

МЕТОД АГРЕГАЦИИ И НОРМАЛИЗАЦИИ ДАННЫХ ГЕТЕРОГЕННОЙ СЕНСОРНОЙ СЕТИ

М. А. Колчин, И. А. Шилин, Н. В. Климов,
Д. С. ГАРАЙЗУЕВ, Д. И. МУРОМЦЕВ, Д. А. ЗАКОЛДАЕВ

*Университет ИТМО, 197101, Санкт-Петербург, Россия
E-mail: kolchinmax@niuitmo.ru*

Представлен метод агрегации данных сенсорной сети, включающей в себя электронные устройства, работающие на основе различных протоколов и моделей данных. Метод заключается в применении модульной архитектуры программной системы и верхнеуровневых онтологий. Описаны архитектура программной системы и инструмент для полуавтоматической генерации RDF-описаний электронных устройств.

Ключевые слова: сенсорные сети, Интернет вещей, онтологический инжиниринг, технологии семантического веба, архитектура OSGi, верхнеуровневые онтологии, интеграция данных.

Введение. В настоящее время повышается доступность электронных устройств с ограниченными вычислительными ресурсами, в которых реализована поддержка передачи данных по беспроводным или проводным сетям. Кроме того, появление новых протоколов идентификации и передачи данных, таких как IPv6, RFID, 6LoWPAN, ZigBee и т.д., привело к формированию и развитию концепции Интернета вещей [1]. Такие электронные устройства применяются для создания огромного количества сенсорных сетей и внедрения их в различных областях производства и жизнедеятельности человека, например, таких как сельское хозяйство, пищевая и перерабатывающая промышленность, экологический мониторинг, системы безопасности, энергоснабжение и многих других [2—5]. Использование технологий Интернета вещей обеспечивает возможность получения городскими и муниципальными органами информации о состоянии окружающей среды (воздух и вода) [6] и различных элементов городской инфраструктуры, что позволяет прогнозировать пути развития города и принимать обоснованные решения по их реализации.

К данному моменту существует большое количество разработок в области обеспечения взаимодействия на коммуникационном уровне электронных устройств с ограниченными вычислительными ресурсами, а также их взаимодействия с программными системами, однако на прикладном уровне, где объединение различных устройств от разных производителей образует небольшие и обособленные кластеры со своим набором форматов, моделей и стандартов, по-прежнему существуют нерешенные задачи. Одной из задач является отсутствие механизмов обеспечения семантической интероперабельности (от англ. interoperability — способность системы к взаимодействию) на уровне моделей и понятий, что затрудняет разработку информационных систем, использующих разные типы и модели устройств.

В настоящей статье рассматривается метод решения проблемы агрегации данных сенсорной сети, включающей в себя электронные устройства, работающие с использованием различных протоколов передачи данных и разнородных структур данных. Решение состоит из двух шагов: 1) агрегация данных, которая подразумевает применение модульной архитектуры системы, 2) нормализация данных, которая основана на применении верхнеуровневых онтологий.

Агрегация данных гетерогенной сенсорной сети. Текущей проблемой сенсорных сетей является многообразие программного и аппаратного обеспечения, используемого для их

построения. Эта проблема решается путем построения модульной архитектуры программной системы для сбора данных сенсорных сетей. Модульная архитектура содержит модули, каждый из которых реализует алгоритм сбора данных для определенного типа электронных устройств. Такая программная система, пример которой описан в работе [7], позволяет декларативным способом (с использованием XML-файлов) конфигурировать процесс сбора данных, поступающих с различных устройств. Однако данная система поддерживает только ограниченный набор типов устройств, так как для реализации алгоритма работы с новым типом устройства необходимо перепрограммирование всей системы. В отличие от существующих, предлагаемый в настоящей статье метод агрегации данных сенсорной сети позволяет реализовать поддержку новых типов устройств в виде самостоятельного модуля, который подключается к системе без необходимости ее перепрограммирования или даже остановки.

Метод агрегации данных гетерогенной сенсорной сети заключается в использовании модульной архитектуры, основанной на стандарте OSGi*. На рис. 1 представлена диаграмма архитектуры программной системы SemIoT Platform, которая реализует данный метод. Система состоит из 5 модулей, 2 баз данных и набора драйверов устройств:

Drivers — набор драйверов *A—D*, реализующих алгоритмы сбора данных с использованием определенных протоколов передачи данных и моделей данных;

Driver Manager — модуль, отвечающий за контроль жизненного цикла электронного устройства; предоставляет драйверам устройств программный интерфейс для регистрации, изменения статуса соединения и т.д.;

Configuration Manager — модуль, отвечающий за конфигурирование драйверов устройств;

Metadata Writer — модуль, предоставляющий программный интерфейс для записи метаданных (например, местоположения, времени установки, измеряемых показаний и т.д.) электронных устройств в Triplestore — специализированную базу данных для хранения RDF**-данных;

Observation Writer — модуль, предоставляющий программный интерфейс для записи метаданных электронных устройств в Time Series Database*** (TSBD) — специализированную базу данных для хранения данных, привязанных ко времени (например, измерение температуры воздуха в некоторый момент времени); в качестве такой базы данных в программной системе SemIoT Platform используется база данных OpenTSDB****;

Hydra/REST API — REST-подобный программный интерфейс доступа к данным электронных устройств: метаданным устройства и его показаниям; Hydra — это стандарт проектирования REST-интерфейсов, использующий верхнеуровневые онтологии для его аннотирования и не требующий дополнительного программирования (подробнее об этом стандарте см. работу [8]).

Драйвер устройства — это файл в формате JAR (Java Archive), содержащий его исходный код и MANIFEST.MF-файл, в котором прописано название драйвера, его версия и полное имя класса, реализующего специальный интерфейс *Activator*, описанный в стандарте OSGi. На рис. 2 представлены интерфейсы классов, используемые для работы драйвера устройства. Каждый из драйверов *A—D* реализует интерфейс *Driver*, и модуль *Driver Manager* реализует соответствующий интерфейс, который используется для регистрации устройства, изменения его статуса, публикации показаний и т.д.

* OSGi — стандарт построения модульной архитектуры программного обеспечения, работающей на платформе Java Virtual Machine, подробнее см.: <<http://www.osgi.org>>.

** Resource Description Framework, см.: <<http://www.w3.org/TR/rdf-primer>>.

*** Time Series Database, см.: <https://en.wikipedia.org/wiki/Time_series_database>.

**** OpenTSDB, см.: <<http://opentsdb.org>>.

Один драйвер используется для работы с несколькими устройствами, каждое из устройств представлено с помощью класса *Device*, а его показания — с помощью класса *Observation*. Особенностью архитектуры OSGi является возможность загрузки и выгрузки любого из модулей без остановки всей системы, соответственно драйвер устройства можно конфигурировать и подключать в любое время ее работы.

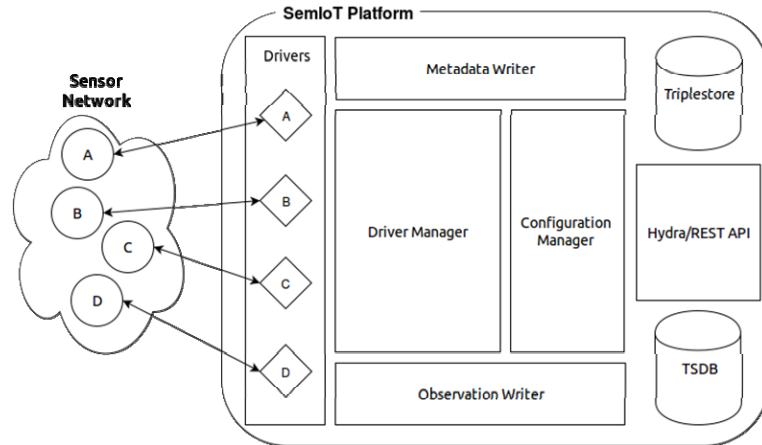


Рис. 1

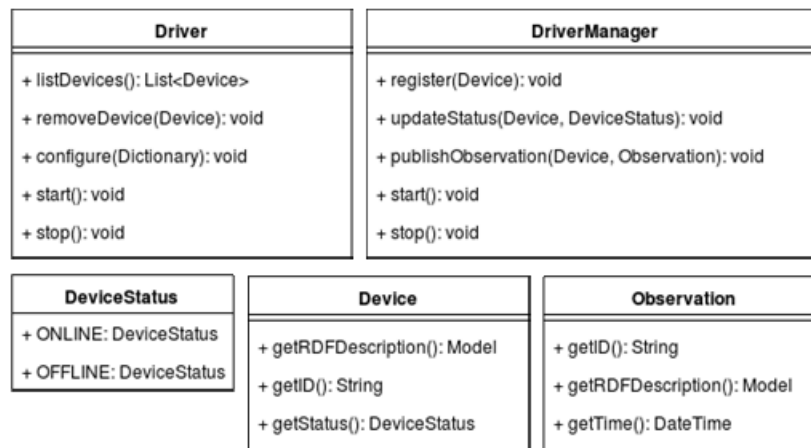


Рис. 2

Метаданные устройства и показаний представляются в модели данных RDF с использованием верхнеуровневых онтологий, что позволяет преобразовать модели и структуры данных, используемые различными электронными устройствами, в универсальную модель.

Нормализация данных гетерогенной сенсорной сети. Если с помощью описанного выше подхода можно решить проблему агрегации данных гетерогенной сенсорной сети на коммуникационном уровне, то для обеспечения семантической интероперабельности на уровне модели и понятий предлагается использовать верхнеуровневые OWL-онтологии [9]. OWL-онтологии предназначены для декларативного описания концептов (понятий) предметной области на основе логики первого порядка и дискрипционной логики.

В области сенсорных сетей уже существуют работы по созданию верхнеуровневых онтологий, которые позволяют описать базовые концепты данной предметной области. В 2011 г. рабочей группой W3C Semantic Sensor Network Incubator Group была представлена верхнеуровневая онтология для описания таких концептов, как *сенсор*, *показание*, *точность показания*, *рабочие условия сенсора*, *предельные показания сенсора* и т.д. Эта онтология получила название Semantic Sensor Network Ontology (SSN) [10].

Метод нормализации данных гетерогенной сенсорной сети основан на использовании онтологии SSN как одного из элементов архитектуры программной системы. В отличие от существующих подходов, которые исследуются в работе [11], данная онтология служит инструментом не только для публикации данных сенсорной сети, который позволяет реализовать бесшовную интеграцию информационных систем, но и для хранения показаний сенсоров.

Структура данных, используемая в OpenTSDB, не поддерживает возможность записи RDF-данных, поэтому был разработан модуль *Observation Writer*, предназначенный для преобразования метаданных показаний в структуру OpenTSDB. Данное преобразование основано на том, что все показания сенсоров аннотированы с помощью онтологии SSN или других онтологий, которые расширяют ее. Концепты онтологии SSN используются как „метки“ для выделения из набора триплов, описывающих какое-либо показание, тех, в которых записана информация для хранения в OpenTSDB. Если же показания аннотированы с помощью других онтологий, то для преобразования используется дополнительный шаг, на котором выполняется логический вывод (reasoning). На рис. 3 представлен пример метаданных показания сенсора температуры окружающей среды. В этом примере используется онтология, расширяющая онтологию SSN путем введения следующих концептов:

ex:AirTemperatureObservation (является подклассом *ssn:Observation*),
ex:AirTemperatureSensorOutput (подкласс — *ssn:SensorOutput*),
ex:AirTemperatureValue (подкласс — *ssn:ObservationValue*),
ex:hasQuantityValue.

```
@prefix ex: <http://example.org/ontology#> .
@prefix ssn: <http://purl.oclc.org/NET/ssnx/ssn#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

:1434916797482 a ex:AirTemperatureObservation ;
  ssn:observationResultTime
'2015-02-05T13:15:30Z'^^^xsd:dateTime ;
  ssn:observedBy <coap://10.0.0.7:60000/sensor> ;
  ssn:observationResult :1434916797482-result .
:1434916797482-result a ex:AirTemperatureSensorOutput ;
  ssn:isProducedBy <coap://10.0.0.7:60000/sensor> ;
  ssn:hasValue :1434916797482-resultvalue .
:1434916797482-resultvalue a ex:AirTemperatureValue ;
  ex:hasQuantityValue '0.636'^^^xsd:float .
```

Рис. 3

После того как аннотация показания приведена с помощью логического вывода к концептам онтологии SSN, происходит преобразование расширенной аннотации в структуру данных OpenTSDB на основе следующих правил:

— в поле *metric_uid* записывается строка, являющаяся результатом замены знаков „/“ и „:“ (слэш и двоеточие) на знак „.“ (точка) в URI сенсора, заданном свойством *ssn:observedBy*: например, *coap://10.0.0.7:60000/sensor* преобразуется в *10.0.0.7.60000.sensor*;

— в поле *timestamp* записываются дата и время регистрации показания в системе, которые задаются свойством *ssn:observationResultTime* и преобразуются в формат POSIX-времени*: например, *'2015-02-05T13:15:30Z'* преобразуется в *1423142130*;

— если присутствует свойство *ex:hasQuantityValue*, тогда его значение записывается в поле *value*; иначе, в тег с именем *value* записывается полный URI класса, являющегося подклассом *ssn:ObservationValue*;

— в тег с именем *type* записывается полный URI класса, являющегося подклассом *ssn:Observation*.

* POSIX-время, см. <<https://ru.wikipedia.org/wiki/UNIX-%D0%B2%D1%80%D0%B5%D0%BC%D1%8F>>.

Например, для показания, приведенного на рис. 3, RDF-данные будут преобразованы в следующий вид:

```
metric_uid = 10.0.0.7.60000.sensor, timestamp = 1434916797482, value = 0.636, type =  
= http://example.org/ontology#AirTemperatureObservation.
```

В работе [12] приведены результаты исследования существующих онтологий для представления концептов сенсорной сети, где показано, что несмотря на большое количество верхнеуровневых онтологий для этой области качество их публикации и отсутствие инструментов автоматической или полуавтоматической генерации RDF-описаний для электронных устройств затрудняет их широкое применение. Поэтому в рамках данного исследования было разработано веб-приложение WoT SemDesc Helper для полуавтоматической генерации таких RDF-описаний. На рис. 4 представлен скриншот его пользовательского интерфейса. WoT SemDesc Helper предоставляет набор веб-форм, используя которые, пользователь вводит необходимую информацию о своем электронном устройстве, и в результате заполнения этих форм генерируется соответствующее RDF-описание.

Рис. 4

Веб-приложение разделено на три части: генератор RDF-описания для модели устройства, генератор RDF-описания конкретного устройства на базе существующей модели и каталоги моделей устройств. Каждая модель устройства имеет свой URI, который позволяет встраивать в устройство не полное его описание, а краткую версию с указанием этого URI. Впоследствии по данному URI может быть получено более полное описание из каталога WoT SemDesc Helper.

Заключение. Предложенный метод агрегации и нормализации данных гетерогенной сенсорной сети основан на применении элементов архитектуры OSGi и верхнеуровневых онтологий, цель которых предоставить универсальную модель и структуру данных. Описаны архитектура программной системы SemIoT Platform, в которой применяется данный метод, и инструмент для полуавтоматической генерации RDF-описаний электронных устройств.

Исходные коды программной системы SemIoT Platform и веб-приложения WoT SemDesc Helper являются открытыми и опубликованы в сети Интернет: <http://github.com/semiotproject/semiot-platform> и <https://github.com/semiotproject/wot-semdesc-helper> соответственно.

Статья подготовлена по результатам работы, выполненной при поддержке Министерства образования и науки Российской Федерации, грант № RFMEFI57514X0101.

СПИСОК ЛИТЕРАТУРЫ

1. Ashton K. The Internet of things // RFID Journal. 2009. N 50.
2. Atzori L. et al. The Internet of things: A survey // Computer Networks. 2010. Vol. 54, N 15. P. 2787—2805.
3. Liu D. L., Zhu X. B., Xu K. L., Fang D. M. The design of modern agriculture control system based on Internet of things // Applied Mechanics and Materials. 2014. Vol. 517. P. 1519—1522.
4. Bainbridge S., Steinberg C., Furnas M. GBROOS—an ocean observing system for the Great Barrier Reef // Proc. of the Intern. Coral Reef Symposium. 2010. P. 529—533.
5. Колчин М. А., Фенцель А., Муромцев Д. И., Попов С. О., Павлов Д. С., Климов Н. В., Андреев А. А., Гарайзуев Д. С. Мониторинг потребления энергии в умных сетях электроснабжения (Smart grid) на основе семантического анализа потоковых данных // Научно-технический вестник информационных технологий, механики и оптики. 2015. № 2 (96). С. 285—292.
6. Oliveira L., Rodrigues J. Wireless sensor networks: A survey on environmental monitoring // J. of Communications. 2011. Vol. 6, N 2. P. 143—151.
7. Aberer K., Hauswirth M., Salehi A. A middleware for fast and flexible sensor network deployment // Proc. of the 32th Intern. Conf. on Very Large Data Bases, Seoul, Korea, 12—15 Sept. 2006. P. 1199—1202.
8. Lanthaler M., Guetl C. Hydra: A vocabulary for hypermedia-driven Web APIs // Proc. of the WWW2013 Workshop on Linked Data on the Web, Rio de Janeiro, Brazil, 14 May 2013.
9. Barnaghi P. et al. Semantics for the Internet of things // Intern. J. on Semantic Web & Information Systems. 2012. Vol. 8, N 1. P. 1—21.
10. Compton M., Barnaghi P., Bermudez L. et al. The SSN ontology of the W3C semantic sensor network incubator group // Web Semantics: Science, Services and Agents on the World Wide Web. 2012. Vol. 17. P. 25—32.
11. Noy N. Semantic integration: A survey of ontology-based approaches // SIGMOD Rec. 2004. Vol. 33, N. 4. P. 65—70.
12. Kolchin M., Klimov N., Andreev A., Shilin I., Garayzuev D., Mouromtsev D., Zakoldaev D. Ontologies for Web of things: a pragmatic review // Proc. of Knowledge Engineering and Semantic Web Conf., Moscow, Russia, 30 Sept. — 2 Oct. 2015.

Сведения об авторах

Максим Александрович Колчин	— Университет ИТМО; кафедра информатики и прикладной математики-1; ассистент; E-mail: kolchinmax@niuitmo.ru
Иван Андреевич Шилин	— Университет ИТМО; кафедра информатики и прикладной математики-1; инженер; E-mail: shilininivan@corp.ifmo.ru
Николай Васильевич Климов	— Университет ИТМО; кафедра информатики и прикладной математики-1; инженер; E-mail: nikolay.klimov@niuitmo.ru
Даниил Сергеевич Гарайзуев	— Университет ИТМО; кафедра информатики и прикладной математики-1; инженер; E-mail: garayzuev@corp.ifmo.ru
Дмитрий Ильич Муромцев	— канд. техн. наук, доцент; Университет ИТМО; кафедра информатики и прикладной математики-1; E-mail: mouromtsev@mail.ifmo.ru
Данил Анатольевич Заколдаев	— канд. техн. наук, доцент; Университет ИТМО; кафедра проектирования и безопасности компьютерных систем; E-mail: d.zakoldaev@mics.spb.ru

Рекомендована кафедрой
информатики и прикладной математики-1

Поступила в редакцию
31.08.15 г.

Ссылка для цитирования: Колчин М. А., Шилин И. А., Климов Н. В., Гарайзуев Д. С., Муромцев Д. И., Заколдаев Д. А. Метод агрегации и нормализации данных гетерогенной сенсорной сети // Изв. вузов. Приборостроение. 2015. Т. 58, № 11. С. 945—951.

**METHOD OF AGGREGATION AND NORMALIZATION
OF HETEROGENEOUS SENSOR NETWORK DATA****M. A. Kolchin, I. A. Shilin, N. V. Klimov, D. S. Garayzuev,
D. I. Muromtsev, D. A. Zakoldaev***ITMO University, 197101, St. Petersburg, Russia
E-mail: kolchinmax@niuitmo.ru*

A method is proposed for aggregation of sensor networks data that includes electronic devices, using different protocols and data models. The method is based on application of module architecture of the program system and upper ontologies. The program system architecture is presented; an instrument for semi-automatic generation of RDF- description of electronic devices is characterized.

Keywords: sensor networks, Internet of things, ontology engineering, semantic web technologies, OSGI architecture, upper ontologies, data integration.

Data on authors

- | | | |
|----------------------------|---|--|
| Maxim A. Kolchin | — | ITMO University, Department of Computer Science and Applied Mathematics-1; Engineer; E-mail: kolchinmax@niuitmo.ru |
| Ivan A. Shilin | — | ITMO University, Department of Computer Science and Applied Mathematics-1; Engineer; E-mail: shilininivan@corp.ifmo.ru |
| Nikolay V. Klimov | — | ITMO University, Department of Computer Science and Applied Mathematics-1; Engineer; E-mail: nikolay.klimov@niuitmo.ru |
| Daniil S. Garaysuev | — | ITMO University, Department of Computer Science and Applied Mathematics-1; Engineer; E-mail: garayzuev@corp.ifmo.ru |
| Dmitry I. Muromtsev | — | PhD, Associate Professor; ITMO University, Department of Computer Science and Applied Mathematics-1; E-mail: d.muromtsev@gmail.com |
| Danil A. Zakoldaev | — | PhD, Associate Professor; ITMO University, Department of Computer System Design and Security; E-mail: d.zakoldaev@mics.spb.ru |

For citation: Kolchin M. A., Shilin I. A., Klimov N. V., Garayzuev D. S., Muromtsev D. I., Zakoldaev D. A. Method of aggregation and normalization of heterogeneous sensor network data // Izvestiya Vysshikh Uchebnykh Zavedeniy. Priborostroenie. 2015. Vol. 58, N 11. P. 945—951 (in Russian).

DOI: 10.17586/0021-3454-2015-58-11-945-951