта: 2011 – 7 с.
2. Doto B. A System for Writing: How an Unconventional Approach to Note-Making Can Help You Capture Ideas, Think Wildly, and Write Constantly – A Zettelkasten Primer / B. Doto – NY: New Old Traditions, 2024. – 211 с.
3. Zoho Notebook [Электронный ресурс] – Режим доступа: <https://www.zoho.com/notebook/notebook-for-windows.html>.
4. Obsidian [Электронный ресурс] – Режим доступа: <https://obsidian.md/>.
5. Техническая документация по NW.js [Электронный ресурс]. – Режим доступа: <https://nwjs.readthedocs.io/en/latest/>
6. Аспин, А. MySQL. Практические рецепты / А. Аспин. – СПб: БХВ-Петербург, 2024. – 592 с.

UDK 004.4+681.518

SOFTWARE TOOL FOR WORKING WITH A PERSONAL KNOWLEDGE BASE WITH AUTOMATIC CLASSIFICATION OF TEXT INFORMATION

Verevkin A.A., student

*Institute of Information Technologies of the Belarusian State University of Informatics and Radioelectronics,
Minsk, Republic of Belarus*

Kuntsevich O.Yu. – Candidate of Pedagogic Sciences, Associate Professor

Annotation. The article deals with the development of a software tool for maintaining and managing a personal knowledge base, presents the concept of the user interface of the application, justifies the choice of development tools, lists the advantages and disadvantages of using the development in the form of a desktop application. The JavaScript language is chosen as a programming language. The distinctive features of the PS from analogues are data visualization and automatic data classification.

Keywords. Software, application development, data visualization.