

05.13.11

В.В. Ланин

Пермский филиал национального исследовательского университета «Высшая школа экономики», факультет бизнес-информатики, кафедра информационных технологий в бизнесе, Пермь, vlanin@live.com

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ПОДДЕРЖКИ РАБОТЫ АНАЛИТИКА ПРИ ПРОЕКТИРОВАНИИ ИНФОРМАЦИОННОЙ СИСТЕМЫ

В статье описан подход к реализации системы, автоматизирующей обработку проектной документации аналитиком при анализе предметной области и проектировании информационной системы. Анализ документации осуществляется на основе обработки естественного языка, специально разработанного предметно-ориентированного языка и онтологических ресурсов. В результате работы системы выделяются связи между документами и производится семантическая индексация документов, на основе которой аналитик сможет легко ориентироваться в документах, быстро найти необходимые разделы документов, проверить непротиворечивость и полноту требований.

Ключевые слова: *проектная документация, разработка информационных систем, обработка естественного языка.*

Введение

Техническая документация играет важную роль на протяжении всего жизненного цикла информационной системы (ИС). Особое значение документация имеет на стадии проектирования, т.к. в ней описываются требуемые свойства будущей системы, и чем эффективнее будет управление требованиями, тем выше будет вероятность успешного завершения проекта. В документации требования чаще всего представлены на естественном языке, и в задачу аналитика входит их формализация. В рамках проекта предлагается разработка инструментального средства, автоматизирующего работу аналитика по анализу проектной документации.

Системы анализа документации при проектировании ИС

Поддержка этапа проектирования ИС включает целый ряд задач, традиционно наибольшее внимание уделяется задаче управления требованиями. Проблематику управления требованиями можно рассматривать с разных ракурсов, с точки зрения менеджмента и подходов, а также с точки зрения применения инструментов и средств для упрощения и совершенствования работы с требованиями. Значительное количество исследований, ведущихся в данной области, направлены на анализ проектной и технической документации. Из российских исследований можно отметить А.В. Заболееву-Зотову [1] и Ю.А. Орлову [2], в работах которых предлагается подход к анализу технического задания на основе специально разработанной нечеткой атрибутивной грамматики, позволяющей отобразить содержимое технического задания. Среди зарубежных исследователей можно выделить исследовательскую группу университета Concordia [3]. В соответствии с их подходом для трансформации требований, выраженных на естественном языке, строятся концептуальные модели на специально разработанном языке. Исследованием и разработкой подходов и инструментов к решению вышеуказанных проблем занимаются также крупные компании, на рынке присутствуют продукты RequisitePro, DOORS, CaliberRM, служащие для поддержки процесса документирования требований в крупных проектах. Однако ни одним из этих решений не поддерживается автоматический анализ проектной документации.

Модель документа и многоаспектная онтология электронных документов

Прежде всего для анализа документации необходимо определить модель документа. Согласно предлагаемой модели электронный документ может быть представлен четверкой вида:

$$d = (S(F, R), C, o, M).$$

Здесь $S(F, R)$ – ориентированный гиперграф, вершинам которого сопоставлены элементы множества F (множество F – это множество фрагментов документа, а R – это множество ребер графа, соответствующее связям между фрагментами); элементы множества C представляют информационное содержание документа (его контент); o – онтология документа, M – отображение множества F на концепты онтологии o . Учитывая специфику решаемых в данной работе задач, конкретизируем понятие *онтологии*:

$$o = (C, R, A),$$

где C – множество *понятий (концептов)* онтологии, R – множество *отношений* между концептами, A – множество *аксиом*, заданных на онтологии. Концептами могут быть как классы, так и экземпляры этих классов, а аксиомы используются для задания ограничений и правил, которые не могут быть выражены через отношения.

Для описания метаданных документы предлагается использовать многоаспектную онтологию электронных документов [4], которая представляет консолидированные знания об их структуре и содержании (формат, тип и структура). В нее включены понятия, относящиеся ко всем трём выделенным аспектам представления информации о документах, каждый из которых описывается онтологией, однако понятия, относящиеся к различным аспектам, связаны между собой. Таким образом, создаётся единая онтология электронных документов. Ресурс поддерживает возможность расширения и уточнения для настройки на решение задач, возникающих при обработке документов в различных ИС в течение всего их жизненного цикла. К настоящему моменту в онтологии описываются стандарты, законодательные акты, связанные с разработкой и функционированием ИС и техническое задание.

Предметно ориентированный язык описания проектной документации

Для описания документации, применяемой в процессе проектирования ИС, был разработан визуальный предметно-ориентированный язык. Предметно-ориентированный язык (Domain-Specific Language, DSL) – это язык моделирования, предназначенный для решения определенного класса задач в конкретной предметной области. В отличие от языков моделирования общего назначения, DSL более выразителен, прост в применении и понятен различным категориям специалистов, поскольку язык оперирует привычными для них терминами предметной области. Для поддержки процесса разработки и сопровождения DSL используется специальный вид программного обеспечения, получивший название DSM-платформа. Система MetaLanguage – DSM-платформа, предназначенная для создания визуальных предметно-ориентированных языков, построения моделей с использованием этих языков и преобразования созданных моделей в иные текстовые и графические нотации [4].

Поскольку при описании проектных документов важно не только определить их структуру, но и задать связи между ними, то разработаем двухуровневый предметно-ориентированный язык описания документов, используемых при проектировании ИС. Первый уровень языка предназначен для описания всего множества документов и взаимосвязи между отдельными документами (рис. 1), а второй – для описания структуры конкретного документа.

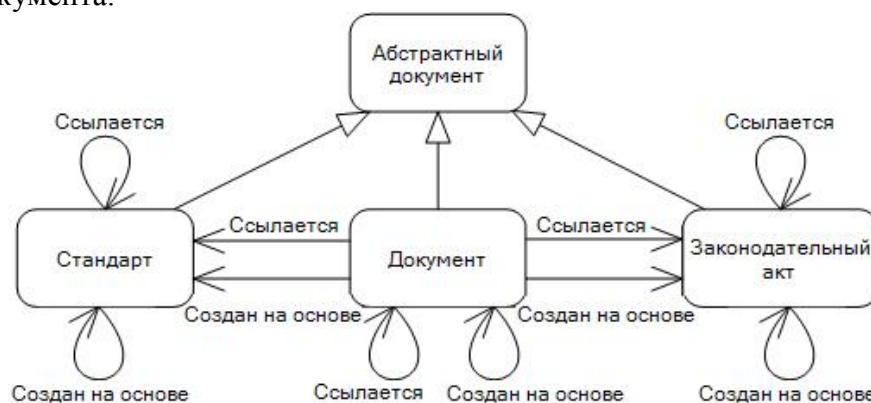


Рис. 1. Метамоделл языка описания связей между документами Архитектура системы

Основные компоненты системы представлены на рис. 2. Исходные документы хранятся в *репозитории документов*, поддерживается работа с форматами OpenXML, ODF, HTML и PDF. В *репозитории моделей* хранится описание типов документов на разработанном предметно-ориентированном языке. Документы, проиндексированные понятиями предметной и многоаспектной онтологиями помещаются *хранилище семантических индексов*.



Рис. 2. Репозитории и программные компоненты системы

Заключение

В настоящий момент реализован прототип системы на платформе Java с использованием библиотек платформы GATE [7], онтологические ресурсы разрабатываются в системе Protege 4.2. Проводятся работы по наполнению базы лексико-семантических шаблонов на языке JARE и уточнение моделей описания документов.

Исследование выполнено при финансовой поддержке РФФИ в рамках проекта № 14-07-31273-мол_a и Научного фонда НИУ ВШЭ (проект № 14-09-0228).

Список литературы

1. *Заболеева-Зотова А.В.* Автоматизация процедур семантического анализа текста технического задания // Известия ВолгГТУ. Сер. «Актуальные проблемы управления, вычислительной техники и информатики в технических системах»: межвуз. сб. науч. ст. / 2007. Вып. 2 №2. С.39-42.
2. *Орлова Ю.А.* Методика анализа текста технического задания // Информатика, вычислительная техника и управление // Известия ТулГУ. Сер. «Техническиенауки»: межвуз. сб. науч. ст. / 2011. Вып. 3. С.213-220.
3. *Zeng Y.* Recursive Object Model, Modelling of linguistic information in engineering design. Computers in Industry 59. 2008. pp.612-625.
4. *Ланин В.В., Лядова Л.Н.* Онтология электронных документов как основа для извлечения знаний о предметной области при ее моделировании // Материалы Всероссийской конференции с международным участием «Знания-Онтологии-Теории». Новосибирск, 2013, Т.1, с. 245-251.
5. *Замятина Е.Б., Лядова Л.Н., Сухов А.О.* Мультиязыковое моделирование с использованием DSM-платформы MetaLanguage // Информатизация и связь. – 2013. – № 5. – С. 11-14.
6. *Агалакова Е.З., Ланин В.В.* Предметно-ориентированный язык описания структуры и содержания электронных документов // Технологии разработки информационных систем ТРИС-2014: сборник статей международной научно-практической конференции. – Таганрог: Издательство ЮФУ, 2014. – С. 70-75.
7. *Cunningham H., Maynard D., Bontcheva K.* Text Processing with GATE. – Gateway Press CA, 2011.