

УДК 004.822

## KNOWLEDGE NET: МОДЕЛЬ И СИСТЕМА НАКОПЛЕНИЯ, ПРЕДСТАВЛЕНИЯ И ИСПОЛЬЗОВАНИЯ ЗНАНИЙ И ДАННЫХ

О.Н. Тушканова<sup>1,2</sup>, В.В. Самойлов<sup>3</sup>

<sup>1</sup> Санкт-Петербургский институт информатики и автоматизации РАН, Санкт-Петербург, Россия

<sup>2</sup> Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия  
tushkanova.on@gmail.com

<sup>3</sup> ООО «Информационные крылья», Санкт-Петербург, Россия  
vsamoilov@infowings.ru

### Аннотация

Описана разрабатываемая авторами модель Knowledge Net, предназначенная для формализации накопления, представления и использования знаний и данных единого информационного пространства интеллектуального предприятия, в том числе в рамках концепции Индустрия 4.0. В основу модели Knowledge Net положена графовая модель данных. Модель Knowledge Net является частью прикладных разработок цифровой платформы предприятия. Разработка программной системы, поддерживающей модель Knowledge Net, выполняется в концепции открытого программного кода. В работе приведён пример использования программного прототипа для описания объектов производственного предприятия. Ключевой особенностью Knowledge Net, отличающей разработку от известных аналогов, является комбинация таких свойств, как поддержка многоаспектности описания субъектов и объектов предприятия, возможность расширения сети понятий и сущностей вместе с развитием предприятия, динамическое формирование структуры аспектов и их свойств, множественность возможных типизаций сущностей, представленных моделью Knowledge Net, а также поддержка версии модели знаний и данных.

**Ключевые слова:** единое информационное пространство, модель знаний, представление знаний, аспект, онтология, граф знаний, графовая модель данных.

**Цитирование:** Тушканова, О.Н. Knowledge Net: модель и система накопления, представления и использования знаний и данных / О.Н. Тушканова, В.В. Самойлов // Онтология проектирования. – 2019. – Т. 9, №1(31). – С.117-131. – DOI: 10.18287/2223-9537-2019-9-1-117-131.

### Введение

Одной из основных черт современной концепции цифрового производства, называемой Индустрия 4.0, является сращивание промышленного производства, информационно-коммуникационных технологий и технологий искусственного интеллекта. Эта тенденция развития промышленности в настоящее время является одной из наиболее актуальных и часто обсуждаемых тем среди учёных и практиков всего мира. Правительства многих стран, включая Российскую Федерацию [1], объявили концепцию цифровой экономики ключевой стратегией развития в области высоких технологий.

По данным исследования [2] руководители крупнейших компаний в области промышленного производства, добычи нефти и газа, энергетики, горнодобывающей промышленности в Америке, Азии и Европе готовы инвестировать значительные средства в Индустрию 4.0 с целью цифровой трансформации их предприятий. В качестве основной цели инвестирования руководители чаще всего называют стремление превратить свои корпорации в сеть сильно связанных гибких компаний, функционирующих на основе знаний, которые будут способны удовлетворять требованиям нового цифрового мира.

Основная идея концепции Индустрии 4.0 направлена на развитие интеллектуальных фабрик, которые должны функционировать в рамках автоматизированных производств, управляемых программным обеспечением, а также на интеграцию всех информационных и программных компонентов на уровне предприятия в единое информационное пространство (ЕИП). Программные системы интеллектуальных фабрик должны полностью контролировать все физические процессы, происходящие на предприятии, эффективно взаимодействовать и сотрудничать друг с другом и с персоналом, а также быть способными принимать децентрализованные решения. Для большинства отраслей реализация таких преобразований является серьёзной проблемой, так как чаще всего информационные компоненты и программные системы исторически разрабатывались независимо друг от друга.

Обязательной компонентой интеллектуального предприятия, развивающегося в рамках концепции Индустрии 4.0, является система накопления, представления и использования данных и знаний, которая должна формировать семантически однозначное ЕИП для всех его подразделений, а также физических и виртуальных объектов производственной системы [3]. Такое пространство должно агрегировать информацию обо всех субъектах и объектах предприятия на всех этапах жизненного цикла (ЖЦ).

Особенность ЕИП состоит в том, что в нём разные подразделения могут рассматривать один и тот же объект или субъект с разных точек зрения. Например, отдел логистики могут интересоваться массово-габаритные характеристики некоторого изделия, а отдел закупок – экономически выгодные аналоги сборочных единиц, которые обладают сходными функциональными возможностями. Это формирует специфическое требование к системе накопления знаний интеллектуального предприятия, сущность которого состоит в обеспечении возможности описания различных аспектов одного и того же объекта в рамках единой структуры представления знаний и данных информационного пространства. Следствием этого требования является вывод о том, что в структуре данных и знаний, используемой для совместной работы множества подразделений в ЕИП, не может быть единой типизации (таксономии) понятий онтологии.

Другой важной специфической чертой интеллектуального предприятия является необходимость постоянного развития и пополнения данных и знаний ЕИП. Действительно, вместе с развитием предприятия, совершенствованием используемых механизмов управления и соответствующей эволюцией программных компонент семантические модели объектов и субъектов также должны постоянно развиваться. Динамика моделей объектов и субъектов производственной системы определяется ЖЦ производства и производимых продуктов, и для каждого его этапа может потребоваться своя модель информации. Параллельно с эволюцией потребных моделей информации необходимо заботиться о поддержке целостности знаний и данных. Такое требование к динамической поддержке баз знаний и данных требует разработки соответствующих формальных методов и программных средств.

В соответствии с современными представлениями задача формирования и динамической поддержки ЕИП знаний и данных предприятия в рамках концепции Индустрии 4.0 возлагается на инфраструктурную компоненту системы управления предприятием, которую принято называть его цифровой платформой. Основной задачей этой платформы является интеграция гетерогенных компонент системы на всех этапах её ЖЦ.

Целью данной работы является описание предлагаемой модели накопления, представления и использования знаний и данных ЕИП производственной системы с учётом её многоаспектности, динамики и требований по поддержке её целостности, а также соответствующего программного прототипа. Эта модель является частью текущих прикладных разработок цифровой платформы предприятия. Соответствующий проект, программный прототип и описы-

ваемая модель получили название Knowledge Net. Разработки системы в данный момент выполняются в концепции открытого программного кода.

В разделе 1 обсуждаются наиболее близкие аналоги существующих систем представления знаний и данных. В разделе 2 пояснены некоторые понятия, важные для понимания принципов устройства модели Knowledge Net. В разделе 3 формулируются ключевые особенности разрабатываемой модели Knowledge Net. Раздел 4 посвящён общему описанию модели Knowledge Net и методики её применения. В разделе 5 кратко описано программное обеспечение, поддерживающее разработанную модель. В разделе 6 приведён пример описания производственного объекта модели Knowledge Net.

## 1 Краткий обзор систем семантического представления знаний и данных

В разделе представлены разработки, в которых использованы подходы, частично реализующие возможности описанной в данной работе модели и системы Knowledge Net.

В [4] предлагается концепция, получившая название Administrative Shell, которая должна поддерживать цифровое представление всей доступной информации об объектах производства, представляющих собой, например, некие механизмы, их части или программные компоненты. Цифровое представление объектов в рамках этой концепции реализуется с помощью таких семантических средств представления знаний как RDF, RDF-Schema и OWL. Использование RDF-концепции имеет несколько преимуществ, среди которых следует упомянуть встроенную возможность использования глобального уникального идентификатора (URI) объекта, возможность её простой интеграции с уже существующими базами знаний, описанными на RDF, а также возможность использовать язык SPARQL в качестве языка запросов. Следует отметить, что предложенная в [4] схема описания объектов не универсальна и на момент публикации работы поддерживала только некоторые типы объектов. Кроме того, использование этого подхода требует предварительного создания тезауруса и таксономии классов объектов. Это не подходит для развивающегося предприятия, так как в этом случае модель знаний должна эволюционировать вместе с развитием объектов, технологий и организационных структур предприятия в целом. В работе не описаны способы отслеживания истории изменений описаний объектов или схемы. Следует отметить, что языки RDF, RDF-Schema и OWL такую возможность не предусматривают.

В [5] описываются результаты сотрудничества между компанией Siemens и Оксфордским университетом, целью которого было облегчение разработки онтологий и их использования в приложениях компании. Итогом этого сотрудничества стала надстройка над Web Protege – программный прототип Siemens-Oxford Model Manager (SOMM). Этот инструмент рассматривается как помощник инженеров, имеющих небольшой опыт работы с семантическими технологиями, при создании онтологических моделей и заполнении их данными. Онтологии, созданные с помощью SOMM, рассматриваются авторами как ядро ЕИП компании Siemens. SOMM реализует фрагмент OWL 2 RL и поставляется вместе с компонентой поддержки схемы и рассуждений над данными. Основным недостатком такого онтологического подхода для создания ЕИП предприятия, который предлагается также в [6, 7], является трудоёмкость первоначального создания схемы, которая позволяла бы описать все объекты реального предприятия. В случае, если предприятие только начинает развиваться и набор используемых им объектов и компонентов постоянно меняется, разработка единой неизменяемой схемы практически невозможна или сопряжена с большими трудозатратами. Более того, даже крупные компании часто не способны разработать свои онтологии настолько детально, чтобы стало возможно применение в них механизмов логического вывода.

В работе [8] описаны основные принципы устройства семантического ядра цифровой платформы xIRBIS-ML. Авторы позиционируют эту разработку как информационно-аналитическую систему, ориентированную на знания. По утверждению авторов, подсистема сбора и обработки документов системы xIRBIS-ML позволяет извлекать из текстовых документов графы понятий и сопоставлять их с существующими онтологиями предприятия. Однако в работе не описана степень участия эксперта для успешного выполнения операции. Отсутствует подробное описание технологий, используемых для создания и обработки онтологий, а также реализации возможности системой поддержки уникальных идентификаторов для всех сущностей системы и поддержки динамических изменений созданной схемы онтологии в процессе развития предприятия.

В работах [9, 10] описаны некоторые принципы новой технологии компонентного проектирования интеллектуальных систем, в частности предлагается использовать семантическую модель баз знаний и привлекать теорию графов для построения алгоритмов их обработки. В этих работах приводится подробное описание базового универсального языка семантических сетей (SC-кода), который, по мнению авторов [9, 10], должен стать стандартом для разработки унифицированных семантических сетей, обеспечивающих представление и интеграцию различных видов знаний в интеллектуальных системах нового поколения, а также описание этапов разработки таких интеллектуальных систем. Однако из текста работ не ясно, выполнялись ли авторами какие-либо прикладные программные разработки, направленные на реализацию описанного подхода и технологии.

Анализ исследований, посвящённых разработке методов и средств создания и поддержки ЕИП интеллектуального предприятия, показывает высокую заинтересованность индустрии в таких средствах. Но уровень их развития является недостаточным для активного внедрения результатов разработок.

## **2 Ключевые свойства модели Knowledge Net**

Семантическая модель накопления, представления и использования данных и знаний является ключевой компонентой интеллектуального предприятия, развиваемой в рамках концепции Индустрии 4.0 [3]. Основной задачей этой компоненты является предоставление сервиса *создания и поддержки* ЕИП знаний и данных интеллектуального предприятия для всех его подразделений. Можно говорить также о поддержке всего ЖЦ такого пространства.

Модель Knowledge Net разрабатывается в рамках исследовательского проекта, а соответствующее программное обеспечение является вариантом реализации этой модели. Ключевыми особенностями разрабатываемой модели ЕИП Knowledge Net, отличающими её от существующих аналогов являются:

- *учёт динамической природы модели* описания, представления и формализации знаний и данных: описание свойств и структуры представления моделей объектов, которые эволюционируют вместе с динамикой развития объектов и предприятия в целом;
- *поддержка многоаспектности модели* описания и представления знаний об объекте и характеризующих его данных: объект может иметь множество различных описаний и свойств, связанных с различными контекстами его использования или “точками зрения” на него (аспектами) со стороны пользователей и задач;
- *графовая модель представления* пространства знаний и данных: аспекты и свойства объекта представляются в форме единой динамически расширяемой модели графа, который образует единую сеть для представления знаний и данных, называемую *метаграфом*, в котором реализация случаев использования знаний и данных сводится к поиску на этом метаграфе;

- *множественность возможных типизаций* (таксономий) объектов, представленных моделью Knowledge Net: формальные источники, как и эксперты-носители знаний, могут формировать различные описания “похожих” объектов, используя множество различных типизаций;
- *многообразие типов и структур данных*, которые могут быть представлены в графе Knowledge Net: свойства объектов, представленных в системе, могут быть описаны атрибутами любых типов, включая слабо структурированные (xml, json и т.п.) и неструктурированные данные, например, САД-модели, PDF-файлы, текстовые описания, изображения, ссылки на другие объекты и их свойства и т.п.;
- *единая ссылочная сеть*: модель даёт возможность выполнения нечёткого поиска, выявления аналогов и кластеризации объектов, субъектов и аспектов-свойств;
- *поддержка специализированных механизмов фасетного поиска* объектов: поиск аналогов, поддержка поиска примеров объектов с заданными свойствами и характеристиками;
- *версионность* модели знаний и данных: возможность хранения истории *всех* изменений *всех* версий сущностей модели знаний Knowledge Net, которые происходили на всех этапах ЖЦ этих сущностей, как это, например, и реализуется в структурах представления знаний и данных типа Data Vault [11].

### 3 Используемая терминология

Под *субъектом* в данной работе понимается некоторый носитель, поставщик или пользователь знаний, активный элемент внешнего (по отношению к системе управления предприятием) мира. Субъект может представлять собой организацию или индивида, обладающего сознанием и волей, который способен к целесообразной деятельности, направленной на тот или иной объект. Примерами субъектов в бизнес-процессах организации являются поставщики, производители, разработчики продукта, сама организация, сервис или программа, осуществляющая добавление данных или знаний в модель.

*Объект* – это физический или виртуальный результат целенаправленной деятельности *субъекта*, предмет или явление реального или виртуального мира. Примерами объектов являются: товар, производимый предприятием, деталь, устройство, сборочная единица и т.д., услуга, сервис и т.п. Объекты могут обладать различными свойствами и могут быть представлены различными описаниями.

Ключевым понятием разрабатываемой модели является понятие *аспекта*. Согласно [12], под аспектом следует понимать “взгляд, точку зрения, с которой рассматривается какое-либо явление, понятие, перспектива”. *Аспектом* некоторого объекта будем называть одно из однотипных проявлений свойств его модели, общее для всех его описаний. Аспект обладает *наименованием*, возможно, *единицей измерения*, может характеризоваться *типами данных*, описывающих *свойства*, относящиеся к аспекту, и *доменами* их допустимых значений. В качестве домена допустимых значений аспекта могут выступать, например, *словари*, численные интервалы или формулы, задающие некоторые ограничения. Аспекты могут быть общими (глобальными) либо существовать только в некотором контексте, который задаётся определённым *субъектом*. Примерами аспектов могут выступать, например:

- *Вес* (вещественный тип, один из вариантов единицы измерения - килограмм, домен значений -  $[0; +\infty)$ );
- *Объём* (вещественный тип, один из вариантов единицы измерения - это литр, домен значений -  $[0; +\infty)$ );
- *Количество фаз переменного тока* (целый тип, единица измерения – штуки, домен значений - все целые неотрицательные числа);

- *Цвет* (строковый тип, нет единицы измерения, возможный домен значений – множество всех цветов) и т.д.

В рамках предлагаемой модели аспекты могут объединяться в иерархии. В этом случае дочерние аспекты будем называть *свойствами* по отношению к родительским аспектам. *Свойство аспекта* – это проявление некоторого дочернего аспекта в контексте родительского аспекта – владельца свойства. Например, *Вес* является самостоятельным аспектом, но он также может быть свойством других аспектов. Аспект *Массогабаритные показатели* может иметь свойства *брутто Вес* и *нетто Вес*, которые ссылаются на аспект *Вес*. Тот же аспект *Вес*, при этом, может входить и в свойство *максимальный для размещения Вес* в контексте родительского аспекта *Характеристики устройства хранения* (стеллажа или полки).

*Свойство объекта* – это проявление аспекта в том или ином контексте, задаваемом конкретным объектом. Свойство, в общем случае, обладает ссылкой на аспект, к которому оно относится, ролевым именем свойства (возможно пустым), мощностью множества допустимых значений и, возможно, конкретными значениями. Мощность свойства означает допустимое число экземпляров значений этого свойства у объекта. Это понятие аналогично общепринятому понятию *мощность отношений* (кратность).

*Словарь* в контексте разрабатываемой системы – это допустимое множество значений некоторого аспекта, имеющего строковый тип. Примером словаря может выступать список возможных цветов для аспекта, описывающего цвет объекта; иерархия классификации объектов по функциональному назначению согласно стандарту [13] также может быть представлена с помощью словаря, а затем использована для указания функционального назначения объекта.

#### **4 Описание модели Knowledge Net и методики её применения**

Основными отличительными особенностями разрабатываемой модели Knowledge Net является поддержка многоаспектности описания сущностей, возможность расширения сети понятий и сущностей, динамического формирования структуры аспектов и их свойств. Выполнение этих требований подразумевает отсутствие жёсткой типизации сущностей, что приводит к сложности хранения данных и знаний в традиционной реляционной или объектной структуре. В связи с этим в основу модели Knowledge Net положена графовая модель данных. Под графовой моделью данных обычно понимают модель, в которой схема и экземпляры данных представляются в структурированном виде как ориентированный размеченный граф, а все манипуляции с данными и ограничения целостности данных могут быть выражены через операции на этом графе [14]. Примерами операций являются нахождение кратчайшего пути (например, при оценке семантической близости понятий онтологии, представленной графом), выделение компонент связности (при поиске независимых компонент знаний и в задачах их кластеризации) и т.д.

В качестве системы управления графовой базой данных выбрана СУБД OrientDB с открытым исходным кодом, которая объединяет в себе возможности документо-ориентированной и графовой базы данных. Кроме того, OrientDB поддерживает интерфейс объектно-ориентированной базы данных, который работает поверх документо-ориентированного слоя. OrientDB написана на языке Java и распространяется под лицензией Apache [15]. Важным преимуществом OrientDB является поддержка протокола Apache's Tinker Pop, что позволит в будущем реализовать импорт знаний из существующих RDF-хранилищ, например, из ресурсов Linked Data. Поддержка интерфейса объектно-ориентированной базы данных означает, что СУБД OrientDB позволяет задавать классы вершин и рёбер будущего графа, аналогичные классам в объектно-ориентированной модели.

На логическом уровне модели следует выделить некоторые основные сущности схемы данных, которые являются понятиями метаонтологии или метаграфа, а именно, *Субъект*, *Аспект*, *Объект*, *Единица измерения*. К вспомогательным сущностям будем относить такие понятия как *Запись в истории*, *Свойство аспекта*, *Справочник*, *Элемент Справочника* и некоторые другие.

Логическая модель данных, представленная графом (рисунок 1), является ядром модели Knowledge Net. Для представления этой логической модели в Orient DB были созданы следующие классы вершин: *Subject*, *Object*, *Aspect*, *Aspect Property*, *Object Property*, *Object Property Value*, *Measure*, *Measure Group*, *History Event*, *Reference Book Item*, *GUID* (англ. Global Unified Identifier) и др.

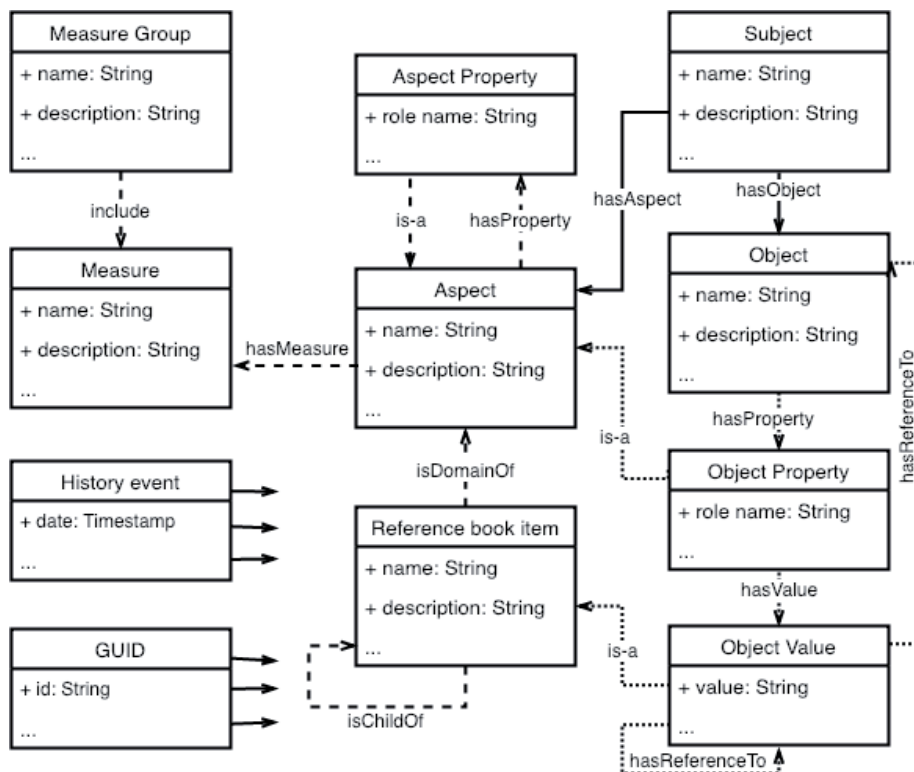


Рисунок 1 - Основные сущности графа знаний и связи между ними

Описанная структура модели Knowledge Net удобна тем, что она позволит организациям самостоятельно или с привлечением экспертов создавать графы знаний, которые могут быть положены в основу модели ЕИП интеллектуального предприятия.

Ключевыми вершинами этого графа знаний являются узлы *Aspect* и *Aspect Property*, предназначенные для описания аспектов и свойств этих аспектов. Сами свойства аспектов также являются аспектами, но с некоторыми дополнениями, например, с уточняющими ролевыми именами. Пара *аспект – свойство* чаще всего используется для группировки связанных аспектов. Например, свойствами аспекта *Геометрия объекта* могут быть аспекты *Длина*, *Ширина*, *Высота*.

Вершины типа *Measure* и *Measure Group* позволяют описывать единицы измерения, которые могут быть ассоциированы с аспектами. Узлы типа *Reference Item* служат для представления словарей, которые описывают допустимое множество значений некоторого строкового аспекта и их элементов. Штриховым пунктиром на рисунке 1 выделены отношения, которые являются основными при описании аспектов графа знаний предприятия.

Вершины типа *Subject* предназначены для описания основных свойств субъектов графа знаний. Объекты могут существовать только в контексте принадлежности некоторому субъекту, в то время как аспекты могут быть и глобальными, то есть не относиться к какому-либо субъекту. Узлы типа *Object*, *Object Property* и *Object Value* позволяют описывать объекты и их свойства. Каждый узел типа *Object Property* всегда связан с аспектом, который он реализует.

Описанный подход к представлению и хранению знаний и данных позволяет поддерживать многоаспектность описания объекта графа знаний, так как для добавления нового аспекта описания объекта, рассматривающего этот объект с другой точки зрения, потребуется всего лишь добавить новый узел графа *Object Property* и связать его с этим объектом. Точечным пунктиром на рисунке 1 выделены отношения, которые позволяют описывать взаимосвязи объектов и их свойств.

Граф знаний, который может быть разработан с использованием модели Knowledge Net, рассматривается авторами как общее информационное ядро (или часть такового) для множества систем в рамках как одной, так и нескольких организаций. Эти системы, с одной стороны, должны иметь возможность ссылаться внутри себя на сущности графа, а с другой стороны, иметь возможность обмениваться между собой этими ссылками без необходимости непосредственного доступа к графу знаний. Такие ссылки должны иметь глобальный контекст и обеспечить для любой внешней системы возможность получить по ним одинаковую непротиворечивую информацию из графа. В качестве ссылок на все типы сущностей для внешних и внутренних систем используется уникальный идентификатор *GUID*, представленный в виде отдельных узлов графа, связанных с каждой сущностью. В течение всего ЖЦ сущности в графе знаний её *GUID* остаётся неизменным. Наличие такого глобального идентификатора делает возможным хранение всей истории изменений всех сущностей в системе на протяжении всего их ЖЦ. Хранение истории изменений реализуется с помощью узлов типа *History Event* и нескольких вспомогательных типов узлов, не представленных на рисунке 1. Узлы *GUID* и *History Event* связаны с каждой сущностью графа (на рисунке 1 эти связи не представлены). Предполагается, что наполнение соответствующего графа предприятия в рамках модели Knowledge Net может выполняться как вручную, так и автоматически с привлечением методов машинного обучения для извлечения знаний.

## **5 Программное обеспечение, поддерживающее модель Knowledge Net**

Развитие модели Knowledge Net выполняется в рамках исследовательского проекта и поддерживается конкретными программными разработками. Программный код выложен в открытый доступ на хостинге GitHub [16].

Программная система, поддерживающая модель Knowledge Net, разрабатывается как веб-приложение с упрощённым пользовательским интерфейсом (рисунок 2), который позволяет вводить, искать и просматривать сущности разработанного графа знаний организации. На данный момент этот интерфейс является основным средством пользовательского доступа к описываемым сущностям. С ростом объёма данных основными средствами доступа к объектам станут API для внешних и внутренних подсистем предприятия, функционирующих в контексте единой платформы и использующих граф, а также специализированные “продвинутые” пользовательские интерфейсы с расширенными средствами поиска, навигации и отображения. После этого роль существующего пользовательского интерфейса станет скорее служебной, используемой при администрировании и контроле информационного пространства интеллектуального предприятия.



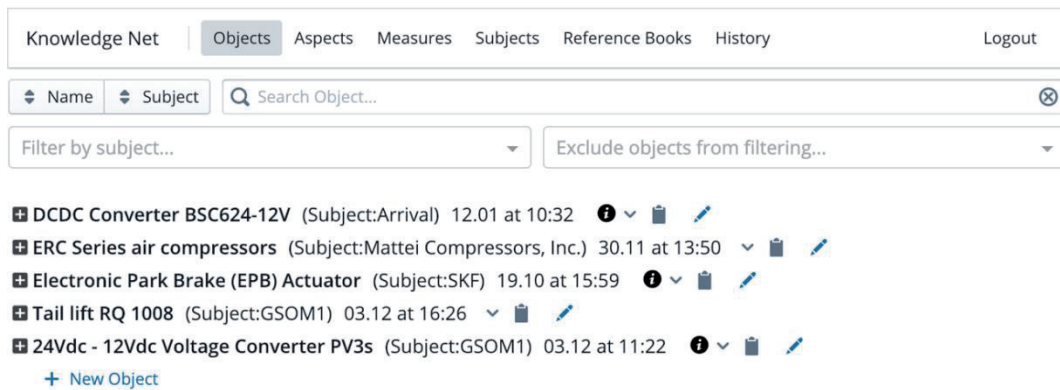


Рисунок 2 – Пользовательский интерфейс базовой версии программной системы модели Knowledge Net

На рисунке 2 представлена подсистема для ввода и отображения объектов организации. В верхней части страницы представлены кнопки переключения на другие разделы интерфейса, поддерживающие работу с аспектами (*Aspects*), единицами измерения (*Measures*), субъектами (*Subjects*), словарями (*Reference Books*) и историей изменений (*History*). Основная часть страницы подсистемы отведена под список объектов, каждый из которых может быть развернут для просмотра его свойств. Создание нового объекта выполняется на отдельной странице. Раздел работы с аспектами (рисунок 3) имеет схожую структуру.

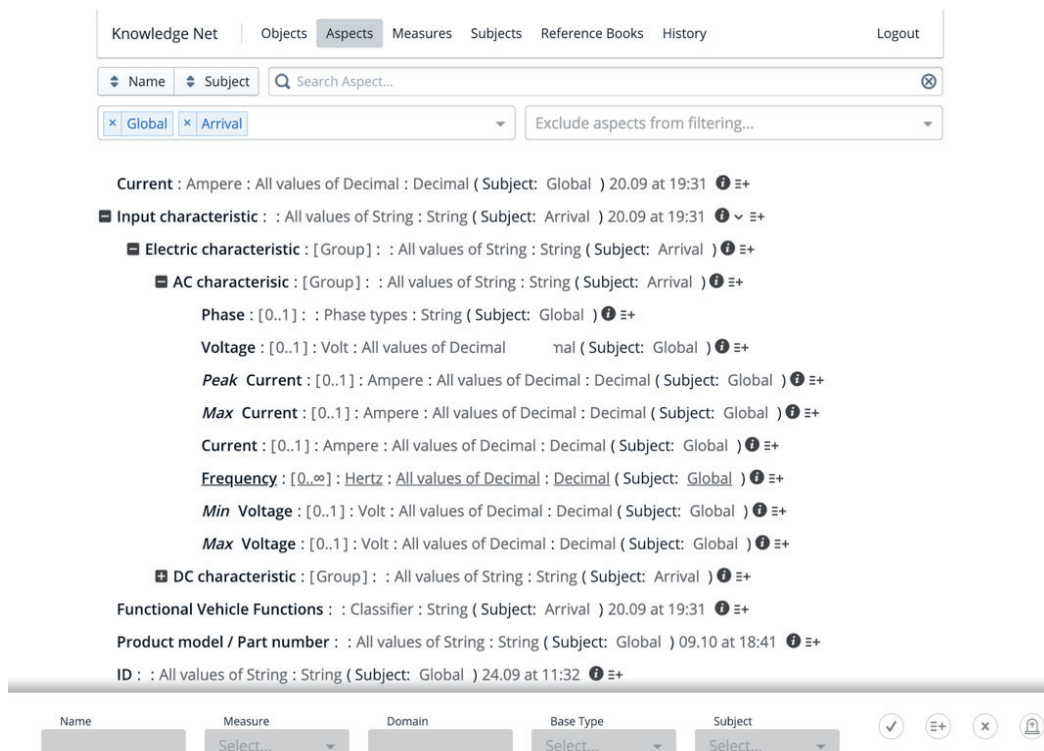


Рисунок 3 – Пользовательский интерфейс базовой версии для работы с аспектами

Разработанное программное обеспечение модели Knowledge Net, обладает следующими основными функциональными возможностями.

- Поддержка основных единиц измерения из общероссийского классификатора единиц измерения [17], а также механизма нечёткого поиска по ним.
- Создание уникального идентификатора (GUID) всех сущностей графа знаний внутри системы Knowledge Net.

- Просмотр и редактирование субъектов, что позволяет вносить в граф знаний информацию о субъектах организации (в первой версии системы субъекты могут обладать только уникальным именем, а интерфейс позволяет просмотреть все сущности, которые связаны с каждым субъектом).
- Создание, редактирование и просмотр аспектов, в том числе, поддержка ввода и редактирования наименования, единицы измерения, типа и домена значений. Аспекты могут иметь свойства, которые образуют деревья вложенности (дочернее свойство является фактически новой сущностью, представляющей собой комбинацию “ролевого” имени, по умолчанию пустого), ссылки-связи на другой аспект, который включается как дочерний, и мощности значения. Аспекты-свойства не могут в глобальном графе образовывать циклические ссылки по связям Аспект → Аспект-свойство → Аспект. Это ограничение проверяется при попытке сохранить новое свойство аспекта. Поиск аспектов-свойств осуществляется фасетным способом по наименованию, единицам измерения или типу.
- Поддержка аспектов, имеющих тип *связь* (значением для свойств, реализующих такие аспекты, может быть ссылка на другие сущности системы. В текущей версии программного обеспечения поддержаны ссылки на объекты и на значения других свойств).
- Фильтрация аспектов (в графе знаний количество используемых аспектов будет достаточно большим, при этом разнообразие будет по количеству субъектов, по тематике аспектов и по их структурной сложности).
- Нечёткий поиск по аспектам в графе (в текущей версии программного обеспечения используется свободно распространяемая библиотека для высокопроизводительного полнотекстового поиска Lucene, которая интегрирована в OrientDB).
- Просмотр и редактирование объектов, что позволяет описывать объекты организации и их свойства с использованием введённых заранее аспектов.
- Просмотр связей объекта и поиск связанных сущностей (разработанный интерфейс позволяет по запросу показать связанные с объектом или субъектом сущности).

Таким образом, в базовом программном обеспечении поддержаны основные особенности модели Knowledge Net, связанные с накоплением и представлением знаний, которые позволяют ввести семантические описания объектов в систему и начать использовать полученный граф как основу для ЕИП предприятия.

## 6 Пример описания объекта

Программное обеспечение, поддерживающее модель Knowledge Net, было апробировано на примере создания фрагмента графа знаний для компании-производителя электромобилей. Всего было обработано около 50 объектов разного уровня сложности, которые были отмечены компанией, как наиболее приоритетные. Основные аспекты (около 200), необходимые для описания этих объектов, были выделены и структурированы в рамках другого исследовательского проекта, результаты которого частично описаны в [18]. Такого количества объектов и описывающих их аспектов оказалось достаточно, для того чтобы сделать первые выводы о работоспособности модели.

С целью демонстрации возможностей модели и программного прототипа рассмотрим описание объекта *DCDC Converter BSC624-12V* – конвертера постоянного тока, – выполненное в системе Knowledge Net (рисунок 4).

Для описания объекта в основном использованы аспекты, обладающие свойствами, то есть родительские и дочерние аспекты с несколькими уровнями вложенности, объединенные в иерархии. Например, аспект *Geometry* обладает аспектами-свойствами *Height*, *Width*, *Length*, и все они использованы для описания свойств объекта *DCDC Converter BSC624-12V*.

При этом сам аспект *Geometry* в контексте объекта обладает ролевым именем *outer*, то есть он использован для описания внешней геометрии этого объекта.

- DCDC Converter BSC624-12V (Subject:Arrival) 20.01 at 19:47 ⓘ ⌵ ⌵ ⌵
  - internal ID (Arrival) 724509 ⌵
  - outer Geometry (Arrival) ⌵
    - Height (Arrival) 7 cm ⌵
    - Width (Arrival) 15 cm ⌵
    - Length (Arrival) 34.82 cm ⌵
  - net Weight (Arrival) 4.8 kg ⌵
  - Colour (Arrival) Red ⌵
  - Colour (Arrival) Black ⌵
  - Electric characteristic (Arrival) ⌵
    - DC characteristic (Arrival) ⌵
      - Low voltage (Arrival) ⌵
        - Nominal Voltage (Arrival) 14 V ⌵
        - Max Current (Arrival) 2.5E+2 A ⌵
        - Voltage (Arrival) [ 8 : 16 ] V ⌵
      - ⊕ High voltage (Arrival) ⌵
    - IP-protection (Arrival) IP65 ⌵
    - Cooling system (Arrival) ⌵
      - Cooling system type (Arrival) Liquid cool ⌵
      - Coolant Flow rate (Arrival) 4 ⌵
      - Temperature (GSOM1) [-4E+1 : 65] C ⌵
    - Operation External conditions characteristics (GSOM1) ⌵
      - Function core type (IEC 81346) T transforming object → TA electric energy transforming object → TAB DC/DC converter ⌵
      - Max Power (Arrival) ⓘ 3.5 W ⌵
      - Efficiency (GSOM1) 94.4 % ⌵

Рисунок 4 – Описание объекта *DCDC Converter BSC624-12V* в системе Knowledge Net

На рисунке 5 представлена полная иерархия аспекта *Electric characteristic*. Для каждого аспекта указан тип и домен значений. Для аспектов-свойств в квадратных скобках также указана мощность. Мощность *Group* означает, что аспект по умолчанию может не обладать значением и выступает в качестве объединяющей характеристики для своих дочерних свойств.

- Electric characteristic : : All values of String : String ( Subject: Arrival ) 20.09 at 19:31 ⓘ ⌵ ⌵ ⌵
  - ⊕ AC characteristic : [Group] : : All values of String : String ( Subject: Arrival ) ⓘ ⌵ ⌵
    - Current : [0..1] : Ampere : All values of Decimal : Decimal ( Subject: Global ) ⓘ ⌵ ⌵
    - Max Current : [0..1] : Ampere : All values of Decimal : Decimal ( Subject: Global ) ⓘ ⌵ ⌵
    - High Voltage : [0..1] : Volt : All values of Decimal : Decimal ( Subject: Global ) ⓘ ⌵ ⌵
    - Low Voltage : [0..1] : Volt : All values of Decimal : Decimal ( Subject: Global ) ⓘ ⌵ ⌵
  - Low voltage : [Group] : : All values of String : String ( Subject: Arrival ) ⓘ ⌵ ⌵
    - Nominal Voltage : [0..1] : Volt : All values of Decimal : Decimal ( Subject: Global ) ⓘ ⌵ ⌵
    - Max Current : [0..1] : Ampere : All values of Decimal : Decimal ( Subject: Global ) ⓘ ⌵ ⌵
    - Voltage : [0..1] : Volt : All values of Decimal : Decimal ( Subject: Global ) ⓘ ⌵ ⌵
  - ⊕ High voltage : [Group] : : All values of String : String ( Subject: Arrival ) ⓘ ⌵ ⌵
    - Operating Voltage : [0..∞] : Volt : All values of Decimal : Decimal ( Subject: Global ) ⓘ ⌵ ⌵

Рисунок 5 – Свойства аспекта *Electric characteristic*

Отметим, что при использовании некоторого аспекта со свойствами для описания объекта не все дочерние свойства аспекта обязательно должны быть заполнены. Например, для объекта *DCDC Converter BSC624-12V* не использован аспект *AC characteristic* и его свойства.

Аспект *Function core type* (см. рисунок 6) представляет собой пример аспекта с доменом значений *Словарь*. Этот словарь содержит иерархию классификации объектов по функциональному назначению согласно стандарту [13].

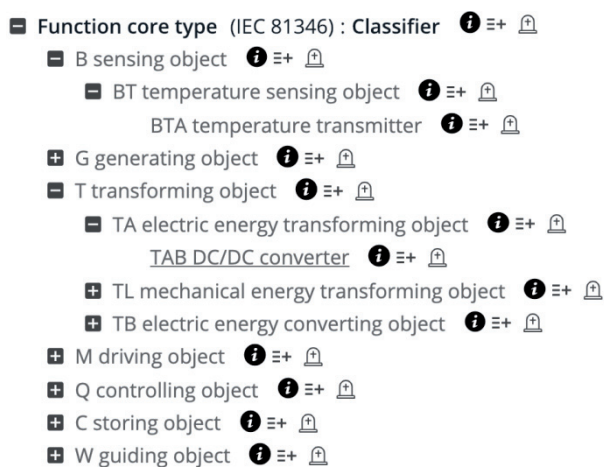


Рисунок 6 – Фрагмент словаря аспекта *Function core type*

Разработанный программный прототип положительно оценен специалистами в области формализации знаний.

## Заключение

Представлена семантическая модель накопления, представления и использования знаний и данных Knowledge Net и поддерживающий её программный прототип, который разрабатывается в концепции открытого программного кода. Структура модели Knowledge Net позволит организациям самостоятельно или с привлечением экспертов создавать графы знаний, которые могут быть положены в основу модели ЕИП предприятия, являющегося ключевой компонентой интеллектуального предприятия нового поколения, развиваемой в рамках концепции Индустрия 4.0.

Основными отличительными особенностями разрабатываемой модели Knowledge Net является поддержка многоаспектности, возможность расширения сети понятий и сущностей вместе с развитием предприятия и динамическое формирование структуры аспектов и их свойств. В основу модели Knowledge Net положена графовая модель данных. Программное обеспечение позволяет ввести семантические описания объектов в систему и использовать полученный в результате граф как основу для представления ЕИП предприятия. В последующих релизах программного обеспечения предполагается значительно расширить функциональные возможности системы Knowledge Net.

## Благодарности

Работа выполнена при финансовой поддержке компании ООО «Информационные крылья» в сотрудничестве с учёными Санкт-Петербургского института информатики и автоматизации РАН и Высшей школы менеджмента СПбГУ.

## Список источников

- [1] Государственная программа “Цифровая экономика Российской Федерации”. Утверждена распоряжением Правительства Российской Федерации от 28 июля 2017 г. №1632-р.

- [2] **Hanley, T.** The Industry 4.0 paradox. Executive summary / T. Hanley, A. Daecher, M. Cotteleer, B. Sniderman - <https://www2.deloitte.com/insights/us/en/focus/industry-4-0/challenges-on-path-to-digital-transformation/summary.html>.
- [3] **Adolphs, P.** Status report - Reference architecture model Industrie 4.0 (rami4.0) / P. Adolphs, H. Bedenbender, D. Dirzus, et al. – VDI, 2015. - [https://ec.europa.eu/futurium/en/system/files/ged/a2-schweichhart-reference\\_architectural\\_model\\_industrie\\_4.0\\_rami\\_4.0.pdf](https://ec.europa.eu/futurium/en/system/files/ged/a2-schweichhart-reference_architectural_model_industrie_4.0_rami_4.0.pdf).
- [4] **Grangel-González, I.** Towards a Semantic Administrative Shell for Industry 4.0 Components / I. Grangel-González, L. Halilaj, G. Coskun, S. Auer, D. Collarana, M. Hofmeister // Proc. of IEEE Tenth International Conference on Semantic Computing (ICSC). – 2016. - P. 230-237.
- [5] **Kharlamov, E.** Capturing Industrial Information Models with Ontologies and Constraints / E. Kharlamov, B. Cuenca Grau, E. Jiménez-Ruiz, S. Lamparter, G. Mehdi, M. Ringsquandl, Y. Nenov, S. Grimm, M. Roshchin, I. Horrocks // Proc. of International Semantic Web Conference. – 2016. - P.325-343.
- [6] **Hauptert J., Schneider M.** The object memory server for semantic product memories. SemProM, Springer, 2013. - P.175–189.
- [7] **Kharlamov, E.** Optique: Ontology-Based Data Access Platform / E. Kharlamov, E. Jiménez-Ruiz, C. Pinkel, M. Rezk, M.G. Skjæveland, A. Soyly, G. Xiao, D. Zheleznyakov, M. Giese, I. Horrocks, A. Waaler // Proc. of ISWC (P&D). - 2015. - [http://ceur-ws.org/Vol-1486/paper\\_24.pdf](http://ceur-ws.org/Vol-1486/paper_24.pdf).
- [8] **Максимов, Н.В.** Семантическое ядро цифровой платформы / Н.В. Максимов, О.Л. Голицина, М.Г. Ганченкова, Д.В. Санатов, А.В. Разумов // Онтология проектирования. - 2018. - № 3 (29). - С. 412-426. – DOI: 10.18287/2223-9537-2018-8-3-412-426.
- [9] **Голенков, В.В.** Проект открытой семантической технологии компонентного проектирования интеллектуальных систем. Часть 1: Принципы создания / В.В. Голенков, Н.А. Гулякина // Онтология проектирования. – 2014. - № 1(11). - С.42–64.
- [10] **Голенков, В.В.** Проект открытой семантической технологии компонентного проектирования интеллектуальных систем. Часть 2: Унифицированные модели проектирования / В.В. Голенков, Н.А. Гулякина // Онтология проектирования. – 2014. - № 4(14). - С.34–53.
- [11] **Linstedt, D.** A short intro to #datavault 2.0 – <https://danlinstedt.com/allposts/datavaultcat/a-short-intro-to-datavault-2-0>.
- [12] Философский энциклопедический словарь (2010). – <http://philosophy.niv.ru/doc/dictionary/philosophy/index.htm>.
- [13] IEC 81346-2:2009. Industrial systems, installations and equipment and industrial products - Structuring principles and reference designations - Part 2: Classification of objects and codes for classes.
- [14] **Renzo A.** Survey of graph database models / A. Renz, C. Gutierrez // ACM Computing Surveys. – 2008. - №40 (1). - P.1–39.
- [15] OrientDB Community Edition. – <https://orientdb.org>.
- [16] Source Code of the Knowledge Net. – <https://github.com/InfoWings/Knowledge-Net/>.
- [17] ОК 015-94 (МК 002-97). Общероссийский классификатор единиц измерения (утв. Постановлением Госстандарта России от 26.12.1994 N 366).
- [18] **Kudryavtsev, D.** Mind Mapping and Spreadsheets in Collaborative Design of Knowledge Graphs / D. Kudryavtsev, T. Gavrilova, I. Leshcheva, A. Begler, M. Kubelskiy, O. Tushkanova // Joint Proceedings of the BIR 2018 Forum, Workshops and Doctoral Consortium, September 24-26 2018, Stockholm, Sweden. P. 82-93.

## KNOWLEDGE NET: MODEL AND SYSTEM FOR ACCUMULATION, REPRESENTATION AND USE OF KNOWLEDGE AND DATA

O.N. Tushkanova<sup>1,2</sup>, V.V. Samoylov<sup>3</sup>

<sup>1</sup> St. Petersburg Institute for Informatics and Automation of RAS, St. Petersburg, Russia

<sup>2</sup> Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia  
tushkanova.on@gmail.com

<sup>3</sup> Info Wings LLC, St. Petersburg, Russia  
vsamoilov@infowings.ru

### Abstract

The paper describes the Knowledge Net data and knowledge accumulation model developed by the authors for the company Info Wings LLC, proposed as a model for formalizing, presenting and using of knowledge and data for integrated information space of an intelligent enterprise, e.g. within the framework of Industry 4.0. The Knowledge Net is based on a graph data model and is a part of the ongoing application development of a digital enterprise platform. The code of a software system that supports the Knowledge Net model is open sourced. The paper provides an example of using a software prototype to describe objects of a manufacturing enterprise. A key feature of Knowledge Net, which distinguishes development from well-known counterparts, is the combination of properties such as support for multi-aspect descriptions of subjects and objects of an enterprise, the possibility of expanding the network of concepts and entities along with the development of the enterprise, dynamic formation of the structure of aspects and their properties, multiplicity of possible typifications presented by the Knowledge Net model, as well as support for the versioning knowledge and data models.

**Key words:** unified information space, knowledge model, knowledge representation, aspect, ontology, knowledge graph, graph data model.

**Citation:** Tushkanova ON, Samoylov VV. Knowledge Net: model and system for accumulation, representation and use of knowledge and data [In Russian]. *Ontology of designing*. 2019; 9(1): 117-131. - DOI: 10.18287/2223-9537-2019-9-1-117-131.

### Acknowledgment

The research was done in the interests and with the financial support of the company Info Wings LLC in collaboration with scientists from the St. Petersburg Institute of Informatics and Automation of the Russian Academy of Sciences and the Graduate School of Management at St. Petersburg State University.

### References

- [1] State program “Digital economy of the Russian Federation” [In Russian]. Approved by the order of the Government of the Russian Federation dated July 28, 2017. №1632-r.
- [2] Hanley T, Daecher A, Cotteleer M, Sniderman B. The Industry 4.0 paradox. Executive summary. <https://www2.deloitte.com/insights/us/en/focus/industry-4-0/challenges-on-path-to-digital-transformation/summary.html>.
- [3] Adolfs P, Bedenbender H, Dirzus D. et al. Status Report - Reference architecture model Industrie 4.0 (rami4.0), 2015. - [https://ec.europa.eu/futurium/en/system/files/ged/a2-schweichhart-reference\\_architectural\\_model\\_industrie\\_4.0\\_rami\\_4.0.pdf](https://ec.europa.eu/futurium/en/system/files/ged/a2-schweichhart-reference_architectural_model_industrie_4.0_rami_4.0.pdf).
- [4] Grangel-González I, Halilaj L, Coskun G, Auer S, Collarana D, Hofmeister M. Towards a Semantic Administrative Shell for Industry 4.0 Components. Proc. of IEEE Tenth International Conference on Semantic Computing (ICSC); 2016: 230-237.
- [5] Kharlamov E, Cuenca Grau B, Jiménez-Ruiz E, Lamparter S, Mehdi G, Ringsquandl M, Nenov Y, Grimm S, Roshchin M, Horrocks I. Capturing Industrial Information Models with Ontologies and Constraints. International Semantic Web Conference; 2016: 325-343.
- [6] Hauptert J, Schneider M. The object memory server for semantic product memories. SemProM, Springer; 2013: 175–189.

- [7] **Kharlamov E, Jiménez-Ruiz E, Pinkel C, Rezk M, Skjæveland MG, Soylu A, Xiao G, Zheleznyakov D, Giese M, Horrocks I, Waaler A.** Optique: Ontology-Based Data Access Platform. Proc. of ISWC (P&D); 2015. - [http://ceur-ws.org/Vol-1486/paper\\_24.pdf](http://ceur-ws.org/Vol-1486/paper_24.pdf).
- [8] **Maksimov NV, Golitsina OL, Ganchenkova MG, Sanatov DV, Razumov AV.** Semantic core of the digital platform [In Russian]. *Ontology of designing*. 2018; 3 (29): 412-426. – DOI: 10.18287/2223-9537-2018-8-3-412-426.
- [9] **Golenkov VV, Gulyakin NA.** The project is an open semantic technology component design of intelligent systems. Part 1: Creation principles [In Russian]. *Ontology of designing*. 2014; 1 (11): 42–64.
- [10] **Golenkov VV, Gulyakin NA.** The project of an open semantic technology component design of intelligent systems. Part 2: Unified design Models [In Russian]. *Ontology of designing*. 2014; 4(14): 34–53.
- [11] **Linstedt D.** A short intro to #datavault 2.0. - <https://danlinstedt.com/allposts/datavaultcat/a-short-intro-to-datavault-2-0>.
- [12] Philosophical Encyclopedic Dictionary (2010). [In Russian]. - <http://philosophy.niv.ru/doc/dictionary/philosophy/index.htm>.
- [13] IEC 81346-2:2009. Industrial systems, installations and equipment and industrial products -- Structuring principles and reference designations - Part 2: Classification of objects and codes for classes.
- [14] **Renzo A, Gutierrez C.** Survey of graph database models. *ACM Computing Surveys* 2018; 40 (1): 1–39.
- [15] OrientDB Community Edition. - <https://orientdb.org>.
- [16] Source Code of the Knowledge Net. - <https://github.com/InfoWings/Knowledge-Net>.
- [17] OK 015-94 (MK 002-97). All-Russian classifier of measurement units [In Russian]. (app. Resolution of state standard of Russia of 26.12.1994 N 366).
- [18] **Kudryavtsev D., Gavrilova T., Leshcheva I., Begler A., Kubelskiy M., Tushkanova O.** Mind Mapping and Spreadsheets in Collaborative Design of Knowledge Graphs Joint Proceedings of the BIR 2018 Forum, Workshops and Doctoral Consortium, September 24-26 2018, Stockholm, Sweden. - P.82-93.

## Сведения об авторах



**Тушканова Ольга Николаевна**, 1988 г. рождения. Окончила Южный федеральный университет в 2011 г. Научный сотрудник Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН), доцент Санкт-Петербургского политехнического университета Петра Великого. Кандидат технических наук (2017). Количество печатных работ 28. Область научных интересов: интеллектуальный анализ данных, машинное обучение, большие данные, онтологии, извлечение знаний, семантические технологии.

**Olga Nikolaevna Tushkanova** (b. 1988). Researcher in St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences, St. Petersburg, assistant professor in Peter the Great St. Petersburg Polytechnic University, St. Petersburg. Received PhD in computer science from St. Petersburg Institute for Informatics and Automation of Russian Academy of

Sciences in 2017. Received MS degree in system analysis from Southern Federal University in Rostov-on-Don (2011). Main publications: 28. Current scientific interests: data mining, machine learning, Big Data, ontologies, knowledge engineering, semantic technologies.



**Самойлов Владимир Владимирович**, 1971 г. рождения. Окончил Высшее военноморское инженерное училище им. Ф.Э. Дзержинского в 1993 г. Ведущий аналитик ООО “Информационные крылья”. В списке научных трудов более 30 работ в области много-агентных приложений, систем поддержки принятия решений, онтологий, механизмов машинного обучения и объединения гетерогенных данных из различных источников.

**Vladimir Vladimirovich Samoylov** (b. 1971) graduated from Higher Naval Engineering School named after F.E. Dzerzhinsky in 1993. Leading Analyst at Info Wings LLC. The list of scientific papers includes more than 30 works in the field of multi-agent applications, decision support systems, ontologies, machine learning mechanisms and combining heterogeneous data from various sources.