

АВТОМАТИЗАЦИЯ АНАЛИЗА ТЕКСТА ДОКУМЕНТА «ТЕХНИЧЕСКОЕ ЗАДАНИЕ НА АВТОМАТИЗИРОВАННУЮ СИСТЕМУ»

Сочнев Р.С.

(Рыбинская Государственная Авиационная Технологическая Академия, Рыбинск)

Трудоемкость и значительная доля повторяющихся операций процесса анализа текста технического задания приводит к необходимости и возможности его автоматизации. В данной работе описывается вариант автоматизации анализа текста технического задания (ТЗ) на автоматизированную систему (АС). Анализ выполняется с помощью смысловых (семантических) сетей действий АС (сравнение с образцовой сетью).

Ключевые слова: техническое задание, автоматизированная система, организация баз данных, обработка текста.

1. Введение

Заказчики не всегда обладают высокой квалификацией в области автоматизации, неся при этом ответственность за объект автоматизации в целом. Автоматизация анализа ТЗ может им помочь сосредоточиться на нетиповых требованиях ТЗ с учетом особенностей технологического объекта управления и вопросов обеспечения оптимальной реализации АСУ, освободив от выверки формализуемой части ТЗ.

Основная проблема формализации анализа текста ТЗ состоит в том, что часть компонентов ТЗ содержит информацию, которая по своему характеру является некорректной, что обусловлено отсутствием специалистов высокой квалификации, разнообразием форм передачи мысли естественным языком и сложностью процесса формализации описываемых явлений. Требования ТЗ могут трактоваться по-разному различными специалистами. Очень важным моментом является полнота юридической ответственности разработчика АС при невыполнении каких-либо требований, чтобы можно было бы классифицировать это не как изменение ТЗ, а как недоработка разработчика.

Примеры повторяющихся операций процесса анализа текста технического задания: сравнение с образцом (проверка выполнения требований нормативной документации), проверка прослеживаемости требований во всех частях ТЗ (например, обработка в алгоритмах сигналов из перечня сигналов), проверка наличия параметров требований (например, время цикла работы АС от начала сбора данных о сигналах до выдачи управляющего воздействия, время работы на собственном ИБП (при его наличии)) и т.п.

Для семантического анализа текста ТЗ необходима разработка средства моделирования данных (модели представления данных), которое позволит автоматизировать работу с техническим заданием и обеспечит полный охват предметной области.

2. Краткий обзор опубликованных решений анализа ТЗ

По данной тематике опубликовано несколько статей.

В статье Татьяны Карловой «Формирование образа изделия в предпроектной деятельности» одной из основных проблем разработки ТЗ считается разработка формальной модели представления знаний. Эта модель, рассматриваемая как ядро экспертной системы, должна определять тип используемых знаний, методы их структурирования и способы задания логической связи между ними. Задача анализа ТЗ сводится Татьяной Карловой к формализации знаний эксперта, причем под знаниями понимаются не только голые факты, но и алгоритм их обработки. В данной статье не описана модель представления знаний, т.е. осуществлена только постановка задачи.

В статье Орловой Ю.А. «Грамматика для анализа текста формального документа «техническое задание» описана попытка автоматизировать разработку технического задания с помощью семантического анализа. Для этого предлагается создание унифицированной структуры текста технического задания, то есть создание такой грамматики, которая позволит наиболее полно отобразить содержимое технического задания.

Орловой Ю.А. разработаны следующие требования к грамматике:

возможность представления неточной информации и гибкость - грамматика должна позволять осуществлять процесс обработки ТЗ с минимально возможными потерями.

Для построения грамматики в вышеуказанной статье предложено использование нечетких порождающих атрибутивных грамматик. Описание атрибутивной грамматики состоит из раздела описания атрибутов и раздела правил. Раздел описания атрибутов определяет состав атрибутов для каждого символа грамматики и тип каждого атрибута. Правила состоят из синтаксической и семантической части. Семантическая часть правила состоит из локальных объявлений и семантических действий. В качестве семантических действий допускаются как атрибутивные присваивания, так и составные операторы. Атрибутивная грамматика ТЗ представляет собой формализм, описывающий структуру ТЗ, и является главным элементом, предназначенным для преобразования данных из «сырого текста» ТЗ к виду списка предложений ограниченного естественного языка и дальнейшего его разбора.

Фактически грамматика ТЗ используется для разбиения исходного текста документа на разделы и обработки наиболее важных из них для анализа ТЗ. Для этого Орлова Ю.А. требует четкое соблюдение структуры документа. ТЗ в ее представлении представляет собой структурированный текст, состоящий из последовательности заранее заданных разделов. Но жесткая структура текста ТЗ противоречит пункту 2.2 ГОСТ 34.602 «В зависимости от вида, назначения, специфических особенностей объекта автоматизации и условий функционирования системы допускается оформлять разделы ТЗ в виде приложений, вводить дополнительные, исключать или объединять подразделы ТЗ».

Отнесение разделов ТЗ к структуре формализованного ТЗ Орлова Ю.А. предлагает выполнить с помощью ключевых слов, что неэффективно из-за устаревших стандартов на термины и определения в области автоматизации и богатство синонимов в русском языке.

3. Повторяющиеся элементы ТЗ

Существуют ли повторяющиеся элементы у разных ТЗ на АС (за исключением букв и т.п.)? Для ответа на этот вопрос рассмотрим пункты «назначение» трех технических заданий на автоматизированные системы, управляющие газотурбинным агрегатом (ГТА) или газотурбинной установкой (ГТУ):

1. Программно-технический комплекс (ПТК) предназначен для выполнения функций автоматического управления, регулирования, контроля и защиты ГТА, состоящего из ГТУ (используемой в качестве привода) и турбогенератора. Конкретное исполнение ПТК определяется Перечнем сигналов ввода/вывода, техническим описанием алгоритмов функционирования, таблицей предупредительных и аварийных уставок и вводятся отдельными утверждаемыми Приложениями к данному ТЗ.

2. САУ предназначена для реализации функций:

- автоматическое, автоматизированное и ручное управление ГТА;
- автоматическое регулирование и поддержание заданных режимов ГТА;
- защита систем ГТА при аварийных ситуациях;
- контроль параметров ГТА, аварийная и предупредительная сигнализация отклонений их от нормы;
- контроль технологической части ГТА и индикация положения исполнительных механизмов;
- диагностирование технического состояния ГТА;
- представление информации о ходе технологического процесса, состоянии систем ГТА и САУ ГТА.

3. САУ ГТУ предназначена для:

- автоматического и ручного управления ГТУ;

- автоматического регулирования заданных режимов ГТУ;
- защиты систем ГТУ при аварийных ситуациях;
- контроля параметров, сигнализации отклонений их от нормы и индикации положения исполнительных органов (ИО);
- диагностики и тестирования систем ГТУ.

Во всех приведенных пунктах присутствуют следующие слова: управление, регулирование, контроль. В пунктах 2 и 3 присутствуют слова «защита», «диагностика». Т.е. в ТЗ на однотипные технологические объекты управления (ТОУ) существуют одни и те же элементы. Этими одинаковыми элементами ТЗ являются действия автоматизированных систем (управление, регулирование, контроль, защита, диагностика и т.п.).

Примечание: наименование и состав АС могут быть разными.

Анализ ТЗ может быть автоматизирован благодаря наличию одинаковых элементов ТЗ - действий автоматизированной системы.

4. Структура повторяющихся элементов ТЗ

Рассмотрим пункты « требования к управлению» трех технических заданий на автоматические системы, управляющие газотурбинным агрегатом (ГТА) или газотурбинной установкой (ГТУ):

1. Функции управления ПТК:

- автоматическая реализация статических и динамических режимов (перевод ГТА из одного статического режима в другой по заданному алгоритму), по команде оператора (в местном режиме управления), по команде АСУ ТП (в дистанционном режиме управления) или при срабатывании ограничительной уставки или аварийной защиты;
- автоматическая проверка предпусковой готовности ГТА;
- автоматический контроль времени выдачи команд на исполнительные механизмы, имеющие сигнализаторы положения - на всех режимах работы ГТА;
- блокировка несанкционированных команд оператора;
- функции по п. 4.6 Правил технической эксплуатации электрических станций и сетей Российской Федерации.

2. САУ ГТА должна осуществлять:

- автоматический пуск и вывод ГТД на заданную частоту вращения по технологической программе;
- управление маслосистемами ГТД и ТГ;
- управление элементами механизации компрессора ГТД;
- изменение технологического режима по командам от ПУ или АСУ ТП;
- управление пусками, остановами и работой под нагрузкой без вмешательства оператора;
- включение/выключение систем и агрегатов ГТД и ТГ:

электромаслонасоса ТГ, вентиляторов, нагревателей, противообледенительной системы по заданной программе.

3. В рабочем режиме САУ ГТУ должна выполнять функцию управления реализуя следующие задачи:

- пуск ГТД и вывод на режим заданной нагрузки по технологической программе;
- управление пусками, остановами и работой под нагрузкой без вмешательства оператора;
- управление режимом охлаждения турбогенератора;
- управление по заданной программе УТО;
- включение противообледенительной системы по заданной программе;
- управление вентиляцией по сигналу загазованности.

В вышеприведенных пунктах присутствуют следующие действия автоматизированной системы: пуск, изменение режима, включение/выключение.

Данные действия являются понятиями, входящими в понятие «управление», т.е. существует структура действий автоматизированной системы.

5. Этапы анализа ТЗ

В данной работе предлагаются следующие этапы анализа ТЗ

1. поиск действий АС в тексте ТЗ;
2. определение структуры действий АС по тексту ТЗ;
3. сравнение структур действий анализируемого и эталонного ТЗ.

6. База данных для анализа ТЗ

Основной проблемой разработки базы данных для анализа ТЗ является выбор средства моделирования данных.

6.1 Выбор средства моделирования данных

К основным средствам моделирования данных относятся:

- логические модели;
- продукционные модели;
- сетевые модели;
- фреймовые модели [5].

Логические модели имеют ясную формальную семантику, что ограничивает их использование для анализа текста на естественном языке, который не имеет формальной семантики.

При большом числе продукций в продукционной модели усложняется проверка непротиворечивости системы продукций, т.е. множества правил. Поэтому число продукций, с которыми работают современные системы ИИ, как правило, не превышают тысячи.

Часть специалистов считают, что нет необходимости выделять фреймовые модели представления знаний, так как в них объединены все основные особенности моделей остальных типов.

Иерархические модели, представляя собой вид семантической сети, с одной стороны обеспечивают уникальность класса, а с другой стороны классификация может быть выполнена по множеству признаков. Неудобство использования иерархической модели видно на примере УДК.

В семантических сетях существует возможность представлять знания более естественным и структурированным образом, чем в других формализмах.

На основании вышеуказанного в качестве средства моделирования данных были выбраны семантические сети.

6.2 СЕМАНТИЧЕСКИЕ СЕТИ

В самом общем случае семантическая сеть представляет собой граф, вершины которого соответствуют объектам предметной области, а дуги – отношениям между ними. В качестве примера рассмотрим предложения "ТУ-160 является дальним бомбардировщиком" и "Каждый дальний бомбардировщик является самолетом". Они могут быть представлены через семантическую сеть (рис 1). В этом примере используется тип дуг "является".

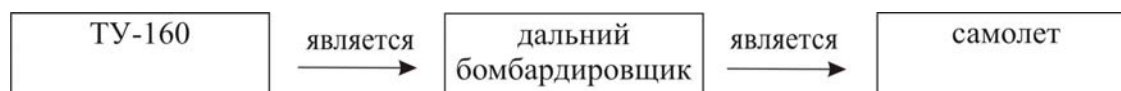


Рис. 1 Пример семантической сети

6.3 МЕТОДИКА АНАЛИЗА ТЕКСТА

6.3.1 Основные понятия

Смысл – внутреннее, логическое содержание (слова, речи, явления), постигаемое разумом, значение [1].

Значение - содержание, связываемое с тем или иным выражением (слова, предложения, знака и т.п.) некоторого языка [2].

Содержание - действие по глаголу содержать [1].

Объект-источник – это основной объект потока действий, который обеспечивает активизацию (запуск) потока действий. Примеры объектов-источников – АС, оператор.

Совместные действия – это действия, образующие общее действие. Например, работаю + работает = работаем.

Не совместные действия – это действия, ослабляющие друг друга.

Поток действий – это вид семантической сети, представляющий собой ориентированный граф, дугами которой являются действия, а узлами – входы/выходы действий.

6.3.2 Метод анализа

Без содержания нет смысла. В данной статье для формализации работы с содержанием и обеспечения однородности данных предлагается использовать действия в качестве основного средства моделирования данных. Возьмем, например, действие «живу». Нет этого действия, и все теряет смысл.

Структура предметной области описывается через построение связей действий. Для этого используются семантические сети.

Существует множество вариантов выполнения того или иного действия. Но большинство связей действий стабильно. Хотя бы часть одинаковых связей действий выполняются однотипными АС. Много ли АСУ ТП, которые не управляют, не регулируют, не сигнализируют, не защищают?

Для защиты от путаницы, возникающей из-за многозначности слов (например, открыть / закрыть окно, дверь, кастрюлю, компьютерный файл), действия рассматриваются по их содержанию (действиям, входящим в них). Для того, чтобы открыть окно или дверь нужно, чтобы их изготовили (резали, шкурили, красили и т.п.), а для открытия компьютерного файла нужно, чтобы его до этого запрограммировали (набрали, скопировали и т.п.).

Для рассматриваемой предметной области разрабатывается эталонная семантическая сеть. Затем выявляется семантическая сеть анализируемого ТЗ. После этого выполняется сравнение эталонной и полученной из ТЗ сетей.

6.4 ПРИМЕНЕНИЕ СЕМАНТИЧЕСКИХ СЕТЕЙ ДЛЯ АНАЛИЗА ТЗ

6.4.1 Правила создания эталонной семантической сети вида «поток действий».

При создании потока действий должны выполняться следующие правила связывания действий:

1. Связь между действиями осуществляется напрямую, а не через другие объекты.
2. Для любого объекта существует как минимум одно действие - «существует» («живет»), с которым непосредственно связывается объект, что обеспечивает единообразие обработки объектов и позволяет установить признак существования объекта.
3. Любое действие в потоке действий должно быть связано хотя бы с одним другим действием или с объектом-источником. Т.е. все действия связаны, нет действия из ниоткуда.
4. Ближайшими к объектам являются те действия, которые их непосредственно создали и в которых эти объекты используются. Действие не заканчивается объектом, т.к. любой объект сам по себе не имеет смысла вне действий. Связь осуществляется через действие «существует» («живет»). Например, программа для компьютера была непосредственно создана действием «программировать».

Для уменьшения количества одинаковых записей предлагается использовать две формы записи связей действий: абсолютная форма записи связей действиями (АФ) и относительная форма записи связей действий (ОФ).

Абсолютная форма представляет собой наследование свойств базового класса остальными действиями. Базовым классом является действие «создавать».

Относительная форма представляет собой запись действий, связанных с объектом-источником, в порядке преобразования ими ресурсов.

Дополнительные правила записи относительной формы:

- Определить объект – источник потока действий.
- Записать основное действие объекта-источника – живет или существует.
- Связывать с основным действием другие действия в порядке преобразования ими ресурсов. Порядок преобразования ресурсов определяется необходимостью для жизни (существования) объекта-источника или назначением объекта-источника. Т.е. если действие создает ресурсы, необходимые не связанным действиям, то оно раньше них связывается с уже связанными действиями.
- Если ресурсов нет, то действия связываются в порядке активизации (выполнения).
- Связь между действиями имеет разную силу. Сила связи определяется количеством активизаций данной связи. Связь с действиями, активизируемыми другими объектами-источниками, слабее связи с действиями, активизируемыми своим объектом-источником.
- Для структурирования потока действий группу действий, связанную с одним действием, можно заменить общим действием, если действия этой группы созданы наследованием от данного общего действия.

Виды связей действий:

Простая связь – связь двух действий между собой.

Кольцевая связь – замкнутая на одно из действий последовательная связь нескольких действий.

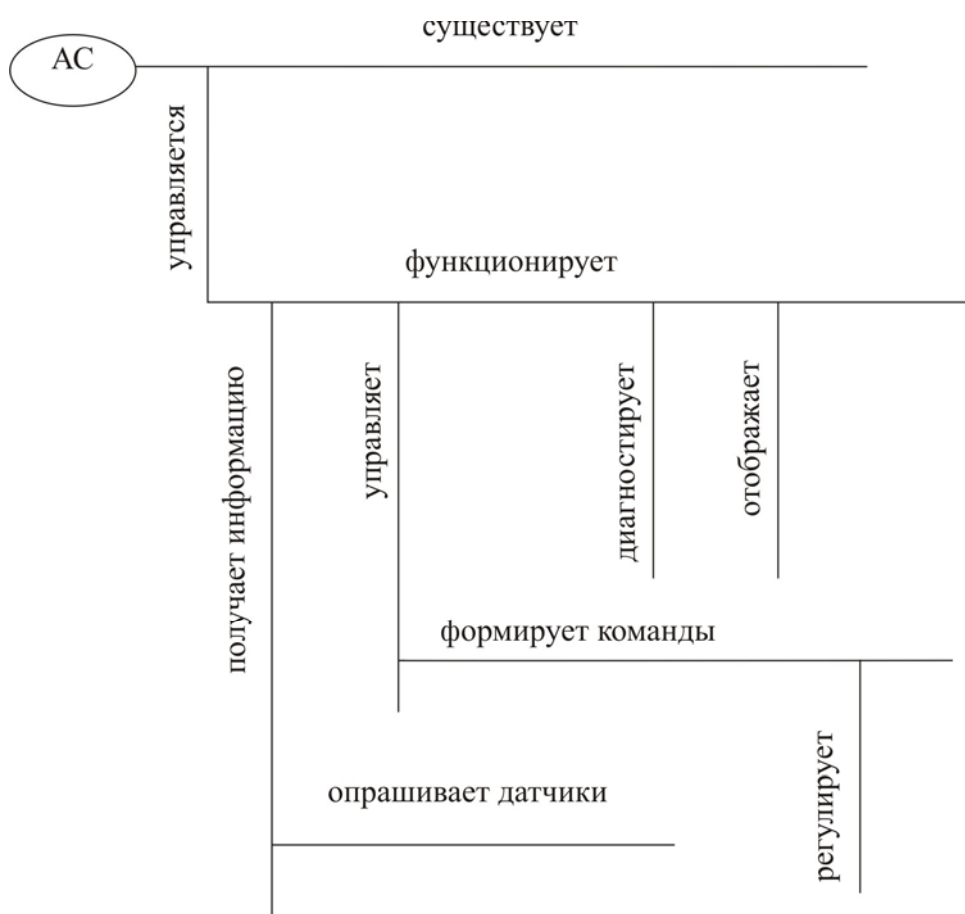


Рис. 2 Пример части связей действий АС



Рис. 3 Шаблон связей действия

6.4.2 Правила преобразования текста ТЗ в семантическую сеть вида «поток действий».

Поиск связей действий в анализируемом ТЗ может быть выполнен следующими способами:

- если есть совпадающие слова в предложениях с этими глаголами;
- если в предложении дается ссылка на другое предложение;
- если предложения находятся в тексте рядом.

Очень часто в тексте ТЗ встречаются существительные, которые могут быть преобразованы в глаголы. Например: «САУ предназначена для выполнения функций автоматического управления, регулирования, контроля и защиты газотурбинного агрегата».

6.4.3 Операции над семантическими сетями вида «поток действий».

Семантическая сеть была реализована с помощью объектно-ориентированного программирования (ООП). Поток действий были реализованы в виде таблиц реляционной базы данных.

Операции над семантической сетью:

копирование, удаление, поиск совпадений, объединение, пересечение, вычитание, отрицание.

6.5 РАБОТА БАЗЫ ДАННЫХ АНАЛИЗА ТЗ (БД)

В БД происходит обработка и накопление вариантов ТЗ. БД хранит отдельные файлы уже разработанных и опробованных на практике частей ТЗ. БД сортирует эти файлы по выбираемой оператором или определяемой автоматически структуре анализируемого ТЗ и на основании анализа требований.

Смысл из файлов БД получает с помощью семантических сетей. Выявление недопустимых сочетаний сигналов выполняется по БД ошибочных действий. БД определяет, какие параметры заданы нетипично.

Самообучение БД выполняется при анализе каждого нового ТЗ. БД выявляет новые связи действий и предлагает оператору включить их в эталонную семантическую сеть.

7. Критерии оценки ТЗ на АС

ГОСТ Р ИСО/МЭК ТО 9294-93 предлагает следующие понятия качества, применимые к содержанию, структуре и представлению документации:

- 1) качество содержания можно измерять в элементах точности, полноты и ясности;
- 2) качество структуры можно измерять легкостью, с которой читатель имеет возможность определить местоположение информации;

3) качество представления должно соответствовать типу проекта. Например, руководство пользователя может иметь форму набора машинописных страниц, скрепленных вместе, или может быть типографской книгой с обширными иллюстрациями, созданной специалистом по графике.

В данной статье дополнительно предлагаются следующие критерии оценки ТЗ:

- прослеживаемость требований во всех частях ТЗ;
- своевременность требований;
- точность параметров требований;
- полнота ответственности разработчика ТЗ;
- реализуемость;
- оптимальность.

Для автоматизации оценки реализуемости и полноты требований использовался метод сравнения с образцами. В качестве образцов были использованы ГОСТ 34.602 и ранее разработанные ТЗ на АС. Примечание: при отсутствии ранее разработанных ТЗ может быть использован шаблон ТЗ.

В соответствии с п.1 Приложения 1 ГОСТ 34.602 проект ТЗ на АС разрабатывает организация-разработчик АС с участием заказчика на основании технических требований (заявки; тактико-технического задания и т. п.).

О полноте ответственности разработчика ТЗ. С одной стороны, п. 1.4.ГОСТ 34.602 гласит: «Задаваемые в ТЗ на АС требования не должны ограничивать разработчика системы в поиске и реализации наиболее эффективных технических, технико-экономических и других решений». С другой стороны, разработчик должен быть лишен возможности схалтурить. Поэтому необходимо автоматизировать проверку выполнения п. 3. 3 ГОСТ 34.602 «Значения показателей, норм и требований указывают, как правило, с предельными отклонениями или максимальным и минимальным значениями. Если эти показатели, нормы, требования однозначно регламентированы НТД, в ТЗ на АС следует приводить ссылку на эти документы или их разделы, а также дополнительные требования, учитывающие особенности создаваемой системы. Если конкретные значения показателей, норм и требований не могут быть установлены в процессе разработки ТЗ на АС, в нем следует сделать запись о порядке установления и согласования этих показателей, норм и требований: "Окончательное требование (значение) уточняется в процессе ... и согласовывается протоколом с ... на стадии ...". При этом в текст ТЗ на АС изменений не вносят».

Для автоматизации оценки оптимальности использовались методы оптимизации графов.

8. Выводы

Синтаксис естественного языка позволяет выделить из текста действия АС. Форматирование текста (на параграфы, абзацы, предложения, списки) позволяет определить структуру действий автоматизированной системы. Формализация определения действий АС и их структуры позволяет автоматизировать анализ ТЗ. Наличие общих элементов ТЗ (действий АС и их структур) позволяет сравнивать ТЗ. Это дает возможность разработки базы данных, выполняющей функции накопления апробированных частей ТЗ и автоматического анализа ТЗ.

Научная новизна заключается в выявлении общих элементов ТЗ -действий АС , в предложении определения структур действий АС и сравнения ТЗ с их помощью.

Литература

1. Толковый словарь русского языка: В 4 Т. / Под ред. Д. Н. УШАКОВА. Т. 1. М., 1935; Т. 2. М., 1938; Т. 3. М., 1939; Т. 4, М., 1940. (Переиздавался В 1947-1948 ГГ.); Репринтное издание: М., 1995; М., 2000.
2. Большая советская энциклопедия. Гл. ред. А.М. Прохоров, 3-е изд. Т. 1-30. М., «Сов.

энциклопедия», 1969-78.

3. Татьяна Карлова «Формирование образа изделия в предпроектной деятельности», журнал «САПР и графика» №7, 2004г.

4. Орлова Ю.А. «Грамматика для анализа текста формального документа «техническое задание», VI Международная научно-методическая конференция “Дистанционное обучение - образовательная среда XXI века“
(<http://www.park.by/education>).

5. Т.А. Гаврилова, В.Ф. Хорошевский «Базы знаний интеллектуальных систем», СПб.: Питер, 2001, 384 с.

ARTICLE TITLE

AUTOMATION of the ANALYSIS of the TEXT of the DOCUMENT "TECHNICAL REQUIREMENT ON AUTOMATED SYSTEM"

Roman Sochnev, state aircraft technological academy, Rybinsk (roman3@rambler.ru).

Abstract: Labour content and significant share reiterative operation process of the analysis of the text of the technical requirement brings about need and possibility to its automations. In given work is described variant to automations of the analysis of the text of the technical requirement (TZ) on automated system (the ACE). The Analysis is executed by means of semantic (semantic) of the networks action ACE (the comparison with exemplary network).

Keywords: the technical requirement, automated system, organization database, processing the text.