

УДК 004.912

## ПРОГРАММНЫЙ КОМПЛЕКС СЕМАНТИЧЕСКОГО ПОИСКА ИНФОРМАЦИИ В ОБЛАСТИ КОСМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

И. М. Бойко, Д. С. Писаревский, А. А. Мацко

УП «Геоинформационные системы» НАН Беларуси, Минск

*Приводится описание программного комплекса, который реализует функции мониторинга и семантического поиска информации в области космической деятельности. Программный комплекс используется для оперативного получения информации пользователями организаций, работающих в области космической деятельности. Программный комплекс развивается и внедряется в рамках программы Союзного государства «Интеграция-СГ».*

### Введение

В настоящее время наиболее используемым источником актуальной и оперативной информации является Интернет. Традиционные поисковые системы применяют процедуры поиска по ключевым словам и предоставляют результаты в виде единого и необозримого списка ссылок на текстовые документы. Избыточность результата затрудняет анализ и оценку получаемой информации. Решением проблемы своевременного и достоверного информационного обеспечения является разработка интеллектуального программного инструмента для профильных специалистов, позволяющего найти и обобщить материал по задаваемой тематике, например в области космической деятельности (КД). Создание такого современного сервиса информационной поддержки организаций НАН Беларуси является актуальной задачей.

### 1. Общая характеристика программного комплекса

Программный комплекс (ПК) представляет собой клиент-серверное приложение. Клиент и сервер взаимодействуют друг с другом в сети Интернет при помощи сетевых протоколов. ПК построен по модульному принципу. Каждый программный модуль (ПМ) выполняет определенную функцию, обеспечивая работу соседнего модуля (рис. 1).

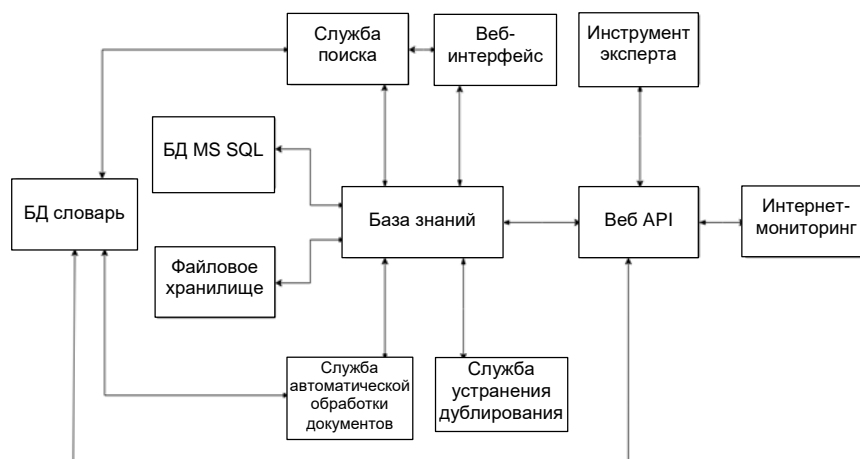


Рис. 1. Структура ПК

ПК обрабатывает информацию на двух языках: русском и английском.

ПМ интернет-мониторинга (блок Интернет-мониторинг) производит регулярный мониторинг текстовой информации в сети Интернет в области КД в соответствии с изначально заданным и постоянно пополняемым списком интернет-ссылок, обновляет базу данных и наполняет ее текстами, извлеченными во время мониторинга.

ПМ верификации (блок Служба устранения дублирования) определяет тематическую схожесть текстов, удаляет дублирование, автоматически группирует тексты по семантическим признакам и распределяет по заданным тематическим категориям, используя уже существующие методы [1]. Тематические категории, например навигационное оборудование космического аппарата, ориентация и стабилизация космического аппарата, система обеспечения температурных режимов космических аппаратов и др., заранее заданы экспертами.

ПМ обработки текстов (блок Служба автоматической обработки документов) автоматически формирует таксономию в виде иерархических отношений космической терминологии, выявленной при анализе тематических текстов (рис. 2). Эксперты проводят выборочную оценку и корректировку таксономии с помощью дополнительного программного инструмента. Кроме того, используется словарь терминов (блок БД словарь) с их определениями, который сливается с извлеченными терминами, добавляя к ним существующие определения для отображения в веб-интерфейсе.

В целом формируется специализированная база знаний в области КД (блок База знаний).

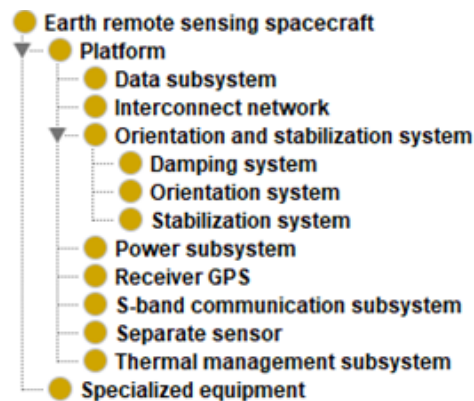


Рис. 2. Фрагмент таксономии терминов ДЗЗ на английском языке

Пользовательский интерфейс (блок Веб-интерфейс) предоставляет сервис поиска релевантной информации в сформированной базе. Также в режиме реального времени осуществляются отслеживание и отображение новостей в области КД, в том числе размещаемых на информационных страницах мировых космических агентств.

ПМ реализованы на языках программирования C#, C++, Python. Формирование, хранение данных и обращение к базе осуществляются с помощью MS SQL. Посредством SQL-таблиц организована и таксономия терминов. В настоящее время таксономия преобразуется в онтологию, в которой задаются не только сами термины, но и семантические и логические связи между ними. Использование онтологии улучшает результаты поиска в базе, предоставляя тексты, содержащие термины из запроса пользователя, а также семантически связанные с ними тексты.

## 2. Перспективы развития программного комплекса

Перспективной представляется разработка ПК на основе технологии Semantic WEB, использующей Web Ontology Language (OWL) и Resource Description Framework (RDF) в соответствии с международными стандартами [2, 3].

Основным компонентом ПК является онтология, состоящая из формального представления тематической терминологии на основе базового словаря и включающая определения и аксиомы, которые делают предполагаемые значения терминов явными и описывают их взаимосвязи. Онтология поддерживает представление сущностей, интерпретацию данных и вывод знаний на основе семантических взаимосвязей. Данные из различных дисциплин, включая информацию в области КД или географическую информацию, могут быть интегрированы в онтологии и способствовать решению различных прикладных задач [2].

Онтологии успешно разрабатываются и внедряются в практику [1, 5]. Как пример из области КД для damping system в редакторе онтологий Protégé [4] отображается и обрабатывается целый ряд взаимосвязей (рис. 3):

- Definition: on-board automatic motion control system for damping spacecraft angular velocity;
- SubClassOf: orientation and stabilization system;
- property change\_parameter: angular velocity;
- property implement\_function: damp.

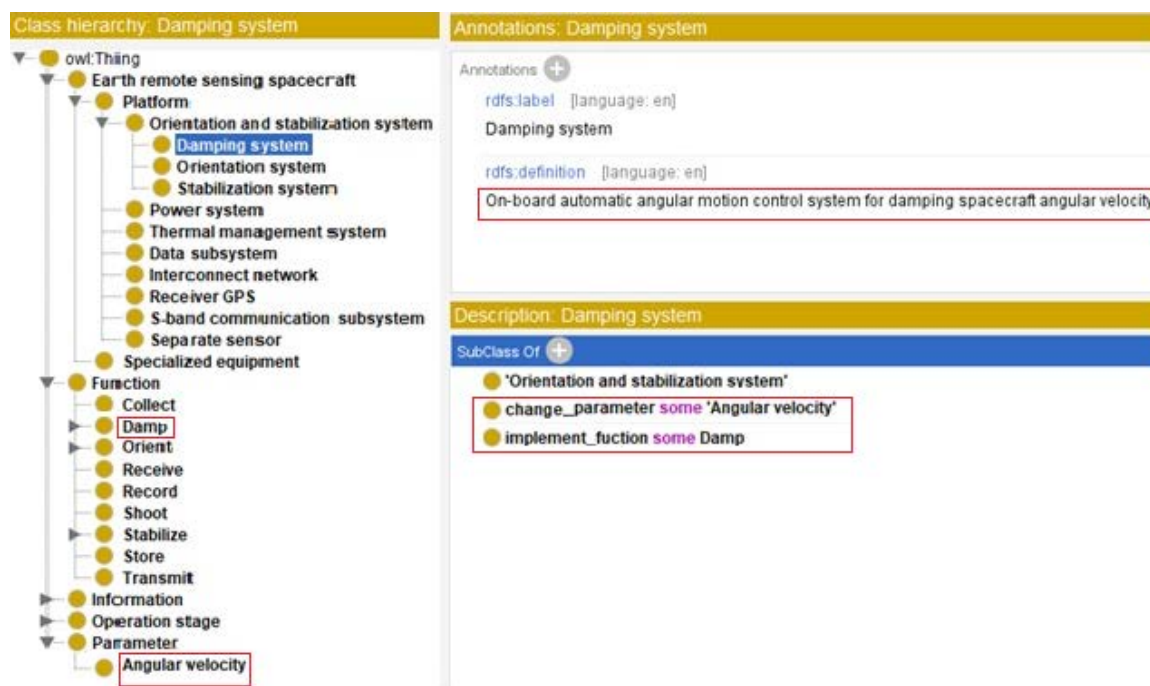


Рис. 3. Фрагмент онтологии терминов ДЗЗ на английском языке

## Заключение

ПК разработан в рамках Государственной программы «Научно-технические инновационные технологии и техника» на 2016–2020 гг. Пользователями ПК являются сотрудники аппарата и организаций НАН Беларуси, участвующих в информационной поддержке деятельности в области КД; сторонние пользователи, а также эксперты.

ПК используется при подготовке обоснований и поддержке принятия целевых решений для перспективных задач в области КД Республики Беларусь и обладает возможностью функционального масштабирования не только на космическую отрасль, но и на любую другую, например авиационную или на сельское хозяйство.

Кроме того, ПК может быть использован в интересах госорганов, формирующих программы и проекты научных исследований и т. д. (например, ГКНТ), а также в интересах юридических и физических лиц.

### Список литературы

1. Буравкин, А. Г. Рубрикация текстовой информации в системе интернет-мониторинга публикаций по космической тематике / А. Г. Буравкин, С. Ф. Липницкий, Л. В. Степура // Седьмой Белорусский космический конгресс : материалы конгресса : в 2 т., Минск, 24–26 окт. 2017 г. – Минск : ОИПИ НАН Беларуси, 2017. – Т. 2. – С. 210–213.
2. ISO/TS 19150-1:2012(E). Geographic information – Ontology – Part 1: Framework [Electronic resource]. – 2012. – Mode of access: <https://www.iso.org/standard/57465.html>. – Date of access: 22.04.2022.
3. ISO 19150-2:2015/Amd.1:2019(E). Geographic information – Ontology – Part 2: Rules for developing ontologies in the Web Ontology Language (OWL) [Electronic resource]. – 2019. – Mode of access: <https://standards.iteh.ai/catalog/standards/iso/f544f434-a3f3-40fe-a540-4a84da8b5cda/iso-19150-2-2015>. – Date of access: 22.04.2022.
4. A free, open-source ontology editor and framework for building intelligent systems [Electronic resource]. – Mode of access: <https://protege.stanford.edu/>. – Date of access: 22.04.2022.
5. Золотой, С. А. Методология построения интеллектуального интернет-сервиса на основе онтологии для использования в области космической деятельности / С. А. Золотой, И. М. Бойко, С. Д. Писаревский // Проблемы инфокоммуникаций. – 2021. – № 2(14). – С. 59–66.