



УДК 811.93

СЕМАНТИЧЕСКИЙ АНАЛИЗ ТЕКСТОВОЙ ЧАСТИ РУКОВОДСТВА ПОЛЬЗОВАТЕЛЯ ПО СБОРКЕ ДВИГАТЕЛЯ С ПОМОЩЬЮ ОНТОРЕДАКТОРА FLUENT EDITOR

Орлова А.А., Боргест Н.М.

*Самарский государственный аэрокосмический университет им. академика С.П. Королева
(национальный исследовательский университет), г. Самара, Россия*

a.aorlova@yandex.ru

borgest@yandex.ru

В работе рассматривается семантический анализ текстовой части руководства по сборке двигателя 2,0 DOHC-16V, подготовленной на естественном английском языке, и преобразование его в онтологию на контролируемом английском языке (Controlled Natural Languages, CNL). Изучена возможность использования редактора Fluent Editor, в котором онтологии создаются с помощью CNL, как для проверки полноты руководств пользователя, так и для их подготовки. Онтология части руководства пользователя по сборке двигателя представлена тезаурусом используемых профессиональных терминов и понятий рассматриваемой предметной области.

Ключевые слова: Controlled Natural Language, онтология, Fluent Editor, семантический анализ.

Введение

В последнее время все чаще для решения проблем неоднозначности используют естественные контролируемые языки. Создаются онтологические редакторы, использующие для построения онтологий CNL (Controlled Natural Languages), например, польский редактор Fluent Editor. Контролируемый язык – упрощенная версия естественного языка, следовательно, CNL можно использовать в качестве языка для описания технических текстов, так как они имеют более строгие структуру и систему понятий. В свою очередь, редакторы, в основу которых заложен CNL можно использовать в качестве инструмента для подготовки технических текстов (в том числе руководств по сборке).

1. Fluent Editor

1.1. Controlled Natural Language

Контролируемый язык – упрощенная версия естественного языка, созданная путем ограничения грамматики, терминологии и речевых оборотов, чтобы снизить или искоренить многозначность и сложность естественного языка [O'Brien, 2003].

Fluent Editor – редактор от польской компании

Cognitum для всеобъемлющего редактирования сложных онтологий, при создании которых используется естественный контролируемый язык. В Fluent Editor контролируемым естественным языком является английский. CNL может служить для представления знаний и в качестве языка интерфейса систем знаний высокого уровня.

2. Семантический анализ

При написании онтологии на CNL необходимо придерживаться правила контролируемого языка, а также использовать тезаурус описываемой предметной области для искоренения многозначности профессиональных терминов. Поэтому сначала рассмотрен вопрос о выявлении экземпляров, отношений и классов руководства пользователя для преобразования текста из естественного английского языка в онтологию на CNL. Приведена технология преобразования тезауруса используемых профессиональных терминов.

2.1. Алгоритм выявления экземпляров и отношений между ними в руководстве пользователя по сборке двигателя

В первоначальном варианте онтологический редактор не распознает текст, так как он не соответствует грамматике и синтаксису CNL. Окно,

визуализирующее иерархию отношения классов и экземпляров онтологии, – окно Дерева Таксономии (от англ. Taxonomy Tree) остается пустым.

Следовательно, необходимо преобразовать текст на естественном английском в CNL. (рисунок 1).

Алгоритм преобразования описания руководства по сборке двигателя из естественного английского языка в CNL состоит из следующих действий:

1. Создание зависимости экземпляр-экземпляр. Экземпляр - частный вид класса, не имеющий подвидов, но имеющий возможность состоять в отношениях с другими экземплярами. Экземпляры на CNL всегда записываются с заглавной буквы, если имя экземпляра состоит из нескольких слов, то они разделяются дефисами и тогда каждое слово в имени экземпляра записывается с заглавной буквы.

2. Отношения необходимо задать зависимостью «Экземпляр отношение Экземпляр», например, «Step-1 cleaning Engine.». Экземпляр «Step-1» вводится по причине того, что как правило все предложения в руководствах пользователя, употребляются в указательной форме «Сказуемое дополнение», а для задания предложения на CNL необходима форма «Подлежащее сказуемое

дополнение», в данном примере экземпляр «Step-1» является подлежащим. Отношение «cleaning» осталось от предыдущей формы предложения, но так как необходимо в предложении ответить на вопрос «Очистить что?», дополнительно введен экземпляр «Engine».

3. Если встречаются предложения с однородными членами «Measure the diameters of the main and big-end bearing journals using a proprietary micrometer screw.» необходимо провести преобразование типа «Step-2 measure Diameters-Of-The-Main-Bearing and measure Big-End-Bearing.», т.е. описать, что в шаге №2 проводится две подобных операции, союзом «and» обозначается принадлежность двух отношений и экземпляров к одному экземпляру-подлежащему. Произведя пошаговую доработку обычного текста в CNL, получаем онтологию, представленную на рисунке 2.

В Дереве Таксономии в пункте «relation» отображены все, описанные отношения. Для лучшего восприятия и проверки написанного, визуализируем построенную онтологию с помощью встроенного в Fluent Editor инструмента графического представления онтологий – CNL-диаграммы, представленной на рисунке 3.

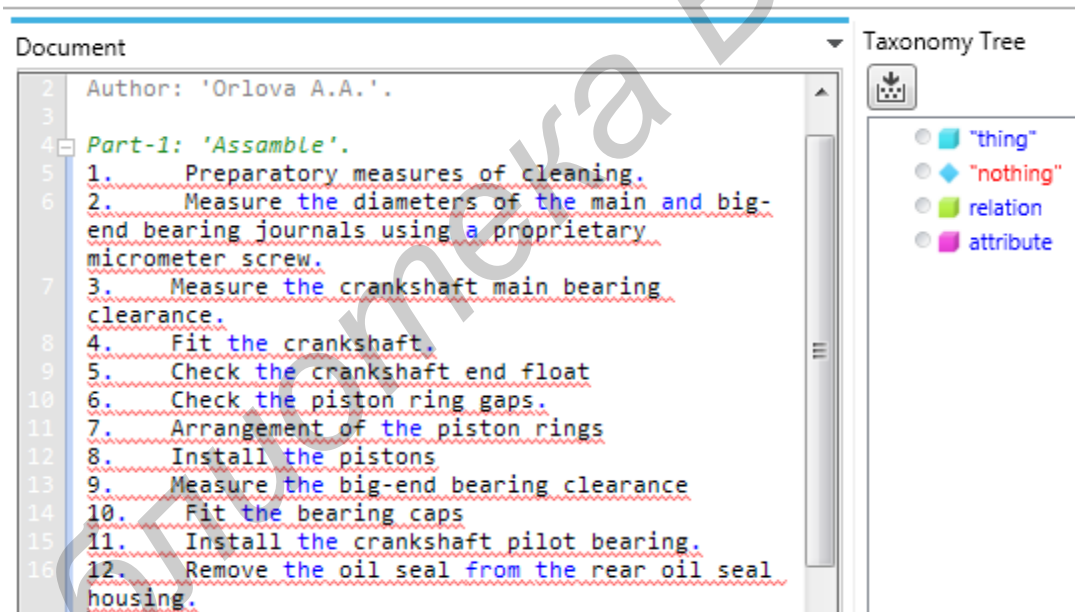


Рисунок 1 – Часть руководства пользователя по сборке двигателя на естественном английском языке

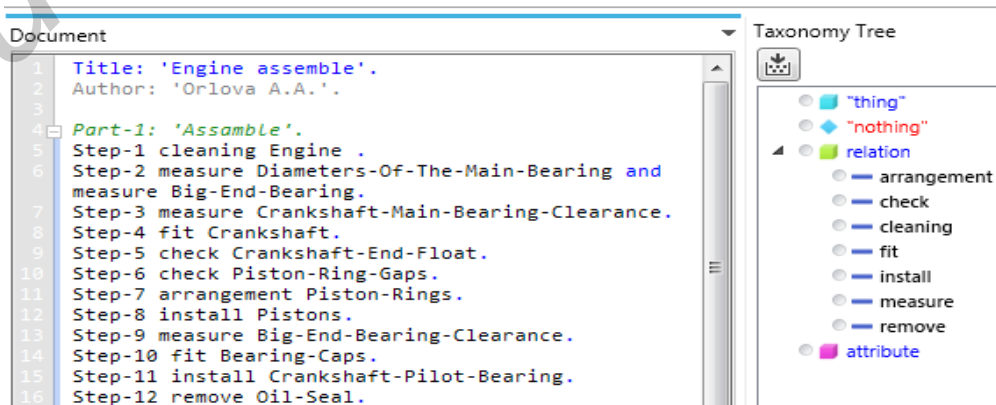


Рисунок 2 – Часть руководства пользователя по сборке двигателя , представленная на CNL, после 1,2 и 3 преобразования

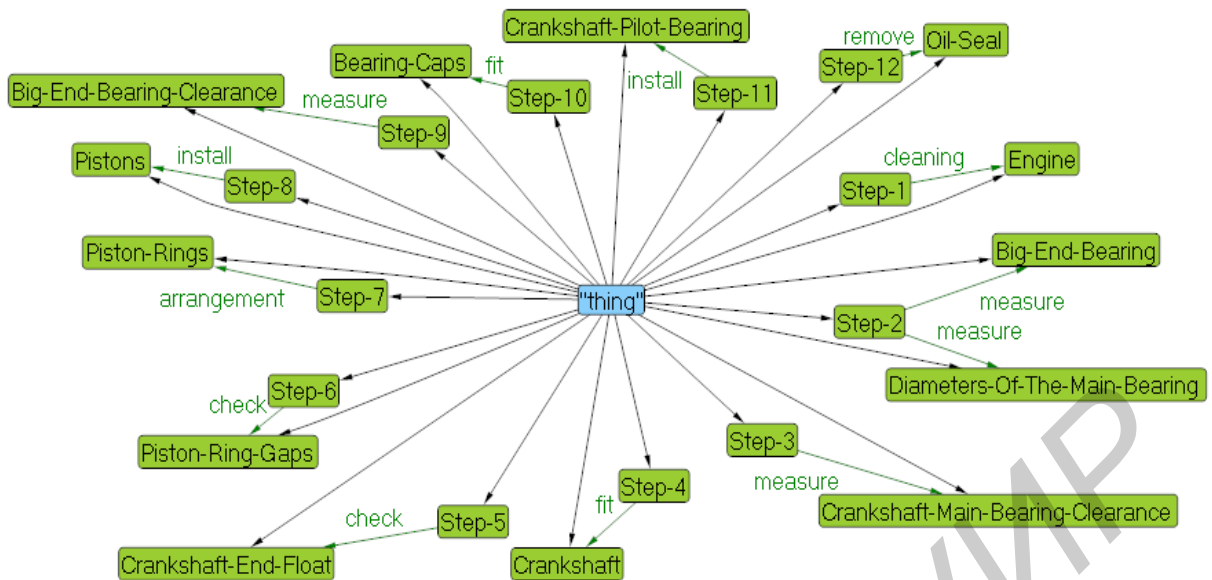


Рисунок 3 – Визуализация части руководства по сборке двигателя, представленная в виде CNL-диаграммы

2.2. Технология преобразования тезауруса к руководству пользователя по сборке двигателя в тезаурус на CNL

Продолжим алгоритм тем, что добавим в онтологию тезаурус профессиональных терминов для искоренения многозначности [Габдрахманов и др., 2012]. Основой многих прикладных онтологий является тезаурус, разработка которого позволяет упорядочить знания о предмете, «договориться» о понимании сути обозначений и понятий, определиться и разобраться с терминологией [Боргест, 2010]. При введении определений терминов в их начальном виде на естественном английском редактор не воспринимает текст, Дерево Таксономии остается пустым (рисунок 4).

4. Выявление ключевых слов в каждом определении, например, в определении понятия «engine» ключевой будет являться фраза «is a machine to convert form of energy», «crankshaft» - «is a mechanical part to perform of motion», «oil seals» -

«is a spaces for helping prevent lubricant escape». Данные ключевые фразы являются основой для написания на CNL, необходимо лишь воспользоваться правилами правописания имен экземпляров и классов. При задании ключевых фраз понятий как классов следует соблюдать несколько правил правописания CNL. Во-первых, пробелы в имени класса так же, как и в имени экземпляра, заменяются на дефисы. Во-вторых, имя класса всегда начинается с маленькой буквы, а если имя класса состоит из нескольких слов, то каждое из них начинается с маленькой буквы (рисунок 5). На рисунке 6 представлена CNL-диаграмма онтологии тезауруса руководства пользователя по сборке двигателя.

Целиком алгоритм последовательного выполнения семантического анализа руководства пользователя по сборке двигателя из текста на естественном английском языке в CNL представлен на рисунке 7 в виде блок-схемы.

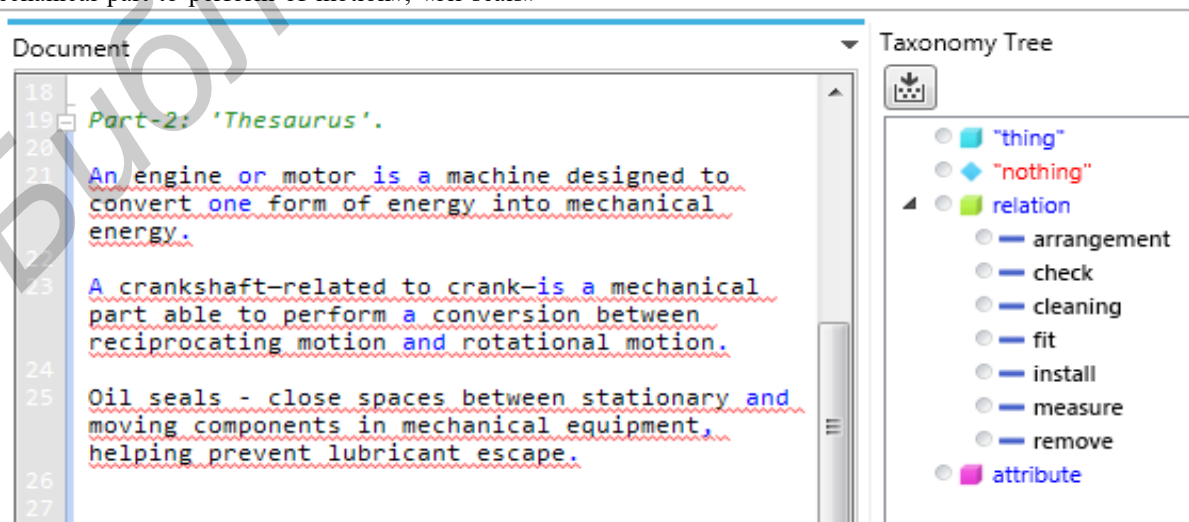


Рисунок 4 – Тезаурус к части руководства пользователя по сборке двигателя на естественном английском языке

Document

```

17
18
19 Part-2: 'Thesaurus'.
20
21 Engine is a machine-to-convert-form-of-energy.
22
23 Crankshaft is a mechanical-part-to-perform-of-motion.
24
25 Oil-Seal is a spaces-for-helping-prevent-lubricant-escape.
26
27
28
29
30
31
32
33

```

Document Document Diagram

Taxonomy Tree

- thing
 - machine-to-convert-form-of-energy
 - Engine
 - mechanical-part-to-perform-of-motion
 - Crankshaft
 - spaces-for-helping-prevent-lubricant-escape
 - Oil-Seal
 - "nothing"
 - relation
 - arrangement
 - check
 - cleaning
 - fit
 - install
 - measure
 - remove
 - attribute

Рисунок 5 – Тезаурус к части руководства пользователя по сборке двигателя, представленная на CNL, после четвертого преобразования

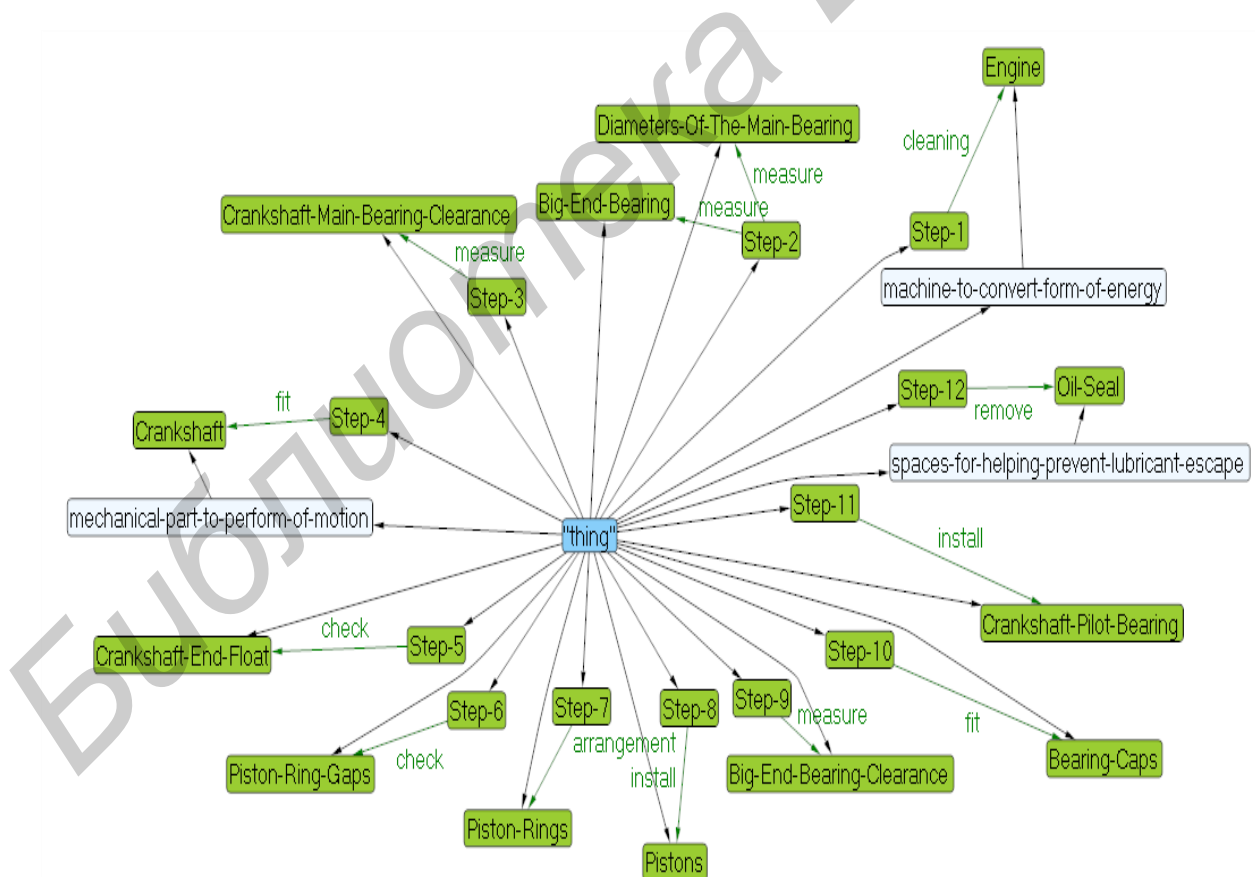


Рисунок 6 – Визуализация тезауруса к части руководства по сборке двигателя, представленная в виде CNL-диаграммы

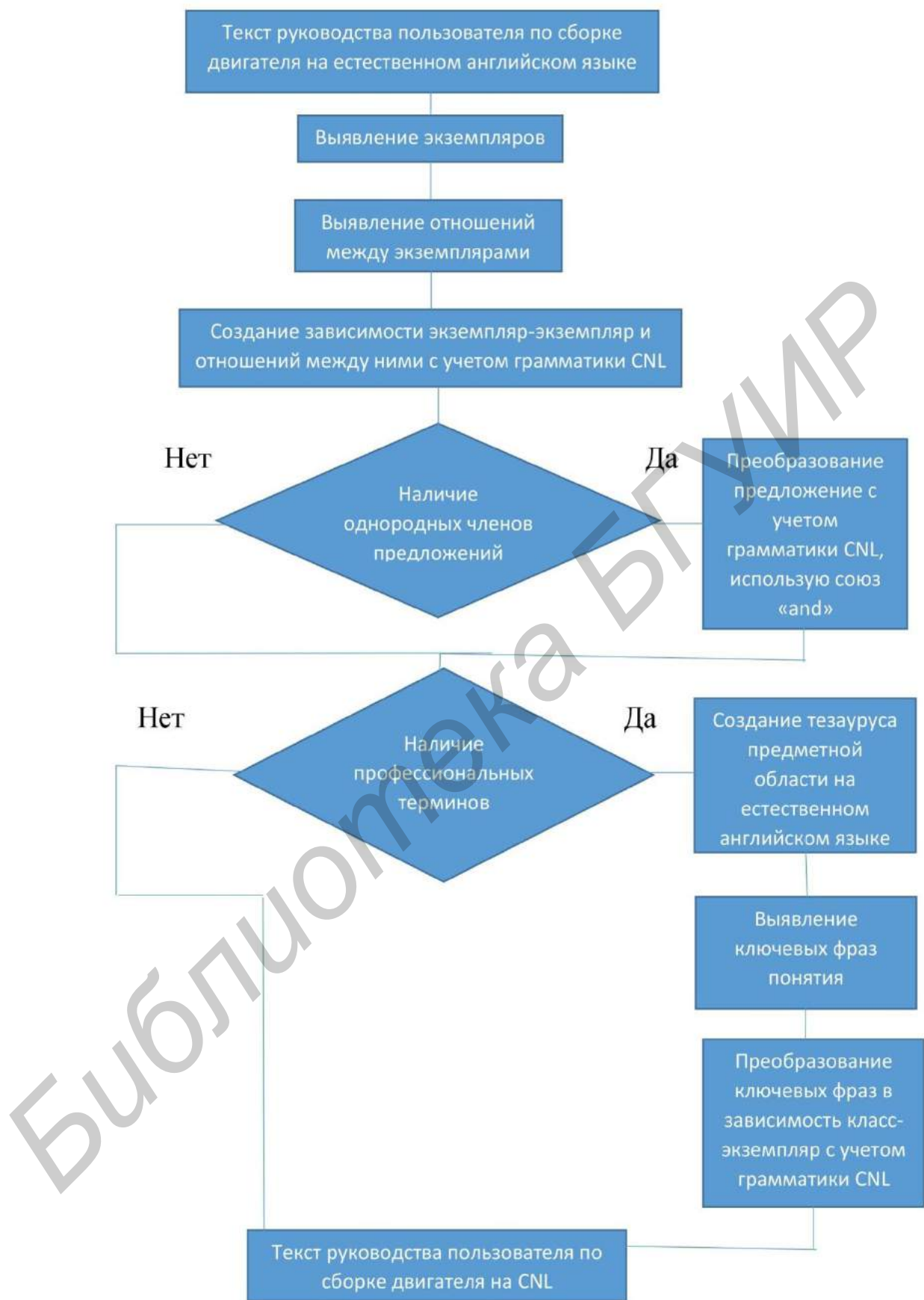


Рисунок 7 – Блок-схема алгоритма преобразования текста из естественного английского языка в CNL

Заклучение

Проведен семантический анализ технического текста на естественном английском для преобразования его в онтологию на CNL, путем некоторых упрощений, но с соблюдением профессиональных терминов. Соблюдение профессиональной терминологии является важным обстоятельством при создании онтологии, так как это язык межперсонального взаимодействия [Ландэ и др., 2014]. Функционал и возможности онтологического редактора Fluent Editor позволяют осуществлять семантический анализ текстов на естественном английском. Следует отметить, что инструменты визуализации редактора удобны для проверки полноты руководств пользователей и отношений в рассматриваемой предметной области.

Библиографический список

[2,0 DOHC-16V engine] 2,0 DOHC-16V engine. Engine Assembly – Dismantle and Assemble (engine removed). Scorpio '95. 09/97 [Электронный ресурс] - <http://www.fordscorpio.co.uk/manual/engines/dohc16v/DOHC16VDismant.pdf>. – с.30

[Боргест, 2010] Боргест Н.М. Онтология проектирования / уч. пособие Самарского государственного аэрокосмического университета, 2010. – с.91

[Габдрахманов и др., 2012] Габдрахманов И.Н., Кучуганов В.Н., Медведев Д.В. и др. Технология генерации онтологии базы данных с помощью онтологического толкового словаря, «Онтология проектирования» №4, 2012. – с.106

[Ландэ и др., 2014] Ландэ Д.В., Снарский А.А. Подход к созданию терминологических онтологий, «Онтология проектирования» №2, 2014. – с. 114

[O'Brien, 2003] Sharon O'Brien Controlling controlled English an analysis of several controlled language rule sets/ School of Applied Language and Intercultural Studies Dublin City University Dublin 9, Ireland. [Электронный ресурс] – <http://mt-archive.info/CLT-2003-Obrien.pdf> - с.10

SEMANTIC ANALYSIS OF TEXT USER MANUAL ENGINE ASSEMBLY WITH ONTOLOGY EDITOR FLUENT EDITOR

Orlova A.A, Borgest N.M.

*Samara State Aerospace University (National
Research University), Samara, Russia*

a.aorlova@yandex.ru
borgest@yandex.ru

In works deals semantic analysis of the engine assembly text in a natural English in the ontology of the Controlled Natural Languages (CNL). The possibility of using the editor Fluent Editor, where ontologies are created by CNL, to verify the completeness of the manuals, as well as for their preparation. Ontology of the manual assembly of the engine filled the thesaurus used by professional terms and concepts.

Introduction

In recent years more and more to address the ambiguity of using controlled natural languages. Created ontology editors, are used to build ontologies CNL, for example, the Polish editor Fluent Editor.

Controlled Natural Language is a simplified version of a natural language. Therefore, CNL can be used as a language for describing the technical texts, as they have a strict structure and system concepts. In turn, the Editors, which laid the basis for CNL used as a tool for the preparation of technical documents (including guidelines on the assembly).

Main Part

When writing ontology CNL must adhere to the rules of the controlled language, and use the thesaurus described domain to eliminate ambiguity professional terms. Therefore, we first address the issue of identifying instances of classes and relations manual for the text of the English language in a natural ontology to CNL. The technology used to simplify the thesaurus professional terms.

The original version of the ontological editor does not recognize the text, since it does not conform to the syntax and grammar of CNL. Window, visualizing the hierarchy relationship of classes and instances of the ontology - Taxonomy tree window (from the English. Taxonomy Tree) remains empty. Therefore, it is necessary to simplify the text in a natural English to CNL.

The conversion algorithm description manual assembly of the natural English in CNL consists of the following:

1. Create an instance and depending on the instance.
2. Creating a relationship between instances.
3. Conversion of homogeneous terms of the proposals.
4. Conversion of the thesaurus to the assembly guide on CNL.

Conclusion

Conducted technical semantic analysis of text in a natural English for its transformation into an ontology at CNL, by some simplifications, but in compliance with the professional terms. Compliance with professional terminology is an important factor in the creation of ontology, as is the language of interpersonal interaction. From the above it can be concluded about the possibilities of ontological editor Fluent Editor serve as a mechanism to semantic analysis texts in natural English. It should also be noted that visualization tools are convenient editor to verify the completeness and uniqueness of user guides and relationships in the subject area.