Веб-сервис для интеллектуального поиска планетных данных на основе пространственного и семантического контекста онлайн-карт небесных тел

Е. В. Матвеев¹, А. Ю. Жаркова^{1,2}, И. П. Карачевцева¹

¹ *Московский государственный университет геодезии и картографии
Москва, 105064, Россия
E-mail: e_matveev@miigaik.ru*

² *Государственный астрономический институт имени П. К. Штернберга
Московского государственного университета имени М. В. Ломоносова
Москва, 119234, Россия
E-mail: a_zharkova@miigaik.ru*

Представлены результаты разработки специализированного веб-сервиса, обеспечивающего интеллектуальный поиск планетной информации. Поисковый сервис создан в рамках формирования информационной системы для обработки космической съёмки небесных тел и доступа к результатам через геопорталы. Разработка специальных средств интеллектуализации поисковых сервисов вызвана огромным объёмом и разнообразием информации, накопленной в различных научных областях, и в частности в планетных исследованиях. При разработке сервиса выполнен анализ цифровых архивов планетных данных и программных средств, обеспечивающих доступ к хранимой в них информации. Показано, что, помимо использования программных средств самих веб-ресурсов (API), применение в качестве дополнения метода тематического индексирования исходных интернет-источников позволяет улучшить доступ к информации, обеспечивая её оптимизированный поиск. Созданный поисковый сервис интегрирован с онлайн-картами небесных тел, реализованными с помощью средств ArcGIS, и обеспечивает поиск данных с учётом пространственного и семантического контекста, представленного на картах. В дальнейшем предполагается использовать созданные поисковые возможности для Геопортала планетных данных Московского государственного университета геодезии и картографии (<http://cartsrv.mexlab.ru/geoportal>).

Ключевые слова: поисковый веб-сервис, интеллектуальный поиск, планетные данные, ГИС, онлайн-карты небесных тел

Одобрена к печати: 10.07.2019

DOI: 10.21046/2070-7401-2019-16-4-19-30

Введение

Современная цифровая эра и интернет привели к накоплению огромных массивов информации. Это касается различных видов деятельности, включая обработку геоданных. В сфере планетных исследований с использованием пространственных данных в результате интенсивного изучения Солнечной системы количество данных дистанционного зондирования небесных тел достигло значительного объёма и продолжает расти в связи с обработкой полученной информации. Так, существующее хранилище НАСА (PDS — Planetary Data System), которое на текущий момент содержит около 1 Пбайт сырых данных, вероятно, достигнет 2 Пбайт в ближайшие пять лет (Hare et al., 2017). Многообразие и объёмы данных, получаемых в рамках многочисленных космических миссий к телам Солнечной системы, таких как Lunar Reconnaissance Orbiter (LRO) (Луна), Mars Express (Марс), Messenger (Меркурий), обуславливает необходимость интеллектуализации поиска текущей и новой информации. Растёт не только массив данных, но также и количество научных публикаций, с ними связанных, включая новые сведения о планетных объектах, в том числе и искусственных, таких как планетоходы. Например, сейчас марсоходы передают уникальные данные с поверхности Марса, включая панорамные изображения, результаты геологических исследований и т. п. В огромном информационном потоке могут быть упущены не только какие-либо важные данные,

но и результаты научных исследований. Поэтому всё необходимое становится полнота описания разнородной и разновременной информации, доступной в сети интернет. Для решения этой проблемы в задачах поиска данных о небесных телах создан специализированный веб-сервис для глобального анализа интернет-ресурсов, связанных с планетной тематикой.

Общая схема функционирования веб-сервиса для интеллектуального поиска планетной информации с помощью онлайн-карт

Результаты планетных исследований, выполненные с помощью данных дистанционного зондирования и результатов их обработки, удобно представлять в картографическом виде. Современные карты часто публикуются с помощью онлайн-сервисов. Поэтому разработанный поисковый модуль интегрирован в онлайн-карты, что обеспечивает поиск данных с учётом пространственного и семантического контекста, представленного на картах.

Исходные пространственные данные и результаты первичной обработки, которые получены командами, работающими непосредственно с инструментами, установленными на орбитальных аппаратах, ведущих изучение небесных тел, хранятся в архивах космических агентств. Данные дистанционного зондирования и результаты их обработки, включая изображения поверхности небесных тел, производные продукты (ЦМР — цифровая модель рельефа; ортомозаики), а также различные тематические карты, могут храниться в архивах планетных миссий космических агентств, таких как, например, НАСА (<https://pds.jpl.nasa.gov/>), архив данных которого (PDS) является крупнейшим в мире на современном этапе. Помимо этих данных, существует множество результатов научных исследований, основанных на тематической обработке и анализе пространственной информации. Информация может быть любого типа или формата, включая космические изображения, ЦМР, карты, а также текстовые описания (научные статьи, технические отчёты, стандарты, метаданные), которые могут находиться на различных веб-ресурсах.

Поэтому для оптимизации работы веб-сервиса предварительно был выполнен анализ возможных источников данных о небесных телах Солнечной системы. Тематические источники планетной информации могут быть разделены на два основных типа:

1. Ресурсы, предоставляющие пространственные данные, которые доступны посредством публичных сервисов: архивы растровых изображений поверхности небесных тел, например Planetary Data System (PDS, <https://pds.jpl.nasa.gov/>), наборы векторных слоёв для ГИС (PIGWAD, <http://webgis.wr.usgs.gov/>), тематическая информация об объектах рельефа (каталоги названий с координатами объектов на поверхности небесных тел (Gazetteer of Planetary Nomenclature, <http://planetarynames.wr.usgs.gov/>)).
2. Ресурсы, предоставляющие семантическую (текстовую) информацию об объектах исследований: различные спецификации и стандарты (в частности, стандарт описания растровых данных в системе PDS (<https://pds.jpl.nasa.gov/tools/standards-reference.shtml>)), описание исследований (например, проводимых в Японском аэрокосмическом агентстве (<https://repository.exst.jaxa.jp/dspace/jaxapress?select=jaxapress>)), а также научные статьи, отчёты, новости и т. д.

Функциональность поискового веб-сервиса реализована в онлайн-картах планетных тел, опубликованных с помощью программного обеспечения (ПО) ArcGIS (рис. 1, см. с. 21). Считывание информации является основополагающей задачей работы системы поиска. Реализация этого процесса основана на пространственной выборке объекта (как векторного, так и растрового) в окне онлайн-карты.

Интеллектуализация работы поискового алгоритма реализована методом нейронных сетей и заключается в определении типа векторных объектов пользователем в окне онлайн-карты (например, кратер определённого типа на определённой территории небесного тела и т. д.), осуществлении поиска на внешних ресурсах и в конкретизации ответов поискового запроса на внешних ресурсах на основе «умной» фильтрации и сортировки массива входных параметров (Матвеев и др., 2016). Нейронные сети эффективны для решения задач поиска

в сети интернет (Ferreira, 2006) и обеспечивают быстрое обучение даже на неполных данных, фильтруют «шумы» входных параметров и легко адаптируются к изменениям этих параметров (Graves et al., 2014). Нейронные сети использованы при формировании списка взаимосвязанных входных параметров, которые отвечают персональным потребностям пользователя. Этот список является динамическим, так как с течением времени алгоритм самонастраивается через «обучение» (запросы по схожей тематике составляются с учётом истории использования системы поиска), предоставляя пользователю более персонализированные данные с каждым новым запросом. То есть алгоритм учитывает при выдаче информации статистику предыдущих запросов по схожей тематике (например, если раньше пользователь смотрел информацию по морфологии какого-либо объекта, то алгоритм поиска будет стараться подбирать данные по схожей тематике, но при условии, что такие данные будут содержаться в общем массиве информации, выданной при запросе к внешнему ресурсу). Также стоит отметить, что сохранена возможность просмотра всех результатов поиска. В списке ссылок при совместной работе модулей веб-скрапинга и поиска с API (application programming interface — программный интерфейс приложения) указываются ссылки на результат работы обоих модулей (см. рис. 6 и 7).

Рис. 1. Обобщённая схема функционирования веб-сервиса интеллектуального поиска планетной информации с помощью онлайн-карт

Входным параметром для работы модуля интеллектуального поиска является список взаимосвязанных атрибутов выбранного на карте объекта (кратера, борозды, векторной границы снимков космического аппарата, равнины и т.д.): семантическая (имя объекта, тип, морфометрические характеристики и т.д.) и пространственная (координаты объекта) информация, отфильтрованная по тематике всех атрибутов, а также по частоте обращения. Результат, полученный с помощью фильтрации в модуле интеллектуального поиска, реализованного с помощью созданного нами алгоритма на основе нейронных сетей (Матвеев и др., 2016), затем используется для формирования запросов во внешних веб-ресурсах для поиска планетных данных.

Алгоритм поиска информации планетных данных с использованием программных средств веб-ресурсов (API)

Космические агентства предоставляют доступ к цифровым хранилищам информации посредством интерфейсов прикладного программирования API — набора готовых классов, процедур, функций, которые обеспечивают доступ к своим данным или программным продуктам (Clarke, 2004). Выполнение API-запросов к внешним ресурсам может выполняться как процедурно (с помощью предоставляемых готовых программных библиотек или модулей для языков программирования), так и удалённо на сервисах внешних ресурсов (с помощью HTTP-запросов).

Также с помощью API организован доступ к различным веб-ресурсам, содержащим научные работы, результаты исследований, новости с сайтов космических агентств, а также описательную информацию (метаинформацию), связанную с искомым объектом (небесным телом, элементом рельефа на его поверхности и т. д.).

Рис. 2. Алгоритм поиска планетной информации на внешних ресурсах с помощью API

Модуль поиска данных с помощью API (рис. 2) реализован на языке программирования Python 2.7 (<https://www.python.org/>), поскольку он совместим с архитектурой ПО ArcGIS (<http://resources.arcgis.com/ru>), используемого нами для составления и публикации онлайн-

карт. Кроме того, для работы с ArcGIS создано огромное количество постоянно обновляемых программных библиотек Python для работы с пространственными данными. Для разработки программного модуля поиска данных с помощью API использованы следующие общедоступные библиотеки Python: NumPy (<http://www.numpy.org>), SciPy (<https://www.scipy.org>), requests (<http://docs.python-requests.org/en/master/>), psycopg2 (<http://initd.org/psycopg/>), Scrapy (<https://scrapy.org/>), urllib2 (<https://docs.python.org/2/library/urllib2.html>). Библиотеки psycopg2, NumPy и SciPy необходимы для организации доступа к базе данных PostgreSQL с сервисом статистики и обновления данных в её служебных таблицах. Функционал API и библиотек requests и Scrapy требуется для организации работы с веб-содержимым интернет-страниц, а именно: для передачи запроса к данным внешних источников информации, проверки доступности как ресурса, так и результата запроса к нему (ответ на запрос пользователя).

Отличительной особенностью работы модуля поиска объекта на веб-ресурсах с помощью API является то, что запрос к внешним источникам составляется по «восходящему» алгоритму (см. *рис. 2*). Первоначально поиск осуществляется по конкретизированным параметрам (используются, например, координаты объекта, его имя и т.д.), и при условии, что ответ на запрос пришёл не пустым, запрос с более обобщёнными параметрами производиться не будет. В случае если запрос не имеет результата, используется более обобщённая информация, например тип объекта (кратер, борозда, плато), его территориальное расположение на небесном теле и т.д. Поскольку, как правило, растровые объекты менее информативны и не содержат семантических атрибутов, то сначала обрабатывается атрибутивная информация векторных пространственных объектов, а затем метаданные растровых объектов.

Рис. 3. Пример работы веб-сервиса интеллектуального поиска планетной информации в интерфейсе онлайн-карты с использованием API

Итоговый список внешних источников информации отражает результаты работы «восходящего» алгоритма поиска: выводимые ссылки предоставляют одновременно как конкретную информацию об объекте, так и обобщённые данные.

Одним из важных этапов работы сервиса является проверка выполнения запросов поиска на ресурсах и доступности самих исходных источников. Доступность внешнего ресурса проверяется его откликом на запрос API. Если после запроса был выдан «пустой» ответ, т.е. по результатам запроса не было найдено никакой информации или ресурс по каким-то

причинам стал недоступен, этот источник не будет указан в результирующем списке ссылок (см. рис. 2).

Завершающим этапом работы модуля поиска с помощью API является обработка полученной от внешних источников информации и составление списка ссылок на эти ресурсы с последующим отображением в интерфейсе онлайн-карты (рис. 3, см. с. 23).

Алгоритм поиска планетной информации на веб-ресурсах с помощью веб-скрапинга

Веб-скрапинг — это сбор данных из сети интернет (Митчел, 2016) с помощью считывания и последующего индексирования собственными программными средствами исходных HTML-документов без использования API веб-ресурсов. Этот метод позволяет реализовать средства для автоматического опроса веб-серверов источников данных, запрашивать исходные данные и выполнять отбор нужной информации. В отличие от сбора данных с помощью API, когда структура возвращаемой информации определяется конкретным ресурсом, использование веб-скрапинга позволяет разработать собственную структуру, подходящую для заданных целей.

Рис. 4. Алгоритм поиска информации на внешних веб-ресурсах методом веб-скрапинга

Алгоритм поиска информации на внешних веб-ресурсах методом веб-скрапинга состоит из нескольких последовательных процедур (рис. 4, см. с. 24):

- считывание содержимого «сырой» веб-страницы;
- отфильтровка её HTML-кода (выделение элементов разметки страницы, в которых может находиться полезная информация);
- анализ полученных данных и нахождение в них необходимой информации;
- проверка актуальности ссылки (производится запрос на сервис индексации и в случае нахождения похожей ссылки проверяется, что дата публикации ссылки (при условии наличия такой даты) или дата прошлого запроса не старше одного года);
- в случае отсутствия искомой информации на веб-странице выполнение перехода на следующую доступную URL-ссылку, которая была найдена в HTML-коде или в массиве ссылок внешних веб-ресурсов, хранящихся в коде модуля.

Входным параметром алгоритма является начальная URL-ссылка внешнего веб-ресурса, которая была определена заранее и по которой в дальнейшем происходит переход на веб-страницу и запускается поиск. Выходным параметром является список ссылок на внешние источники информации.

Существенное ограничение использования метода заключается в зависимости времени работы сервиса поиска от количества информации на веб-ресурсе: чем больше данных хранится в исходном источнике, включая как пространственные и/или семантические данные, так и всё наполнение источника (текст веб-страниц, файлы на ftp-серверах, информация о перенаправлении на другие веб-ресурсы и т. д.), тем дольше будет работать поиск информации (Zheng et al., 2007). Важно отметить, что веб-скрапинг создаёт повышенную нагрузку на исходном ресурсе, вследствие чего возможны временные запреты на доступ к информации. К ограничениям стоит отнести и то, что сайты с динамическим отображением данных требуют более глубокого изучения и индивидуального подхода к считыванию с них информации. Также лимитирующим фактором этого метода являются юридические ограничения использования некоторых сайтов. В рамках планетных исследований юридические аспекты несущественны, поскольку используемые веб-ресурсы не имеют указанных ограничений на сбор информации.

Рис. 5. Схема работы модуля веб-скрапинга (индексации) внешних веб-ресурсов

На основе предложенного алгоритма разработан модуль веб-скрапинга (индексации) внешних веб-ресурсов, схема работы которого приведена на *рис. 5* (см. с. 25). При работе модуля используется СУБД PostgreSQL (<https://www.postgresql.org/>), в которой хранятся индексы веб-страниц исходных источников (полученных в ходе работы модуля поиска информации с помощью API), отсортированных по каким-либо условиям (например, по исследуемому небесному телу, типу искомого объекта, типу данных и т. п.). Модуль выполняет регулярное обновление индексов для поддержания их актуальности.

При работе модуля используются несколько общедоступных библиотек. Библиотека BeautifulSoup (<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>) является основой при сборе информации. Она позволяет фильтровать, искать и анализировать данные HTML-кода страниц. Совместное использование BeautifulSoup и библиотеки lxml (<http://lxml.de/>) позволяет эффективно и оптимально отсортировать и обработать содержимое веб-страниц. Обеспечением доступа к страницам исходного веб-ресурса, а также проверкой доступности самого источника занимаются библиотеки requests (<http://docs.python-requests.org/en/master/>), Scrapy (<https://scrapy.org/>) и urllib2 (<https://docs.python.org/2/library/urllib2.html>). Библиотека pycorg2 (<http://initd.org/pycorg2/>) позволяет организовать доступ к сервису индексации для СУБД PostgreSQL и обновление её служебных таблиц (см. *рис. 5*).

Тестирование созданного веб-сервиса поиска планетной информации

При реализации работы веб-сервиса поиска планетной информации подготовлено несколько макетов электронных карт небесных тел. Они содержат набор тематических слоёв, настроенных для интернет-доступа, и опубликованы с помощью сервиса ArcGIS-online. Онлайн-карты представляют результаты научных исследований поверхности Ганимеда (спутника Юпитера) (Лазарева, Матвеев, 2016; Zubarev et al., 2015) и Меркурия (Жаркова и др., 2016), выполненных с помощью обработки данных дистанционного зондирования:

- детальные онлайн-карты на отдельные исследуемые участки Ганимеда (<http://arcgis/2gRx1yv>; http://bit.ly/Ganymed_local_morphologmap2);
- глобальная онлайн-карта Ганимеда (<http://bit.ly/2FhTwpR>);
- онлайн-карта северного полушария Меркурия, включая каталог малых плоскодонных кратеров и снимки высокого разрешения, полученные космическим аппаратом MESSENGER (MErcury Surface, Space ENvironment, GEochemistry and Ranging) (<http://arcgis/2gW6zAe>);
- глобальная онлайн-карта поверхности Меркурия, включающая статистические характеристики морфометрических параметров рельефа (<http://arcgis/2mCqppN>).

Поиск планетной информации с помощью веб-сервиса, внедрённого в онлайн-карты небесных тел, осуществляется с использованием различных интернет-ресурсов, отобранных при анализе исходных источников, в том числе таких как:

- система планетных данных НАСА (PDS);
- архив Японского агентства аэрокосмических исследований;
- Европейское космическое агентство;
- Астрогеологический научный центр Американской геологической службы (USGS — United States Geological Survey);
- база данных планетной номенклатуры Международного астрономического союза;
- международные издательства Elsevier, Springer, EDP Sciences;
- библиографическая и реферативная база данных научных изданий Scopus;
- службы новостей космических агентств, включая «Роскосмос», и зарубежные агентства;
- поисковая система полнотекстовых научных публикаций Google.Академия.

Рис. 6. Результат поиска информации на внешних веб-ресурсах с помощью модуля на основе API, внедрённого в онлайн-карту Ганимеда

Рис. 7. Результат поиска информации на внешних веб-ресурсах с помощью онлайн-карты Ганимеда на основе модуля, интегрирующего методы API и веб-скрапинга

При проведении сравнительных тестов на качество и достоверность работы созданного веб-сервиса были использованы две версии онлайн-карты Ганимеда на один из исследуемых участков (http://bit.ly/Ganymed_local_morphologmap2) с внедрением разработанных модулей: один вариант карты обеспечивал поиск с использованием только API, а другой — с комбинированием API и веб-скрапинга. При этом модуль веб-скрапинга настроен на поиск самой современной на текущий период информации. Тестирование показало следующие результаты:

- обе версии веб-сервиса обеспечивают быстрый поиск информации: использование API обеспечивает результат в среднем за 80–110 мс, тогда как поиск комбинированным методом осуществляется с небольшой задержкой, практически не заметной для пользователя, и занимает от 120 до 135 мс;

- при поиске с использованием API пользователю онлайн-карты выдаётся вся информация, найденная за любой период времени по всем источникам. Комбинация методов обеспечивает возможность доступа к новейшей информации, причём той, которая наиболее полно удовлетворяет условиям запроса;
- при поиске с использованием API выдаётся большее количество данных. Например, при запросе информации о геоморфологических зонах Ганимеда, помимо других интернет-источников, найдено 11 веб-страниц на сайте Астрогеологического научного центра Американской геологической службы (рис. 6, см. с. 27). При комбинировании методов по такому же запросу только на одном веб-ресурсе найдена единственная ссылка, наиболее близко соответствующая требованиям запроса: геологическая карта Ганимеда (рис. 7, см. с. 27).

Таким образом, поиск с помощью API обеспечивает широту охвата, не пропуская ни один из веб-ресурсов, на которых может быть найдена любая информация об объекте на карте, тогда как комбинированный метод обеспечивает оптимизацию поиска, выбирая из множества результатов лишь тот, который наиболее близок условиям запроса, отсекая всю другую информацию.

Заключение

Актуальность предложенного веб-сервиса связана с необходимостью удобного поиска и быстрого доступа к планетной информации, которая может быть представлена в картографическом контексте. К отличительным особенностям сервиса можно отнести то, что он обрабатывает и предоставляет информацию быстро, с достаточной полнотой и не требует гигантских вычислительных мощностей, как альтернативные сервисы Google или Яндекса (из-за конечного списка исследуемых ресурсов). Разработанный веб-сервис, основанный на комбинированном использовании методов API и сбора информации в сети интернет с помощью тематического индексирования (веб-скрапинг), внедрён в интерфейс планетных онлайн-карт Ганимеда и Меркурия, которые доступны для широкого круга пользователей.

Метод веб-скрапинга для поиска планетной информации в интернете позволил повысить качество искомых данных, расширить список поставщиков информации, обеспечить поиск на внешних ресурсах, которые не предоставляют ни функционала API, ни приемлемого качества ответов на POST-запросы, существенно улучшив качество ответов на запросы пользователя по нескольким причинам:

- совместное использование веб-скрапинга с API веб-ресурсов обеспечивает уточнение найденной информации, поскольку при использовании API выдача результатов может быть неоднозначной или оказаться «пустой», несмотря на то, что исходные источники всё-таки содержат полезные данные;
- обеспечена возможность фильтрации найденной информации с помощью анализа и машинного обучения на основе применения статистических данных из результатов предыдущих запросов, привлечения дополнительной информации о типе искомых пространственных объектов и их атрибутах;
- результаты запроса ранжированы в окне выдачи пользовательского интерфейса онлайн-карты.

Созданный специализированный веб-сервис, обеспечивающий интеллектуальный поиск планетной информации с использованием пространственного и семантического контекста, будет применяться при публикации создаваемых нами онлайн-карт небесных тел. В дальнейшем предполагается использовать созданные поисковые возможности для Геопортала планетных данных Московского государственного университета геодезии и картографии (<http://cartsrv.mexlab.ru/geoportal>).

Работа выполнена при поддержке Российского фонда фундаментальных исследований по теме «Разработка веб-сервиса для интеллектуального поиска информации о планетных данных на основе пространственного и семантического контекста при онлайн-картографировании» (проект № 16-37-00316 мол_а).

Литература

1. Жаркова А. Ю., Матвеев Е. В., Коханов А. А. Онлайн картографирование Меркурия для представления результатов морфометрических исследований // 13-я конф. молодых ученых: сб. тез. Москва. 13–15 апр. 2016. М: ИКИ РАН, 2016. С. 34.
2. Лазарева М. С., Матвеев Е. В. Разработка крупномасштабных геоморфологических карт Ганимеда для онлайн-картографирования // Всерос. науч. конф. «Международный год карт в России: объединяя пространство и время»: сб. тез. Москва, 25–28 окт. 2016. М.: Изд-во Геогр. фак. Московского гос. ун-та имени М. В. Ломоносова, 2016. С. 181–184.
3. Матвеев Е. В., Брусникин Е. С., Жаркова А. Ю. Анализ методов организации системы интеллектуального поиска семантической и пространственной планетной информации // 13-я конф. молодых ученых: сб. тез. Москва. 13–15 апр. 2016. М: ИКИ РАН, 2016. С. 56.
4. Митчел Р. Скрапинг веб-сайтов с помощью Python / пер. с англ. А. В. Груздев. М.: ДМК Пресс, 2016. 280 с.
5. Clarke S. Measuring API Usability // Dr. Dobb's J. 2004. URL: <http://www.drdoobs.com/windows/measuring-api-usability/184405654>.
6. Ferreira C. Designing Neural Networks Using Gene Expression Programming / eds. A. Abraham, B. de Baets, M. Köppen, B. Nickolay // Applied Soft Computing Technologies: The Challenge of Complexity. Springer-Verlag, 2006. P. 517–536.
7. Graves A., Wayne G., Danihelka I. Neural Turing Machine // arXiv: e-print service. arXiv:1410.5401. 2014. 26 p. URL: <https://arxiv.org/pdf/1410.5401.pdf>.
8. Hare T. M., Rossi A. P., Frigeri A., Marmo Ch. Interoperability in Planetary Research for Geospatial Data Analysis // Planetary and Space Science. V. 150. 2017. P. 36–42. DOI: 10.1016/j.pss.2017.04.004.
9. Zheng S., Wu D., Song R. Joint Optimization of Wrapper Generation and Template Detection // Proc. 13th Intern. Conf. Knowledge Discovery and Data Mining. San Jose, California, USA, 12–14 Aug. 2007. 2007. P. 894–902. URL: <https://pdfs.semanticscholar.org/4fb4/3c5a212df751e84c3b2f8d29fabfe56c3616.pdf>.
10. Zubarev A. E., Nadezhkina I. E., Brusnikin E. S., Kokhanov A. A., Lazareva M. S., Karachevtseva I. P., Oberst J. Studies of the basic geodetic parameters and mapping of Galilean satellites // ISPRS Working Group IV/8 “Planetary Mapping and Spatial Databases” meeting. Berlin, Germany. 24–25 Sept. 2015. URL: http://www.dlr.de/pf/Portaldata/6/Resources/dokumente/abt_pd/isprs/berlin_2015/Zubarev_ISPRS_IV8_2015.pdf.

Web-service of intelligent search for planetary data based on the spatial and semantic context on online maps of celestial bodies

E. V. Matveev¹, A. Yu. Zharkova^{1,2}, I. P. Karachevtseva¹

¹ Moscow State University of Geodesy and Cartography, Moscow 105064, Russia
E-mail: e_matveev@miigaik.ru

² Sternberg Astronomical Institute, Moscow State University, Moscow 119234, Russia
E-mail: a_zharkova@miigaik.ru

We present the results of the development of a specialized web service providing intelligent search for planetary information. The search service was created as part of development of the information system for celestial bodies' space survey and for access to results by geoportals. The creation of special tools for the intellectualization of search services is caused by the huge volume and diversity of information

accumulated in various scientific fields, and in particular, in planetary research. During the development of this service, digital archives of planetary data and software that provide access to the stored information were analyzed. Our results show that the use of implementation of thematic indexing of Internet sources in addition to use of website functions (API), helps improve access to the data providing optimized search. The created web-service for search of planetary data is integrated with on-line maps of celestial bodies by ArcGIS tools and provides data search taking into account the spatial and semantic content presented on the maps. We plan to use the developed search facilities for the MII GAIK Planetary Data Geoportal (<http://cartsrv.mexlab.ru/geoportal>) in the future.

Keywords: web-scraping, intelligent search, planetary data, GIS, online maps of celestial bodies

Accepted: 10.07.2019

DOI: 10.21046/2070-7401-2019-16-4-19-30

References

1. Zharkova A. Yu., Matveev E. V., Kokhanov A. A., Onlain kartografirovaniye Merkuriya dlya predstavleniya rezul'tatov morfometricheskikh issledovaniy (Web-mapping of Mercury for the presentation of the results of morphometric studies), *13-ya konferentsiya molodykh uchenykh* (13th Conf. Young Scientists), Book of abstracts, Moscow, 13–15 Apr., 2016, Moscow: IKI RAN, 2016, p. 34.
2. Lazareva M. S., Matveev E. V., Razrabotka krupnomasshtabnykh geomorfologicheskikh kart Ganimeda dlya onlain-kartografirovaniya (Development of large-scale geomorphological maps of Ganymede for web mapping), *Vserossiyskaya nauchnaya konferentsiya "Mezhdunarodnyi god kart v Rossii: ob"edinyaya prostranstvo i vremya"* (All-Russia Scientific Conf. "International map year in Russia: uniting space and time"), Book of abstracts, Moscow, 25–28 Oct., 2016, Moscow, 2016, pp. 181–184.
3. Matveev E. V., Brusnikin E. S., Zharkova A. Yu., Analiz metodov organizatsii sistemy intellektual'nogo poiska semanticheskoi i prostranstvennoi planetnoi informatsii (Analysis of methods for organizing an intelligent search for semantic and spatial planetary information), *13-ya konferentsiya molodykh uchenykh* (13th Conf. Young Scientists), Book of abstracts, Moscow, 13–15 Apr., 2016, Moscow: IKI RAN, 2016, p. 56.
4. Mitchel R., *Skraping veb-saitov s pomoshch'yu Python* (Scraping for websites by Python), Moscow: DMK Press, 2016, 280 p.
5. Clarke S., Measuring API Usability, *Dr. Dobb's J.*, 2004, URL: <http://www.drdoobs.com/windows/measuring-api-usability/184405654>.
6. Ferreira C., Designing Neural Networks Using Gene Expression Programming, In: *Applied Soft Computing Technologies: The Challenge of Complexity*, A. Abraham, B. de Baets, M. Köppen, B. Nickolay (eds.), Springer-Verlag, 2006, pp. 517–536.
7. Graves A., Wayne G., Danihelka I., Neural Turing Machines, *arXiv:1410.5401*, 2014, 26 p., available at: <https://arxiv.org/pdf/1410.5401.pdf>.
8. Hare T. M., Rossi A. P., Frigeri A., Marmo Ch., Interoperability in Planetary Research for Geospatial Data Analysis, *Planetary and Space Science*, Vol. 150, 2017, pp. 36–42, DOI: 10.1016/j.pss.2017.04.004.
9. Zheng S., Wu D., Song R., Joint Optimization of Wrapper Generation and Template Detection, *Proc. 13th Intern. Conf. Knowledge Discovery and Data Mining*, San Jose, California, USA, 12–14 Aug., 2007, 2007, pp. 894–902, available at: <https://pdfs.semanticscholar.org/4fb4/3c5a212df751e84c3b2f8d29fabfe56c3616.pdf>.
10. Zubarev A. E., Nadezhkina I. E., Brusnikin E. S., Kokhanov A. A., Lazareva M. S., Karachevtseva I. P., Oberst J., Studies of the basic geodetic parameters and mapping of Galilean satellites, *ISPRS Working Group IV/8 "Planetary Mapping and Spatial Databases" meeting*, 24–25 Sept., 2015, Berlin, Germany, available at: http://www.dlr.de/pf/Portaldata/6/Resources/dokumente/abt_pd/isprs/berlin_2015/Zubarev_ISPRS_IV8_2015.pdf.