

СЕМАНТИЧЕСКИЕ ОСОБЕННОСТИ ТЕХНИЧЕСКОЙ ДОКУМЕНТАЦИИ: ОНТОЛОГИЧЕСКИЙ ВЗГЛЯД НА ПРОБЛЕМЫ КОРРЕКТНОСТИ

Сидорова Елена Анатольевна

К.ф.-м.н., с.н.с., e-mail: lsidorova@iis.nsk.su

Гаранина Наталья Олеговна

К.ф.-м.н., с.н.с., e-mail: garanina@iis.nsk.su

Кононенко Ирина Семеновна

Н.с., e-mail: irina_k@cn.ru

Боровикова Олеся Игнатьевна

М.н.с., e-mail: olesya@iis.nsk.su

Институт систем информатики им. А.П. Ершова СО РАН,
630060 г. Новосибирск, пр. Лаврентьева 6

Аннотация. В работе рассматриваются особенности жанра технической документации с точки зрения извлечения необходимой содержательной информации для анализа корректности спецификаций, поставленных перед разработчиками задач. Отличительной особенностью предлагаемого подхода является использование онтологии как связующего звена между текстом и формальной верификацией. Объекты и процессы, описываемые в тексте, определяются набором онтологий: онтологией требований, онтологией процессов и распределенных систем и онтологией объекта управления. Онтология объекта управления определяется в рамках семантического словаря и включает понятия, описывающие специализированные объекты и характеристики, используемые в конкретной системе управления.

Ключевые слова: семантический анализ текста, извлечение информации, спецификация требований, онтология требований, онтология процессов, онтология автоматизированных систем управления.

Цитирование: Сидорова Е.А., Гаранина Н.О., Кононенко И.С., Боровикова О.И. Семантические особенности технической документации: онтологический взгляд на проблемы корректности // Информационные и математические технологии в науке и управлении. 2018. № 4 (12). С. 30–41. DOI: 10.25729/2413-0133-2018-4-03

Введение. Техническая документация является основой при согласовании, проектировании и разработке программных систем. Наличие неточно сформулированных требований, ошибок в описании сценария работы программы или неполнота описания моделей может привести к значительному снижению качества разрабатываемой системы.

Можно назвать целый ряд исследований, посвященных извлечению информации из текстов технических заданий, преимущественно англоязычных, и их дальнейшему анализу [4,8-9]. Что касается технической документации на русском языке, подобные исследования весьма немногочисленны. Наиболее всесторонне данная проблема для русского языка освещается в [2].

В данной работе рассматриваются особенности жанра технической документации с точки зрения извлечения необходимой содержательной информации для анализа корректности описания системы, относительно требований и спецификаций поставленных перед разработчиками задач. Поскольку рассмотреть все многообразие возможных ошибок в произвольных технических документах представляется затруднительным, мы ограничимся, во-первых, документами, написанными в соответствии с известными стандартами (ГОСТы и др.), во-вторых, будет рассматриваться только ограниченный класс распределённых программных систем, в-третьих, будут рассматриваться ошибки, которые можно распознать средствами формальных логических построений. В качестве экспериментального материала для исследования использовались тексты технических заданий на разработку автоматизированных систем управления (АСУ).

1. Жанровые особенности технической документации. Техническая документация (ТД) — набор документов, используемых при проектировании, создании и использовании каких-либо технических объектов, в частности программного и аппаратного обеспечения. На начальном этапе разработки программного обеспечения АСУ составляется техническое задание и набор проектной документации, в которых формулируются требования к программному обеспечению, стадии и этапы разработки и т.п. Функциональные требования представляются в виде спецификации функций и определения потоков входных и выходных воздействий и могут содержать как явные, так и скрытые из-за сложности описываемых процессов ошибки.

При сборе экспериментального корпуса текстов были выдвинуты следующие базовые требования к ТД: а) наличие в тексте описания архитектуры разрабатываемой системы, которая включает описание модулей, их назначение и взаимосвязи, б) наличие текстового описания сценария (описание работы системы в виде последовательности автоматизируемых процессов и соответствующих этим процессам входных и выходных состояний элементов системы).

Исследование особенностей текстов собранной коллекции показало, что ТД на разработку программной системы АСУ является примером текста с ярко выраженными жанровыми особенностями. Как отмечается в [2], ТД представляет собой текст, написанный техническим языком, в связи, с чем он обладает рядом особенностей, свойственных данному функциональному стилю. Так, текст не содержит образных выражений, оценочных прилагательных, почти отсутствуют наречия, естественная языковая неоднозначность компенсируется использованием предварительно определенных терминов.

В рамках рассматриваемого жанра выделяются такие значимые для задачи проверки требований субжанры, как *Назначение* (описание системы и составляющих ее элементов с точки зрения цели и функций) и *Сценарий* (описание работы системы в виде последовательности автоматизируемых процессов и соответствующих этим процессам входных и выходных состояний элементов системы). К числу лексико-семантических особенностей исследуемых субжанров можно отнести присутствие следующих классов лексических единиц:

- лексика для наименования сущностей (предметов, веществ, технических устройств и их частей, программных продуктов и их компонентов)
- лексика для наименования ситуаций:
 - предикаты состояния (*отсутствие, находится, содержать*);

- событийные предикаты для представления автоматизируемых процессов и действий (*перемещаться, вращение, подача, разогрев, включение, остановка*);
- функциональные предикаты (*использоваться для, обеспечивать*);
- ментальные предикаты (*контролировать, измерять, отслеживать, определять*);
- параметрическая лексика – наименования качественных и количественных параметров (*уровень, положение, температура*), чисел и единиц измерения, лексических наименований эталонных оценок (*низкий/высокий, заданное положение*), предикатов количественного изменения (*падать, расти, нормализоваться, увеличение, разогрев*).

Большинство ситуаций, описываемых в рассматриваемых текстах, и соответствующих лексических наименований не специфичны (универсальны). Это такие ситуации как, а) процессы перемещения сущности в пространстве, количественные изменения параметров (например, температуры), передача информации; б) действия, приводящие систему или ее компоненты в рабочее/нерабочее состояние и деятельностные характеристики, отражающие функциональное назначение системы или ее компонентов; в) состояния, отражающие значения количественных или качественных параметров.

Специфика предметной области естественным образом выражается 1) в покрытии универсального пространства ситуаций с точки зрения типов используемых в текстах ситуаций и их относительной представленности¹ и 2) в конкретных предметных типах участников ситуаций, заполняющих универсальные роли.

Общей чертой всех технических текстов является употребление условных обозначений в качестве имен собственных: *Большой солнечный вакуумный телескоп (БСВТ)*. Специфической особенностью текстов ТД на создание АСУ является употребление уникальных числовых идентификаторов, которые сопоставлены интродуктивным дескрипциям и служат для обозначения референтов объектов, представленных на структурно-топологической схеме: (12) - *датчик температуры жидкости*. Эти идентификаторы используются в составе именных групп на базе полных или неполных дескрипций для повторного упоминания объектов по тексту: *Температура жидкости измеряется аналоговым датчиком (12)*. Использование идентификаторов предотвращает референциальные конфликты и позволяет решить задачу установления кореференции объектов на семантическом уровне.

2. Онтологии проблемной области. Объекты и процессы, представленные в содержательных блоках технических документов, определяются набором онтологий: онтологией объекта управления, онтологией процессов и онтологией требований.

Онтология процессов определяет основные понятия и характеристики технологических процессов, на автоматизацию которых направлены разрабатываемые программные системы. Онтология объекта управления включает понятия, описывающие специализированные объекты и характеристики, используемые в заданном классе систем управления. Данная онтология реализуется в рамках онтологии АСУ. Онтология АСУ специализирует основные понятия онтологии процессов и содержит понятия, классификацию объектов и процессов, типы параметров и т.п., соответствующие известным стандартам и ГОСТам. Онтология требований предназначена для представления выявленных в тексте требований в виде набора условий и ограничений, связанных темпоральными и

¹ Так, в данной предметной области для субжанра *Назначение* характерно более частотное описание свойств объектов, характеризующих их с точки зрения функционального назначения.

каузальными отношениями. Извлеченное и представленное в виде контента онтологии содержание документов транслируется в формулы одной из темпоральных логик.

Семантический словарь является необходимым инструментом для исследования особенностей технической документации и обеспечивает связь предметной лексики и лексико-семантических характеристик с элементами онтологий.

Следует отметить, что онтологии процессов и требований являются универсальными и не зависят от предметной области. Онтология АСУ применима для достаточно широкого класса задач, связанных с разработкой различных распределенных программных систем. Данная система трех онтологий может быть разработана заранее. Для настройки системы верификации на конкретную предметную область (по сути, на конкретный набор технических документов для данной АСУ) необходимо расширить предметный словарь терминами, принятыми в данной предметной области или в текстах конкретных ТД для обозначения понятий данных онтологий.

Рассмотрим подробнее каждый из указанных компонентов базы знаний.

2.1. Онтология процессов. Онтология процессов предназначена для описания функционирования системы, элементы которой могут действовать параллельно. В отличие от существующих онтологий процессов [6,10], моделирующих сложные системы, наша онтология ориентирована на их верификацию. Её формальная семантика описана в [5]. Мы рассматриваем модели систем в духе CSP [7] как параллельное исполнение последовательных процессов, каждый из которых сопоставляется объекту системы и задаётся сменой своих состояний. Состояния объекта-процесса определяются набором значений его переменных. Взаимодействие процессов происходит путём обмена сообщениями и через изменение разделяемых переменных. Таким образом, процессы в нашей онтологии характеризуются:

1. наборами разделяемых и внутренних переменных,
2. списком действий над этими переменными, изменяющими их значение,
3. списком каналов, по которым процесс может посылать или принимать сообщения,
4. списком действий отправки сообщений.

В соответствии с данным представлением онтология процессов (Рис.1) содержит пять основных классов понятий: *Процесс*, *Переменная*, *Действие*, *Коммуникация* и *Канал*. Исполнение какого-либо действия зависит от условий на значения переменных определенного процесса и содержания присланных ему сообщений.

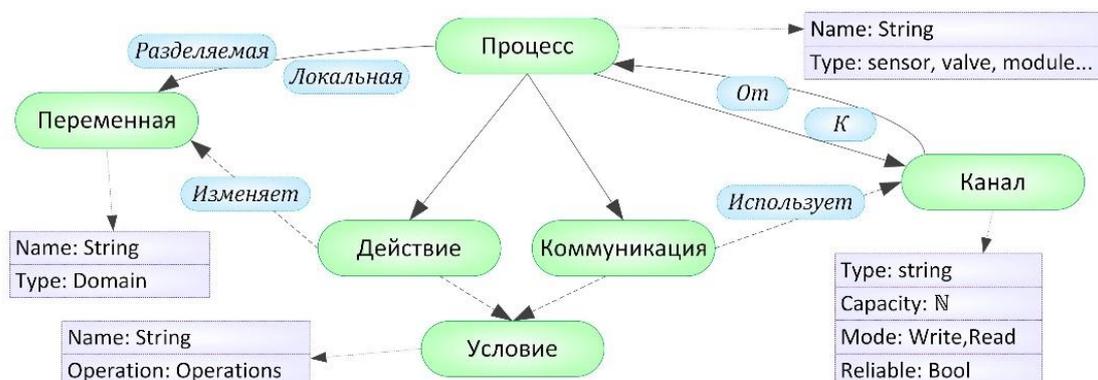


Рис. 1. Базовая онтология процессов.

Знаний, содержащихся в онтологии процессов, оказывается недостаточно для извлечения информации, поскольку такая онтология должна содержать типизацию объектов и процессов и правила их согласования. Такие правила кладутся в основу алгоритма извлечения информации и позволяют формировать экземпляры онтологических классов и выявлять связи на основе лексико-семантической сочетаемости найденных в тексте терминов.

2.2. Онтология АСУ. Онтология АСУ специализирует основные понятия онтологии процессов путем описания атрибутивной структуры основных классов и создание их классификаций.

Понятию *Процесс* базовой онтологии сопоставляются классы *Элемент АСУ* и *Система АСУ*, для которых определена их атрибутивная структура и иерархия типов. Можно выделить четыре основных типа элементов АСУ: управляющее, исполнительное, контрольно-измерительное устройство и объект управления. Семантику данных категорий можно выразить следующей формулой на естественном языке

[Управляющее устройство посылает команду исполнительному устройству на основе данных, полученных от контрольно-измерительного прибора, и исполнительное устройство выполняет определенный набор действий над объектом управления.]

Понятию *Переменной* сопоставляются классы *Контролируемая величина* и *Параметр*. Данные классы описывают изменяемые и неизменяемые характеристики объектов, которые могут быть достаточно разнообразными. Типизация характеристик и структура их описания задана иерархией *Величин*.

Базовые понятия *Канала* и *Коммуникации* выражаются с помощью класса *Сигнал* и парой отношений *посылает* (*Элемент АСУ*, *Сигнал*) и *принимает* (*Сигнал*, *Элемент АСУ*).

Базовое понятие *Действие* выражается с помощью класса *Воздействие* и отношениями *производит* (*Элемент АСУ*, *Воздействие*), *изменяет* (*Воздействие*, *Контролируемая величина*) и *влияет* (*Контролируемая величина*, *Воздействие*). Классификация воздействий опирается на иерархию *Принципов действия*, сопоставляемых каждому элементу АСУ.

Базовое понятие *Условие* является универсальным и не имеет специализации в данной онтологии.

Множество классов и отношений онтологии АСУ согласовано с базовым семантическим словарём для извлечения терминов из технической документации. Отметим, что обычно при извлечении информации опираются на семантический словарь, на основе которого с той или иной степенью общности может быть восстановлен онтологический класс объекта или отношения, описываемого термином словаря.

Для формирования каких-либо утверждений или свойств, которые в дальнейшем должны верифицироваться, в модель знаний необходимо явно внести понятие времени и логических отношений над элементами системы. Эти категории лежат в основе онтологии шаблонов требований.

2.3. Онтология требований. При формулировании требований к системе и сценарию ее работы важным является понятие состояния системы, которое определяется набором состояний объектов-процессов. Состояние объекта – это набор значений его переменных в определенный момент времени. В общем виде требования задаются спецификациями, которые, в свою очередь, строятся из утверждений и шаблонов, описывающих типичные

требования к поведению систем, такие как достижимость критического состояния, стабильное получение ответа на запросы и т.п. Верификация требований с помощью шаблонов является актуальной областью практических исследований [3, 12].

Класс *Утверждение* описывает различные состояния системы. Выделяются утверждения о данных (класс *Данные*) и о событии (класс *Ситуация*), а также их комбинации, описываемой составным событием. Спецификации требований, задаваемые классом *Спецификация*, являются булевыми комбинациями событий-утверждений и строятся с помощью пары отношений, используемых для задания двух операндов, и атрибута, используемого для указания булевого оператора. Более сложные спецификации задаются булевой комбинацией простых спецификаций и шаблонов, представленных классом *Шаблон*. Шаблоны снабжены рядом отношений-спецификаторов, позволяющих задавать различные ограничения, такие как время появления относительно определённых событий-утверждений, время выполнения событий по логическим часам и количество повторов.

Онтология шаблонов спецификаций, задающих требования, подробно описана в [1]. Однако, в явном виде спецификации на естественном языке практически не встречаются. Поэтому и потребовалась разработка системы онтологий и специализированного семантического словаря для того, чтобы обеспечить интеллектуальную систему верификации всеми необходимыми знаниями для вывода точных спецификаций из "неявных" знаний, представленных в тексте.

2.4. Семантический словарь. Информация, хранящаяся в семантическом словаре, должна обеспечить исследование лексико-семантических, семантико-грамматических и логико-композиционных особенностей текстов технических заданий. Словарь содержит термины, с помощью которых в тексте описываются понятия онтологий. Для семантической разметки терминов используется система семантических признаков – атрибутов и классов, организованных в иерархию на базе родовидовых отношений. Семантическая информация о термине кодируется в словаре путем отнесения его к одному или нескольким семантическим классам из заданной в словаре иерархии классов, а лексическое значение представлено в зоне толкования с помощью системы семантических атрибутов.

2.4.1. Иерархия классов словаря. Семантическая информация в словаре разделена на универсальную и предметную части с их последующим объединением либо с помощью механизма множественного наследования на классах, либо с помощью механизма поддержки множественного значения в рамках отдельного термина. Такой подход обеспечивает адаптируемость создаваемого словарного ресурса к другим предметным областям и жанрам текстов. Адаптация осуществляется посредством модификации предметной части иерархии в соответствии с новой предметной областью и пополнения словаря специальной лексикой. Добавлять специальной лексику можно автоматизированно, посредством «прокачки» текстов технической документации с последующей ручной настройкой словаря или путем обработки специализированных справочников или словарей, в которых спискам терминов сопоставлены категории предметной области.

Универсальная часть иерархии словаря содержит таксономию универсальных (не зависящих от предметной области) сущностей и их свойств, ситуаций и отношений.

Иерархия сущностей представлена классами *Физическая сущность*, включая подклассы *Физический объект* (в том числе *Предмет*, *Вещество*, *Место в пространстве*) и *Абстрактная сущность*, в том числе *Свойство*, *Информация*, *Содержание*, *Ментальный*

объект и *Темпоральный объект*. В классе *Предмет* выделяется класс *Приспособления и оборудование* (*Техническое устройство, Технологическое оборудование, Прибор* и т.п.). Класс *Свойство* в иерархии сущностей соответствует признакам выделяемых сущностей и содержит класс *Параметр* с противопоставлением подклассов *Количественный параметр* (*вес, температура, количество* и др.) и *Качественный параметр* (*цвет, должность* и т.п.).

Универсальная иерархия *Ситуаций*, участниками которых являются выделенные сущности, соответствует классам *Состояние, Действие* и *Процесс*. Процессы, как и действия, приводят к изменению состояния некоторых участников ситуации, но процессы дополнительно характеризуются протяженностью во времени и возможностью выделения фаз и соответствующих фазовых состояний.

В классе *Отношение* выделяются типичные отношения между объектами (например, *Использование, Темпоральность, Обусловленность, Сравнение*) и операторы, модифицирующие значение связанного термина (*Отрицание, Фаза, Модальность, Количество*).

Предметная иерархия содержит конкретную таксономию, которая определяется требованиями онтологии АСУ. Наименования сущностей распределяются по классам: *Компоненты АСУ*, включая классы *Аппаратный комплекс* с подклассами *Исполняющий компонент, Управляющий компонент, Контролирующий компонент* и т.д. Важными классами в иерархии сущностей являются *Контролируемый параметр* и *Состояние*.

Приведем пример наполнения иерархии для предметной области “Автоматическая линия розлива бутылок” [11]. К классу *Объект управления* должны быть отнесены релевантные для предметной области универсальные классы сущностей *Резервуар* (которым будут размечены слова *резервуар, бак*), *Сосуд* (*бутылка*) и *Жидкость* (*жидкость, вода*). Контролируемый параметр в этом случае – универсальный класс *Количественный параметр* (*уровень жидкости и температура*) и состояния на базе отношений универсальных классов *Наличие внутри* (*пустой, отсутствие*) и *Положение* (*находиться, положение, заданное положение*). *Компонент АСУ* будет включать в классе *Исполняющие компоненты* универсальные подклассы *Конвейерное оборудование* (*двухсегментный конвейер*), *Запорное устройство* (*клапан*) и *Регулятор температуры* (*пид-регулятор*).

2.4.2. *Лексическое значение слова (толкование)*. Для представления лексического значения слова используются семантические атрибуты. Совокупность значений атрибутов, приписанных слову, в определенной мере моделирует компонентную семантическую структуру слова. Компоненты семантической структуры слова могут рассматриваться как тезаурусные дескрипторы. Например, в структуре всех предикатов перемещения (*перемещаться, ехать, подаваться*) имеется компонент *sem.процесс=перемещение*. Объединение этого компонента со значением свойства *sem.параметр-кач=положение*, описывающего результат этого процесса (изменение положения), соответствует группе результативных предикатов перемещения в конечную точку (*подъезжать, доезжать, подаваться*). Добавление компонента отрицания состояния позволяет выделить группу предикатов перемещения из исходной точки (*удаляться, уезжать*).

3. Архитектура системы поддержки верификации технической документации.

Идея предлагаемого подхода заключается в построении такой технологии, которая бы в процессе семантической обработки текста и последующего анализа результатов выявляла случаи некорректного описания требований и несоответствия описания разрабатываемой

программной системы этим требованиям, в первую очередь для выявления неявных или скрытых ошибок в описании сценария работы системы.

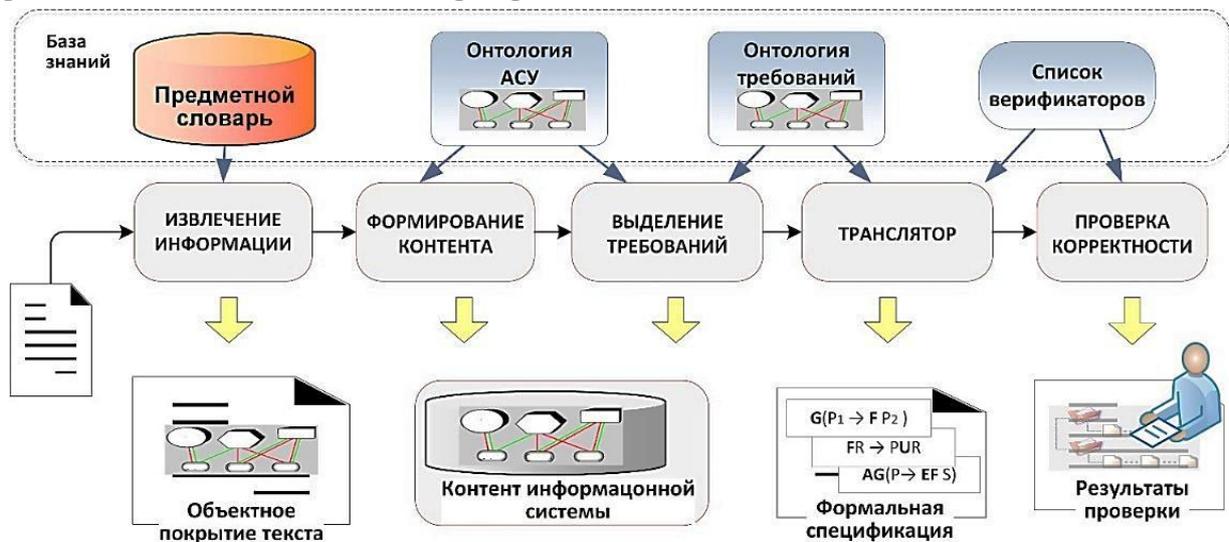


Рис. 2. Схема верификации технической документации.

Общая схема процесса поддержки верификации представлена на Рис. 2.

Интеллектуальная система поддержки верификации технической документации на разработку распределенных систем включает базу знаний и набор исполняемых компонент. На вход системе подается пакет технической документации, результаты работы (в том числе и промежуточные) выдаются пользователю в структурированном виде и предоставляется возможность внести необходимые изменения.

Система знаний, как уже было сказано ранее, представлена набором онтологий. Онтология АСУ является специализацией онтологии объекта управления и онтологии процессов верхнего уровня, позволяющая рассматривать технологические процессы и их взаимодействие. Онтология требований позволяет описать поведенческие свойства системы. Предметный словарь описывает лексическую онтологию для всех остальных онтологий. Отдельным компонентом базы знаний является описание «доступных» верификаторов, которое должно позволять выбрать определенный верификатор для проверки сформулированного требования к системе наилучшим образом.

Система включает следующие программные компоненты.

- 1) Подсистема извлечения информации обеспечивает семантический анализ текста и извлечение информации в виде экземпляров классов онтологии процессов. На данном этапе решаются следующие задачи:
 - а) построение терминологического покрытия текстов на основе предметного словаря;
 - б) построение множества информационных объектов на основе терминологического покрытия текста;
 - в) разрешение кореференции объектов на основе их идентификаторов.
- 2) Модуль формирования контента обеспечивает анализ найденной в тексте информации и наполнение контента информационной системы. Для этого необходимо обеспечить:
 - а) уточнение объектов, параметры которых необходимо заполнить для обеспечения корректного пополнения контента системы верификации;
 - б) снятие неоднозначности (семантических конфликтов);
 - в) проверку совместности найденной информации на основе аксиом онтологии.

- 3) Модуль выделения требований на основе причинно-следственных и временных отношениях процессов, найденных в документе, формирует типичные требования корректности свойств рассматриваемой технологической системы в виде экземпляров класса спецификации онтологии требований.
- 4) Транслятор обеспечивает выбор верификатора и преобразование контента системы, представленного в виде экземпляров классов онтологий, в описание системы на входном языке верификатора и формулы логики спецификаций, поддерживаемой данным верификатором.
- 5) Модуль проверки корректности обеспечивает вызов инструмента верификации и формирует результат в виде списка спецификаций, не прошедших проверку на корректность.

В силу неоднозначности естественного языка и трудоёмкости точного задания требований корректности программных систем все инструменты системы поддержки формальной верификации не являются полностью автоматическими, т.е. результат их работы может требовать дополнительного ручного анализа.

Заключение. Таким образом, рассмотренные в работе инструменты – словарь и набор онтологий – являются семантической базой для исследования проблемы корректности при разработке программных систем. Отличительной особенностью предлагаемого нами подхода является использование онтологии как связующего звена между текстом и формальной верификацией. Такой подход обладает следующими преимуществами:

- семантичность, подход опирается на семантические методы и модели, в первую очередь, на онтологию предметной области объекта управления, относительно которой решается задача извлечения информации;
- масштабируемость, подход к верификации расширяем относительно типов требований и классов верифицируемых ошибок.

За рамками рассмотрения остались вопросы алгоритмической реализации и апробации предложенных моделей, над которыми в данный момент ведется активная работа.

Авторы выражают благодарность фонду РФФИ (проект №17-07-01600) и Президиуму СО РАН (Блок 36.1. Комплексной программы ФНИ СО РАН П.1) за поддержку выполненных исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Гаранина Н.О., Зюбин В. Е., Лях Т. В. Онтологический подход к организации шаблонов требований в рамках системы поддержки формальной верификации программных систем // Системная информатика. 2017. № 9. С. 111–132.
2. Заболеева-Зотова А.В., Орлова Ю.А. Автоматизация семантического анализа текста технического задания. Волгоград. ВолгГТУ. 2010. 155 с.
3. Autili M., Grunske L., Lumpe M., Pelliccione P., Tang A. Aligning Qualitative, Real-Time, and Probabilistic Property Specification Patterns Using a Structured English Grammar // IEEE Transactions on Software Engineering. Volume: 41. Issue: 7. 2015. Pp. 620–638.
4. Deeptimahanti D., Sanyal R. Semi-automatic generation of UML models from natural language requirements // Proceedings of the 4th India Software Engineering Conference. ACM. 2011. Pp. 165–174.

5. Garanina N., Anureev I. Verification Oriented Process Ontology // Proc. of Nine Workshop on Program Semantics, Specification and Verification: Theory and Applications (PSSV 2018). Russia. ЯргУ им. П.Г. Демидова. 2018. С. 58–67.
6. Hepp M., Leymann F., Domingue J., Wahler A., Fensel D. Semantic Business Process Management: A Vision Towards Using Semantic Web Services for Business Process Management // IEEE International Conference on e-Business Engineering (ICEBE 2005). Beijing, China. 2005. Pp. 535–540.
7. Hoare C. A. R. Communicating sequential processes // Communications of the ACM. 21 (8): Pp. 666–677.
8. Kof, L.: Requirements analysis: concept extraction and translation of textual specifications to executable models // Horacek H., Métais E., Muñoz R., Wolska M. (eds) Natural Language Processing and Information Systems. 2009. Pp 79–90. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg. 2010. vol. 5723.
9. Popescu D., Rugaber S., Medvidovic N., Berry D. Reducing ambiguities in requirements specifications via automatically created object-oriented models // Lecture Notes in Computer Science, Springer, Heidelberg. 2008. vol. 1. Pp. 103–124.
10. Schlenoff C., Gruninger M., Tissot F., Valois J., Lubell J., and Lee J. The Process Specification Language (PSL) Overview and Version 1.0 Specification // NIST Report 6459. Jan. 1999.
11. Shanmugham S.G., Roberts C.A. Application of graphical specification methodologies to manufacturing control logic development: a classification and comparison // Int. J. Computer integrated manufacturing. 1998. Vol. 11. Issue 2. Pp. 142–152.
12. Yu J., Manh T. P., Han J. et al. Pattern based property specification and verification for service composition // Proc. of 7th International Conference on Web Information Systems Engineering (WISE). Vol. 4255 of LNCS. Springer-Verlag. 2006. Pp. 156–168.

UDK 004.912, 004.82

**SEMANTIC FEATURES OF TECHNICAL DOCUMENTATION:
ONTOLOGICAL VIEW OF THE PROBLEMS OF CORRECTNESS**

Elena A. Sidorova

PhD, Senior researcher, e-mail: lsidorova@iis.nsk.su

Natalia O. Garanina

PhD, Senior researcher, e-mail: garanina@iis.nsk.su

Irina S. Kononenko

Researcher, e-mail: irina_k@cn.ru

Olesya I. Borovikova

Junior researcher, e-mail: olesya@iis.nsk.su

A.P. Ershov Institute of Informatics Systems SB RAS,
6, Acad. Lavrentjev pr., Novosibirsk, Russia 630090

Abstract. In the paper, the specifics of the genre of technical documentation are examined with regard to extracting the content information necessary to analyze the correctness of the specifications set before the task developers. A distinctive feature of the proposed approach is the use of ontology as a link between text and formal verification. The objects and processes described in the text are determined by a set of ontologies: the ontology of requirements, the ontology of processes and distributed systems, and the ontology of the control object. The ontology of the control object is defined within the semantic dictionary and includes concepts describing special objects and characteristics used in a particular control system.

Keywords: semantic text analysis, information extraction, requirements specification, ontology of requirements, ontology of processes, automated control systems.

References

1. Garanina N., Zyubin V., Liakh T. Ontologicheskii` podhod k organizatscii shablonov trebovaniï` v ramkakh sistemy` pod-derzhki formal`noi` verifikatsii programmny`kh sistem [Ontological Approach to Organizing Specification Patterns in the Framework of Support System for Formal Verification of Distributed Program Systems] // Sistemnaia informatika = System Informatics. 2017. No 9. Pp. 111–132. (in Russian)
2. Zaboleeva-Zotova A.V., Orlova Iu.A. Avtomatizatsiia semanticheskogo analiza teksta tekhnicheskogo zadaniia [Automation of the semantic analysis of the text of technical task] Volgograd. VSTU. 2010. 155 p. (in Russian)
3. Autili M., Grunske L., Lumpe M., Pelliccione P., Tang A. Aligning Qualitative, Real-Time, and Probabilistic Property Specification Patterns Using a Structured English Grammar // IEEE Transactions on Software Engineering. Volume: 41. Issue: 7. 2015. Pp. 620–638.
4. Deeptimahanti D., Sanyal R. Semi-automatic generation of UML models from natural language requirements // Proceedings of the 4th India Software Engineering Conference. ACM. 2011. Pp. 165–174.

5. Garanina N., Anureev I. Verification Oriented Process Ontology // Proc. of Nine Workshop on Program Semantics, Specification and Verification: Theory and Applications (PSSV 2018). Russia. ЯрГУ им. П.Г. Демидова. 2018. С. 58–67.
6. Hepp M., Leymann F., Domingue J., Wahler A., Fensel D. Semantic Business Process Management: A Vision Towards Using Semantic Web Services for Business Process Management // IEEE International Conference on e-Business Engineering (ICEBE 2005). Beijing, China. 2005. Pp. 535–540.
7. Hoare C. A. R. Communicating sequential processes // Communications of the ACM. 21 (8): Pp. 666–677.
8. Kof, L.: Requirements analysis: concept extraction and translation of textual specifications to executable models // Horacek H., Métais E., Muñoz R., Wolska M. (eds) Natural Language Processing and Information Systems. 2009. Pp 79–90. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg. 2010. vol. 5723.
9. Popescu D., Rugaber S., Medvidovic N., Berry D. Reducing ambiguities in requirements specifications via automatically created object-oriented models // Lecture Notes in Computer Science, Springer, Heidelberg. 2008. vol. 1. Pp. 103–124.
10. Schlenoff C., Gruninger M., Tissot F., Valois J., Lubell J., and Lee J. The Process Specification Language (PSL) Overview and Version 1.0 Specification // NIST Report 6459. Jan. 1999.
11. Shanmugham S.G., Roberts C.A. Application of graphical specification methodologies to manufacturing control logic development: a classification and comparison // Int. J. Computer integrated manufacturing. 1998. Vol. 11. Issue 2. Pp. 142–152.
12. Yu J., Manh T. P., Han J. et al. Pattern based property specification and verification for service composition // Proc. of 7th International Conference on Web Information Systems Engineering (WISE). Vol. 4255 of LNCS. Springer-Verlag. 2006. Pp. 156–168.