

УДК 004.85

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СИСТЕМ, ПРЕДОСТАВЛЯЮЩИХ ОТКРЫТЫЕ ДАННЫЕ

Бородин А. Ф.,

студент 2 курса магистратуры

Московский государственный технический университет им. Н. Э. Баумана

Россия, г. Москва

Аннотация

В данной статье производится сравнительный анализ систем, предоставляющих открытые данные на основе выработанной системы критериев.

Ключевые слова: открытые данные, портал, API

COMPARATIVE ANALYSIS OF OPEN DATA SYSTEMS

Borodin A. F.,

master student,

Bauman Moscow State Technical University

Russian Federation, Moscow

Annotation

In this article the comparative analysis of systems provides open data on the basis of the developed system of criteria.

Keywords: open data, site, API

Открытые данные – концепция, отражающая идею о том, что определённые данные должны быть свободно доступны для машиночитаемого использования и дальнейшей повторной публикации без ограничений

авторского права, патентов и других механизмов контроля [1]. Освободить данные от ограничений авторского права можно с помощью свободных лицензий, таких как лицензий Creative Commons. Если какой-либо набор данных не является общественным достоянием, либо не связан лицензией, дающей права на свободное повторное использование, то такой набор данных не считается открытым, даже если он выложен в машиночитаемом виде в Интернет. В сети интернет опубликовано 500 млрд. фактов, которые можно использовать для решения различных классов задач. Существует множество глобальных производственных и научных сообществ. Один из примеров – “DataDock of UK”, в Великобритании. Европейский Союз планирует открыть большой портал открытых данных.

Формат открытых наборов данных – то, как данные структурируются и становятся доступными для людей и машин. Выбор правильного формата позволяет гарантировать, что данные могут быть распознаны и использованы повторно. Для увеличения способов повторного использования данных, разработчикам может понадобиться множество форматов и структур, доступных на разных платформах, в соответствии с потребностями пользователей.

Не все данные могут быть соответствующим образом выражены в виде таблиц. Существует три ключевые структуры представления открытых данных: табличные, иерархические и сетевые. Наиболее распространенная структура данных – табличная. Данные упорядочиваются в строки и столбцы, перечисляя последовательные значения. Если данные основаны на отдельных записях, которые не связаны друг с другом, то табличная структура данных чаще всего предоставляется в формате CSV. Иерархические данные показывают взаимосвязь между минимальными фреймами данных, такими как семейное древо или муниципалитеты в каждой стране. Если набор данных зависит от отношения между точками данных и следует структуре, в которой точки данных связаны в вертикальных «деревьях», иерархическая структура данных наиболее структурировано описывается в формате JSON. Сетевые структурированные

дневник науки | www.dnevniknauki.ru | СВЯЗИ С НАМИ ПОСРЕДСТВОМ ТЕЛЕФОНА 8 (800) 100-8321

данные позволяют поддерживать отношения между любой комбинацией элементов в любых направлениях. Хорошим примером сетевой структуры данных является социальная сеть. Веб – еще один пример сетевой структуры данных, где веб-страницы ссылаются на любое количество других страниц в любом направлении. При этом данные также предоставляются в виде дампов таблиц баз данных, открытых API-интерфейсов, образов приложений, запущенных внутри docker-контейнеров и множеством других способов. Таким образом, можно выделить критерий сравнения источников открытых данных – интерфейс и формат предоставления данных.

Наряду с этим критерием так же можно выделить такой критерий, как наличие и вид документации по источнику данных. Документация предоставляет пользователям и разработчикам информацию обо всех возможностях платформы открытых данных, предлагает алгоритмы получения нужной выборки данных и многое другое. Существует несколько способов создания документации. В одном случае документация составляется разработчиками, создающими системы открытых данных, в другом случае документация составляется компаниями, предоставляющими открытые данные. Также существуют системы, являющиеся самодокументируемыми, например GraphQL. Сервер GraphQL предоставляет интерфейс к получению данных, подобный REST-интерфейсам и предлагает самодокументируемый сервис, описывающий данные, их декомпозицию на объекты и методы их получения. Открытые данные собираются множеством различных способов на протяжении конечных промежутков времени. При сборе такой информации, как, например, ежемесячная сводная статистика в части жилищно-коммунальных услуг по городу Москве, предоставляемой порталом открытых данных data.mos.ru актуальность, является ключевым вопросом. Актуальные и обновляемые данные позволяют как выяснять корреляцию между различными параметрами, так и осуществлять прогнозирование. Следовательно, можно выделить такой критерий сравнения как актуальность данных.

Для сравнения были выбраны крупнейшие источники открытых данных. В их число входят:

- <https://data.mos.ru/> – Портал открытых данных правительства Москвы [2].
- <http://data.gov/> – Портал открытых данных правительства Соединенных Штатов Америки [3].
- <http://data.gov.uk/> – Портал открытых данных правительства Великобритании [4].
- <http://aws.amazon.com/datasets> – Портал открытых данных компании Amazon [5].

Согласно составленной системе критериев информации, можно составить таблицу 1.

Таблица 1 – Сравнительная таблица источников открытых данных

Источник открытых данных \ Критерий сравнения	Интерфейс и формат данных	Наличие и вид документации	Актуальность данных
Портал открытых данных правительства Москвы	csv, json, xml и интерактивные карты	+	+
Портал открытых данных правительства США	csv, xlsx, json	+	+
Портал открытых данных правительства Великобритании	txt	+	+
Портал открытых данных компании Amazon	Снимки баз данных	-	-

Согласно таблице 1 можно сделать следующие выводы:

- Портал открытых данных правительства Москвы предоставляет пользователям и разработчикам наиболее широкий спектр инструментов по работе с открытыми данными, включая интерактивные карты.
- Портал открытых данных компании Amazon удовлетворяет наименьшему числу критериев сравнения открытых данных. При этом, на этом портале данные представлены в формате снимков баз данных, что позволяет использовать их без дополнительной настройки и обработки.
- Портал открытых данных правительства Великобритании предлагает разработчикам и пользователям данные в формате txt-файлов, что требует дополнительной обработки и распознавания.

Библиографический список

1. Открытые данные [Электронный ресурс]. – Режим доступа – https://ru.wikipedia.org/wiki/Открытые_данные (Дата обращения 24.02.2019)
2. Портал открытых данных правительства Москвы [Электронный ресурс]. – Режим доступа – URL: <https://data.mos.ru/> (Дата обращения 10.03.2019)
3. Портал открытых данных правительства США [Электронный ресурс]. – Режим доступа – URL: <https://data.gov/> (Дата обращения 10.03.2019)
4. Портал открытых данных правительства Великобритании [Электронный ресурс]. – Режим доступа – URL: <https://data.gov.uk/> (Дата обращения 10.03.2019)
5. Портал открытых данных компании Amazon [Электронный ресурс]. – Режим доступа – URL: <http://aws.amazon.com/datasets> (Дата обращения 24.04.2019)

Оригинальность 84%