

УДК 629.7

ИЕРАРХИЧЕСКИЕ И АССОЦИАТИВНЫЕ СВЯЗИ МЕЖДУ ТЕРМИНАМИ В ТЕЗАУРУСЕ НА ПРИМЕРЕ СЛОВАРЯ ПРОЕКТАНТА

©2012 Н. М. Боргест, Д. В. Шустова, С. Р. Гиматдинова

Самарский государственный аэрокосмический университет
имени академика С. П. Королёва (национальный исследовательский университет)

Рассматриваются вопросы создания тезауруса предметной области «Проектирование самолёта. Этап разработки технических предложений». Описываются виды связей между терминами в тезаурусе и принципы их создания. Основное внимание уделено созданию ассоциативных связей как наиболее противоречивых и требующих научного обоснования. Приводятся примеры синтаксического представления тезауруса на языке OWL.

Онтология предметной области, тезаурус, проектирование самолёта, родо-видовое отношение, иерархия, семантическая сеть, синтаксическое представление.

При проектировании авиационной техники по мере накопления опыта решения задач на отдельных этапах проектирования и управления разработками появляется возможность всё ближе подходить к использованию интегрированных систем автоматизированного проектирования (САПР) с полным информационным обеспечением в виде информационного тезауруса. Разработка информационного тезауруса по конструкции самолёта обеспечивает удовлетворение одного из важных требований при использовании САПР в практике проектирования – наличие полной информации о проектируемом объекте – самолёте на стадии концептуального проектирования [1]. Разными специалистами, в зависимости от квалификации и технического кругозора, названные выше процедуры формулируются по-разному, а интерпретации одного и того же термина напрямую зависят от сложившегося за многие годы понятия [2].

Визуальные методы проектирования онтологий способствуют более быстрому и полному пониманию структуры знаний предметной области, что особенно ценно для исследователей, работающих в новой предметной области.

Назначение разрабатываемой прикладной онтологии в том, чтобы описать

концептуальную модель предметной области «Проектирование самолёта». Прикладная онтология описывает концепты, которые зависят как от онтологии задач, так и от онтологии предметной области. Целью прикладной онтологии является создание модели данных, которая должна [1, 3]:

- обеспечивать общую терминологию для предметной области, для совместного использования и понимания всеми пользователями;
- давать точное и непротиворечивое определение значения каждого термина;
- обеспечивать задание семантики с помощью множества аксиом, которые автоматически позволяют получать ответ на множество вопросов о предметной области.

Онтология определяет язык, при помощи которого выражается семантическая часть словаря. Для создания тезауруса могут быть использованы как языки программирования, так и инструментальные программные средства, созданные специально для проектирования, редактирования и анализа онтологий и называемые редакторами онтологий или онтологическими системами. Они позволяют применять онтологический подход для решения

широкого круга задач прикладного характера [3-7].

Protégé – одна из наиболее популярных систем работы с онтологиями, созданная в Стэнфордском университете (<http://protege.stanford.edu/>). С момента создания онтологический редактор Protégé многие годы использовался экспертами в основном для концептуального моделирования в области медицины. В последнее время его стали применять и в других предметных областях, в частности, при создании онтологий для Semantic Web. Изначально единственной моделью знаний, поддерживаемой Protégé, была фреймовая модель. Protégé имеет открытую, легко расширяемую архитектуру и, помимо фреймов, поддерживает наиболее распространенные языки представления знаний (SHOE, XOL, DAML+OIL, RDF/RDFS, OWL).

Онтологический редактор Protégé основан на модели представления знаний ОКВС (Open Knowledge Base Connectivity). Все понятия предметной области делятся на классы, подклассы, экземпляры. Экземпляры описываются фреймом и могут быть как у класса, так и подкласса. Разработка онтологий для Protégé состоит из пяти шагов [6]:

- 1) выделение области, масштаба, границ онтологии;
- 2) определение классов;
- 3) организация иерархии классов;
- 4) формирование фреймов для описания классов, подклассов, экземпляров через определение слотов, т.е. свойств;
- 5) определение значений.

В Protégé существует несколько стратегий проектирования онтологий:

- нисходящая (сверху вниз),
- восходящая (снизу вверх),
- комбинированная.

При создании тезауруса «Проектирование самолёта» использовалась комбинированная стратегия.

Для создания информационно-поискового тезауруса существует стандарт [8], который устанавливает правила разработки, структуру, состав и форму

представления тезаурусов, разрабатываемых в рамках автоматизированных информационных систем и сетей научно-технической информации.

Основной единицей тезауруса являются *термины*, которые разделяются на *дескрипторы* (авторизованные термины) и *аскрипторы* (не дескрипторы, квази-или относительные синонимы). *Термин* – это слово или словосочетание, предпочтительно в форме существительного или именной группы, являющееся точным обозначением определённого понятия какой-либо области знания [8]. Таким образом, понятие предметной области имеет несколько возможных вариантов лексического представления в тексте, которые рассматриваются как *синонимы*. Среди синонимов выбирается *дескриптор* – термин, который рассматривается как основной способ ссылки на понятие в рамках тезауруса. Другие термины из синонимического ряда, включённые в тезаурус, являются *аскрипторами*. Они используются как вспомогательные элементы, текстовые входы, помогающие найти подходящие дескрипторы [4]. *Дескрипторы* тезауруса соответствуют выбранной предметной области тезауруса. Каждый дескриптор, внесённый в тезаурус, представляет отдельное понятие данной области.

При создании тезауруса учитывалось, что набор дескрипторов должен удовлетворять следующим требованиям:

- возможностью описывать темы абсолютного большинства текстов предметной области;

- для уменьшения субъективности индексирования множество дескрипторов не должно включать совокупности близких дескрипторов, для этого формируются классы условной эквивалентности, когда совокупности близких, но различных понятий сводятся к одному дескриптору;

- дескриптор должен быть сформулирован однозначно, его подразумеваемое в рамках тезауруса значение должно быть понятно пользователю. Если однозначный и ясный дескриптор подобрать не удаётся, то термин, взятый в качестве дескриптора,

снабжается релятором (краткой пометкой) или комментарием [4].

В соответствии с [8], основными типами отношений, отражаемыми в тезаурусах, являются следующие:

- род - вид;
- часть - целое;
- причина - следствие;
- сырьё - продукт;
- административная иерархия;
- процесс - объект;
- функциональное сходство;
- процесс - субъект;
- свойство - носитель свойства;
- антонимия.

Такие содержательные типы связей между дескрипторами чаще всего не отражаются в подробном перечне отношений тезауруса, а записываются с помощью небольшого набора отношений, которые обычно разделяются на два типа: *иерархические* и *ассоциативные*.

При создании тезауруса необходимо выбрать такой тип отношений между терминами, который бы мог охватить всю предметную область. При несоблюдении этого условия возникнет необходимость вводить в тезаурус «лишние» ассоциативные связи, то есть устанавливать между терминами, которые логически связаны иерархией наследования, вынужденные отношения ассоциации. С целью выбора типа иерархических отношений были созданы несколько пробных версий тезауруса, состоящих из 50 терминов. После проведения сравнительного анализа было принято решение использовать для дальнейшего построения тезауруса отношение «Род-Вид», при котором количество ассоциативных связей оказалось минимальным.

По ГОСТу [8] родо-видовая связь устанавливается между двумя дескрипторами, если объём понятия нижестоящего дескриптора входит в объём понятия вышестоящего дескриптора. Например, «крыло» и «оперение» являются одним

из видов «несущих поверхностей»; «несущие поверхности», в свою очередь, являются одним из видов «конструктивных частей самолёта», а «конструктивные части самолёта» – одним из видов «конструкции» и т.д. (рис.1¹). На рис.1 показана эквивалентная связь между терминами «фюзеляж» и «корпус самолёта».

Одним из важных результатов создания тезауруса в Protégé является его синтаксическое представление на языке OWL (Ontology Web Language).

Структура OWL-онтологии имеет заголовок и тело. В заголовке содержится информация о самой онтологии (версия, примечания). За заголовком следует тело онтологии, содержащее описание классов, свойств и экземпляров. В OWL-онтологии класс обозначается как *owl:Class*. Для организации классов в иерархию используется свойство *rdfs:subClassOf*, например, класс – «параметры», подкласс класса «параметры» – «параметры самолёта», подкласс класса «параметры самолёта» – «параметры двигателя» и т.д. (рис. 2). Термины в тексте языка OWL расположены в алфавитном порядке.

Особое место занимают два взаимодополняющих класса – *owl:Thing* и *owl:Nothing*. Первый из них является надклассом любого класса OWL, второй – подклассом любого класса OWL. Экземпляр любого класса OWL входит в класс *owl:Thing*. Класс *owl:Nothing* является пустым множеством [6].

OWL-класс может быть описан идентификатором класса (URI), этот способ определяет именованный класс OWL. Описания класса являются строительными блоками для определения классов посредством аксиом [5]. Например, простейшая аксиома, определяющая именованный класс, записывается в виде:

owl:Class rdfs:ID="Самолёт"

и постулирует существование класса с именем «Самолёт».

¹ Представленные в статье рисунки получены с помощью средств редактора Protégé.

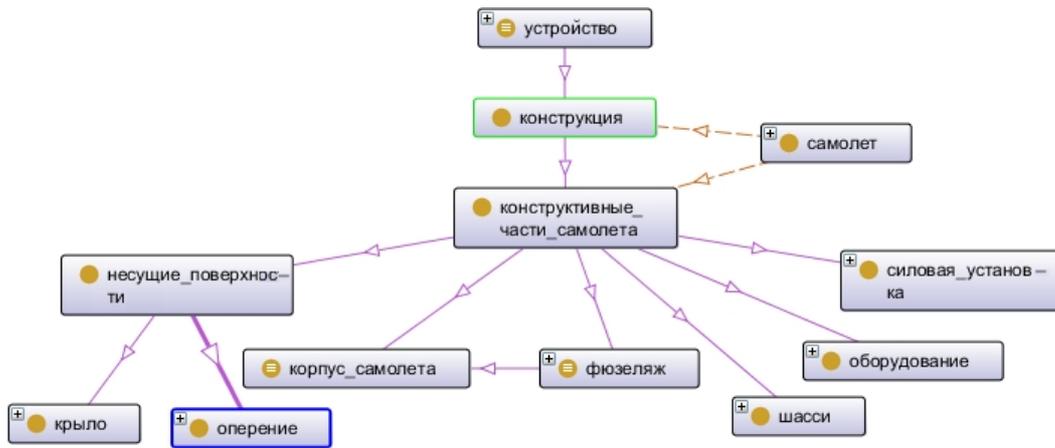


Рис. 1. Установление иерархической родо-видовой связи между терминами

```

Declaration(Class(:параметры))
Declaration(Class(:параметры_двигателя))
SubClassOf(:параметры_двигателя :параметры_самолета)
Declaration(Class(:параметры_крыла))
SubClassOf(:параметры_крыла :параметры_самолета)
Declaration(Class(:параметры_подобия))
EquivalentClasses(:параметры_подобия :критерии_подобия)
SubClassOf(:параметры_подобия :параметры)
Declaration(Class(:параметры_полёта))
SubClassOf(:параметры_полёта :параметры)
Declaration(Class(:параметры_профиля))
SubClassOf(:параметры_профиля :параметры_крыла)
Declaration(Class(:параметры_самолета))
SubClassOf(:параметры_самолета :параметры)
    
```

Рис. 2. Фрагмент организации классов в иерархию на языке OWL

В OWL определены конструкции, комбинируя которые, можно определять более сложные аксиомы классов:

– *rdfs:subClassOf*

говорит о том, что один класс (подкласс) полностью входит в другой (надкласс), например, на языке OWL строка

SubClassOf(:шасси

:конструктивные_части_самолёта),

означает, что шасси является конструктивной частью самолёта;

– *owl:equivalentClass*

говорит о том, что два класса совпадают, например, на языке OWL строка

EquivalentClasses(:фюзеляж :корпус_самолёта),

означает, что термины фюзеляж и корпус самолёта являются синонимичными.

Отношение ассоциации является не-иерархическим и наиболее трудно определяемым. Российский стандарт на создание тезауруса указывает, что ассоциативное отношение является объединением отношений, не входящих в иерархические отношения или в отношения синонимии.

Допускается включать в ассоциативное отношение все виды отношений, кроме синонимии и отношения «род-вид» [8]. Основное назначение установления ассоциативных отношений между дескрипторами тезауруса – указание на дополнительные дескрипторы, полезные при индексировании или поиске.

Например, «самолёт» - «конструкция», «самолёт» - «двигатель», «самолёт» - «параметры» и т.д. Между этими терминами нельзя установить иерархические родо-видовые отношения (например, «конструкция» не является одним из видов «самолёта»), и они не являются синонимами, но состоят в отношениях принадлежности: «самолёт» имеет «конструкцию», «самолёт» имеет «двигатель», «самолёт» имеет «параметры» и т.д. В таком случае устанавливаются ассоциативные связи. В Protégé визуализация отношений ассоциации отображается пунктирными линиями, а стрелки указывают направление, в котором связаны термины (рис. 3).

Ассоциативное отношение между двумя дескрипторами стоит устанавливать, если при употреблении одного термина другой термин подразумевается. Один термин может быть необходимым элементом определения другого термина. Например, термин «конструкция» составляет необходимую часть определения

термина «самолёт».

Отношения ассоциации включались в тезаурус только тогда, когда не было возможности установить иерархические отношения или отношения синонимии.

Случаи, когда необходимо установить ассоциативные связи, могут быть следующими:

- дескрипторы принадлежат одной иерархии (например, «конструктор» участвует в процессе «проектирования», а «проектировщик» обеспечивает процесс «проектирования», и при этом термины «конструктор» и «проектировщик» являются эквивалентными (рис. 4)) или случай, когда термины не являются эквивалентными, но во многих контекстах являются взаимозаменяемыми, такие как «самолёт» и «летательный аппарат»;

- дескрипторы принадлежат разным иерархиям (например, «летательный аппарат» - «деталь», «летательный аппарат» - «сборочная единица», «летательный аппарат» - «полёт» и т.д. В таком случае устанавливаются несколько типов ассоциативных связей: ассоциативные связи принадлежности, например, «летательный аппарат» имеет «деталь», «летательный аппарат» имеет «сборочную единицу»; функциональные ассоциативные связи, например, «летательный аппарат» предназначен для «полёта» (рис. 5)).

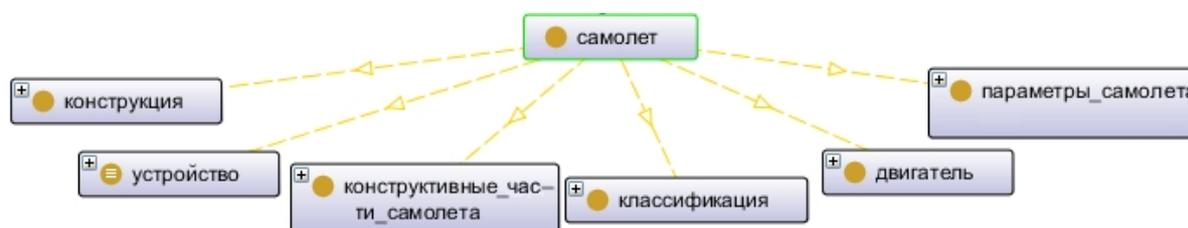


Рис. 3. Установление ассоциативных связей принадлежности между терминами

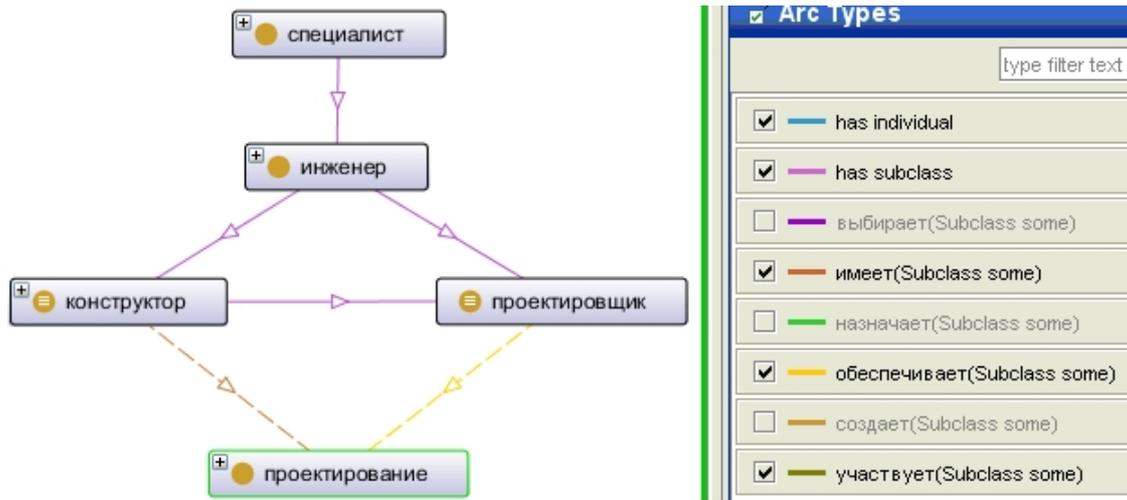


Рис. 4. Пример ассоциативной связи при принадлежности дескрипторов к одной иерархии

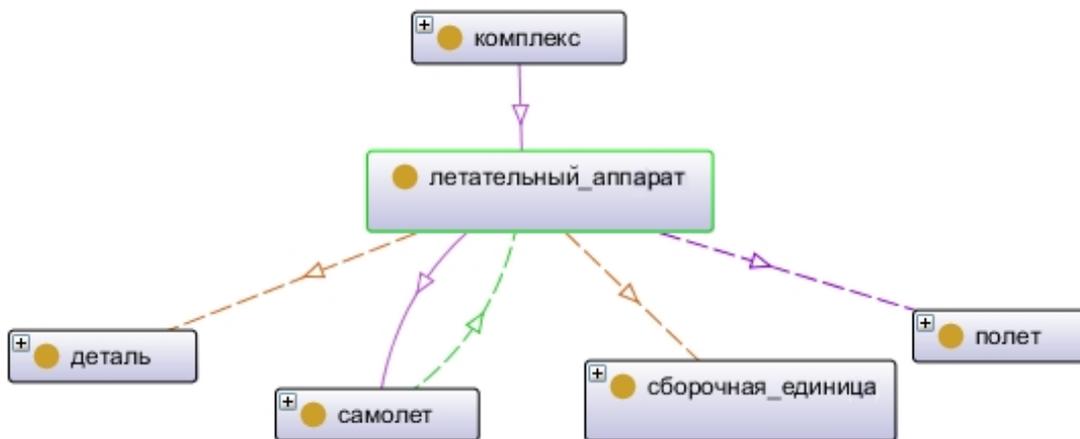


Рис. 5. Пример ассоциативной связи при принадлежности дескрипторов к разным иерархиям

С использованием специально подготовленного словаря рассматриваемой предметной области [9], а также на основе энциклопедических и толковых словарей к настоящему моменту в тезаурус «Проектирование самолёта» включены 15 видов ассоциативных связей (рис. 6). Из них можно выделить несколько типов:

– ассоциативные связи принадлежности: «имеет»;

– ассоциативные функциональные связи: «выполняет преобразование», «обеспечивает», «предназначен для», «разрушает», «создаёт», «назначает», «выбирает», «производится», «вызывает», «участвует»;

– ассоциативные характеристические связи: «характеризует», «показывает», «представляет собой», «является».

- выбирает
- вызывает
- выполняет_преобразование
- имеет
- назначает
- обеспечивает
- показывает
- предназначен_для
- представляет_собой
- производится
- разрушает
- создает
- участвует
- характеризует
- является

Рис. 6. Типы ассоциативных связей в тезаурусе «Проектирование самолёта»

В OWL выделяют две категории свойств: *свойства-объекты* (или объектные свойства), которые связывают между собой *индивиды* (экземпляры классов), и *свойства-значения*, которые связывают индивиды со значениями данных. Оба класса свойств являются подклассами класса *rdf:Property*. Для определения новых свойств как экземпляров *owl:ObjectProperty* или *owl:DatatypeProperty* используются аксиомы свойств, например,

```

Declaration
(ObjectProperty (:обеспечивает)),
.....
SubClassOf(:вертикальное_оперение
ObjectSomeValuesFrom
(:обеспечивает :путевая_устойчивость)).
    
```

Последняя аксиома постулирует существование некоторого свойства «обеспечивает», связывающего экземпляры класса *owl:Thing*, такие как «Вертикальное оперение» и «Путевая устойчивость», друг с другом.

После установления иерархических и ассоциативных связей и проверки этой структуры Protégé даёт возможность расширить границы тезауруса включением в него данных или *Индивидов* (Экземпляров классов). *Индивиды* – это данные базы знаний, которые определяются при помощи аксиом индивидов, например, «Боинг 747» является экземпляром класса «Пассажирский самолёт».

Программы, сгенерированные редактором Protégé, можно использовать и в других редакторах онтологий, имеющих возможность работы с OWL.

Одной из особенностей онтологического редактора Protégé является возможность просмотра созданного тезауруса в виде семантической сети (рис.7).

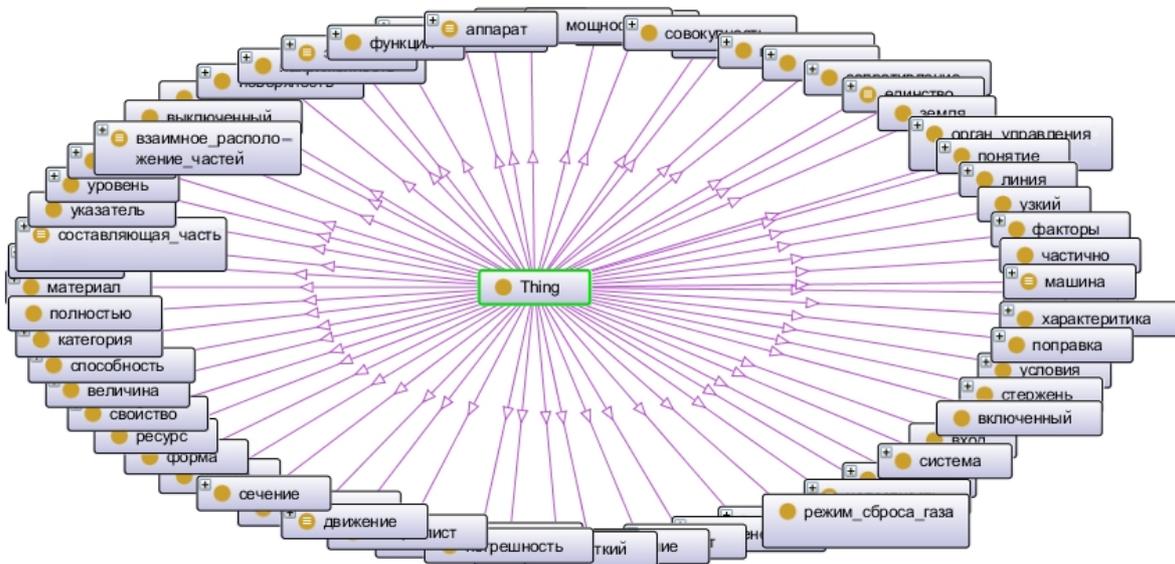


Рис. 7. Семантическое представление тезауруса «Проектирование самолёта»

Знак «+» в верхнем левом углу блока с термином означает, что данный класс (термин) содержит подклассы, состав и связи в которых можно увидеть, нажав на интересующий блок.

Тезаурус «Проектирование самолёта» содержит 644 термина и непрерывно расширяется.

При создании тезауруса конкретной предметной области необходимо использовать не только тексты данной предметной области, но и обращаться к энциклопедиям, терминологическим словарям и справочникам для логического анализа определений терминов, консультироваться со специалистами выбранной предметной области для создания достоверных связей.

Поскольку тезаурус для проектирования самолёта является инновационным, то не существует определённых правил в выборе типа иерархических связей или описания ассоциативных отношений. Предложенные принципы установления связей опираются на ГОСТ [8], а также на опыт специалистов [4].

Библиографический список

1. Боргест, Н.М. Тезаурус как основа онтологии концептуального проектирования самолёта [Текст] / Н.М. Боргест, Д.В. Шустова // Труды Международной конференции с элементами научной школы для молодежи : сб. науч. тр.; Самарский гос. аэрокосм. ун-т им. академика С.П. Королёва (национальный исследовательский ун-т). – Самара: Изд-во Самар. гос. аэрокосм. ун-та, 2010, – С. 353-356.
2. Боргест, Н.М. Автоматизация предварительного проектирования самолёта [Текст] : [учеб. пособие] / Н.М. Боргест. – Самара: Изд-во Самар. авиац. ин-та, 1992. – 92 с.
3. Гаврилова, Т.А. Базы знаний интеллектуальных систем [Текст] : [учеб. пособие для студентов вузов] / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб: Изд-во «Питер», 2001. – 384 с.
4. Лукашевич, Н.В. Тезаурусы в задачах информационного поиска [Текст] / Н.В. Лукашевич. – М.: Изд-во Московского университета, 2011. – 512 с.
5. Клещев, А.С. Математические модели онтологий предметных областей. Существующие подходы к определению понятия «онтология». Часть 1 [Текст] / А.С. Клещев, И.Л. Артемьева // Научно-техническая информация. Сер. 2. – 2001. – № 2. – С. 20-27.
6. Муромцев, Д.И. Онтологический инжиниринг знаний в системе Protege [Текст] : учеб.-метод. материалы / Д.И. Муромцев. - СПб: СПбГУ ИТМО, 2007. - 62 с.
7. Шустова Д.В. Разработка тезауруса прикладной онтологии [Текст] / Д.В. Шустова / Аспирантский вестник Поволжья. Научно-информационный межвузовский журнал, – 2010. – Вып. 1-2, – С.108-110.
8. ГОСТ 7.25-2001. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления [Текст]. – Введ. 2002-07-01. – М.: Изд-во стандартов, 2001. – 13 с.
9. Боргест, Н.М. Краткий словарь авиационных терминов [Текст] / Н.М. Боргест, В.А. Комаров, А.И. Данилин. – М.: Изд-во МАИ, 1992. – 224 с.

HIERARCHICAL AND ASSOCIATIVE RELATIONS BETWEEN THE TERMS IN THE THESAURUS IN A DESIGNER'S DICTIONARY TAKEN AS AN EXAMPLE

©2012 N. M. Borgest, D. V. Shustova, S. R. Gimatdinova

Samara State Aerospace University named after academician S. P. Korolyov
(National Research University)

The paper is devoted to the problems of creating the thesaurus of the domain «Aircraft design. The stage of the development of technical proposals». The kinds of relations between the terms in the thesaurus and the principles of creating them are described. The focus is on the creation of associative relations as the most controversial ones which require scientific substantiation. Examples of syntactic representation of the thesaurus in the OWL language are given.

Domain ontology, thesaurus, aircraft design, group relations, hierarchy, semantic network, syntactic representation.

Информация об авторах

Боргест Николай Михайлович, кандидат технических наук, профессор кафедры конструкции и проектирования летательных аппаратов, Самарский государственный аэрокосмический университет имени академика С. П. Королёва (национальный исследовательский университет). E-mail: borgest@yandex.ru. Область научных интересов: онтология, автоматизация проектирования.

Шустова Дина Владимировна, аспирантка кафедры конструкции и проектирования летательных аппаратов СГАУ. E-mail: ShustovaDV@yandex.ru. Область научных интересов: онтология проектирования.

Гиматдинова Светлана Рашитовна, студентка СГАУ. E-mail: rashitta@mail.ru. Область научных интересов: онтология проектирования.

Borgest Nikolay Mikhailovich, candidate of technical science, professor, the department of aircraft design, Samara State Aerospace University named after academician S.P. Korolyov (National Research University). E-mail: borgest@yandex.ru. Area of research: ontology, computer-aided design.

Shustova Dina Vladimirovna, postgraduate student, the department of aircraft design, Samara State Aerospace University named after academician S.P. Korolyov (National Research University). E-mail: ShustovaDV@yandex.ru. Area of research: design ontology.

Gimatdinova Svetlana Rashitovna, undergraduate student, Samara State Aerospace University named after academician S.P. Korolyov (National Research University). E-mail: rashitta@mail.ru. Area of research: design ontology.