

ТЕХНОЛОГИЯ СОЗДАНИЯ ДОКУМЕНТ-ОРИЕНТИРОВАННЫХ СИСТЕМ, ОСНОВАННЫХ НА МЕТАМОДЕЛИРОВАНИИ

Вячеслав Ланин, Людмила Лядова, Антон Цыбин

Abstract: Описывается подход к созданию динамически адаптируемых информационных систем (ИС), основанный на использовании моделей на всех этапах жизненного цикла. В основе подхода – средства управления документами, позволяющие решать различные задачи, связанные поиском документов, относящихся к предметной области системы, их анализом, классификацией и каталогизацией, созданием на их основе моделей, управляющих функционированием системы, а также подготовкой документации на всех этапах жизненного цикла ИС. Особенность подхода – его ориентация на разработку систем, допускающих динамическую адаптацию к потребностям бизнес-процессов и отдельных пользователей, к меняющимся условиям эксплуатации. Представленные средства основаны на использовании многоуровневых моделей, управляющих созданием и функционированием ИС. Модели разбиваются на несколько уровней, представляющих систему с различных точек зрения и с разной степенью детализации. Базовые модели являются основой для реализации моделей более высоких уровней. Пользователи системы – эксперты в соответствующих предметных областях – участвуют как в процессе создания системы, так и в её настройке в ходе эксплуатации с использованием специальных инструментальных предметно-ориентированных средств. Основа реализации предметно-ориентированных средств разработки адаптируемых ИС – метамоделирование. Мета модель при реализации данного подхода – это язык, с помощью которого описываются модели, лежащие в основе информационной системы. Язык должен отражать специфику предметной области, быть доступным для работы экспертов. Создание системы начинается именно с разработки предметно-ориентированного языка (DSL) с помощью языкового инструментария на основе выполненного анализа документов, описывающих предметную область, регламентирующих работу системы. Предлагается подход к автоматизации анализа документов на основе онтологий, описывающих как сами документы, так и предметную область. При этом извлечённая из документов информация является основой для уточнения построенных моделей и разработки предметно-ориентированных языков. Процесс создания DSL является итерационным. Построенные модели управляют функционированием системы, генерацией отчётов и пр. Кроме того, созданные модели – основа средств автоматизации документирования системы.

Keywords: анализ документов, семантическая разметка, генерация документов, CASE-средства, онтологии, метамоделирование.

ACM Classification Keywords: I.7 DOCUMENT AND TEXT PROCESSING: I.7.2 Document Preparation – Index generation; I.7.1 Document and Text Editing – Document management. D.2 SOFTWARE ENGINEERING: D.2.2 Design Tools and Techniques – Computer-aided software engineering (CASE).

Введение

Задача создания адаптируемых информационных систем (ИС), позволяющих осуществлять гибкую настройку на потребности бизнес-процессов и отдельных пользователей, меняющиеся условия эксплуатации, – одна из критичных задач программной инженерии, так как возможности адаптации определяют «живучесть» системы, эффективность вложений в её создание и стоимость её эксплуатации, сопровождения. Свойство адаптируемости гарантирует возможность развития ИС, её эволюции [Лядова, 2007]. Оно объединяет многие нефункциональные характеристики, обеспечить которые можно только при использовании для создания ИС инструментария, обладающего соответствующими возможностями.

Можно отметить следующие требования к информационным системам, обладающим высокой степенью *адаптируемости*: наличие средств динамической (в ходе эксплуатации системы) настройки, максимально снижающих трудоёмкость и позволяющих выполнять эту работу пользователям системы, для которых должна быть обеспечена возможность работы в привычных терминах предметной области ИС, в которой они работают и являются экспертами.

Все современные технологии и инструментальные средства создания ИС основаны на использовании моделей. Однако максимальная степень адаптируемости ИС достигается, если она основана на *метамоделировании* [Лядова, 2008] и функционирует в *режиме интерпретации моделей*, которые могут изменяться в ходе эксплуатации системы, описывающих предметную область ИС и условия её работы.

На сегодняшний день одним из базовых понятий методологии проектирования ИС является понятие жизненного цикла. *Жизненный цикл* – это непрерывный процесс, который начинается с момента принятия решения о необходимости создания информационной системы и заканчивается завершением её эксплуатации. Структура жизненного цикла включает *три группы процессов*: основные, вспомогательные и организационные. Основные процессы включают этапы создания, внедрения и эксплуатации, сопровождения. Этап создания также является сложным и включает оценку жизненного цикла системы, анализ требований, проектирование структуры компонентов, реализацию проекта. Все этапы создания ИС сопровождаются вспомогательным процессом – *документированием*. В рамках данного процесса производятся работы по созданию различных инструкций и руководств, документации пользователя и программиста.

Из всех описанных этапов жизненного цикла наиболее формализованными и обеспеченными средствами автоматизации являются этапы анализа требований и моделирования предметной области. Меньше внимания в жизненном цикле уделяют этапам тестирования и создания документации. Такая закономерность наблюдается, прежде всего, в небольших компаниях по разработке программного обеспечения (ПО). Причина состоит в сложности автоматизации этих процессов. На выполнение данных этапов вручную требуется большой объём времени и средств, зачастую эти затраты не окупаются для небольших компаний. При этом тестирование – важный этап жизненного цикла, так как недостаточное внимание к нему не позволяет получить качественный, надёжно функционирующий программный продукт. Документирование – также очень важный процесс, т.к. отсутствие документации или её плохое качество затрудняют эксплуатацию и сопровождение системы, возможность её переноса на другие платформы.

Как при создании системы, так и в ходе её эксплуатации решаются задачи работы с документами. Анализ требований к системе связан с необходимостью анализа большого объёма документов, в частности, нормативных, распорядительных и пр., регламентирующих бизнес-процессы и деятельность различных

категорий пользователей создаваемой системы. Анализ документов – основа моделирования предметной области ИС, проектирования её компонентов. Автоматизация этой трудоёмкой задачи – тема исследований многих авторов [Заболеева-Зотова, 2008; Заболеева-Зотова, 2009; Fulford, 2001; Kuo, 2004; Sierra, 2004; Song, 2007 и др.]. В ходе эксплуатации системы требования к ней могут меняться, что отражается либо в изменениях, внесённых в соответствующие нормативные документы, использованные при анализе требований при создании ИС, либо в новых документах, регламентирующих её работу. Таким образом, необходимо автоматизировать отслеживание этих изменений, появление новых документов, которые могут повлиять на разработанную при создании системы модель. Решение этой задачи может быть основано на использовании онтологий, описывающих предметную область и анализируемые документы [Ланин, 2009; Ланин, Лядова, 2010; Лядова, 2010].

При создании системы строится множество моделей различных уровней, которые управляют процессами разработки и функционирования адаптируемой системы, описывая её с различных точек зрения и на разных уровнях детализации [Лядова, 2007; Лядова, 2008]. Причём специфика проектов создания адаптируемых систем, разработки моделей обуславливает необходимость привлечения экспертов – специалистов в соответствующих предметных областях – на всех этапах жизненного цикла ИС. Это требует использования специальных инструментальных средств, языков предметных областей (DSL – Domain Specific Languages), которые обеспечивают экспертам возможность работы в привычных для них терминах при создании или модификации моделей [Лядова, 2010; Лядова, Сухов, 2010]. Основа разработки DSL – анализ предметной области, документов описывающих её. Таким образом, модели, применяемые при анализе документов, и создаваемые модели, управляющие функционированием ИС, объединяются в единую систему. Переходы от одной модели к другой основаны на трансформации моделей, которые описываются также с помощью специальных языков трансформаций предметно-ориентированных языков, представляющих собой метамоделли.

Важность задачи документирования и управления документооборотом при создании программных систем показывает и количество работ, направленных на её решение в различных областях, появившихся в последнее время [Арушанян, 2010; Климов, 2010; Мамаев, 2008; Смирнов, 2012; Сонис, 2006; Суясов, 2007; Тарасенко, 2009 и др.]. Особое значение задача автоматизации документирования ИС, управления документооборотом при реализации программных проектов приобретает при создании систем, допускающих динамическую (в ходе эксплуатации) настройку, т.к. любые изменения системы должны соответствующим образом документироваться [Цыбин, 2007; Tsybin, 2008]. Процесс документирования также базируется на использовании различных моделей.

При документировании системы, разработке руководств и инструкций необходимо сохранить связь с исходными документами, регламентирующими функционирование системы. Эту задачу можно решить, сохранив при создании модели предметной области не только исходные документы, но и внесённую в них специальную разметку, построенную в результате семантического анализа текстов.

Структура системы, ориентированной на документы

Как при создании системы, так и в ходе её эксплуатации решаются задачи управления документами, их поиска, анализа и классификации, каталогизации и эффективного хранения, генерации отчётов и документации, поддержания всего жизненного цикла документов в системе.

Практически все задачи, решаемые при создании и эксплуатации системы, связаны с необходимостью работы с документами в различных форматах:

- На этапе анализа предметной области происходит поиск и изучение документов, описывающих состояние предметной области, свойства различных её объектов, регламентирующих бизнес-процессы, которые должны быть автоматизированы, задающих условия функционирования создаваемой системы, предъявляемые к ней требования. Результаты анализа формализуются в построенных моделях.
- При создании системы каждый процесс должен быть документирован, должна быть подготовлена программная документация, инструкции и руководства по эксплуатации ИС. При этом необходимо сохранить связь с исходными документами, описывающими бизнес-систему, регламентирующую все процессы в ней.
- В ходе эксплуатации системы пользователи получают отчёты, генерируют документы на основе информации, хранящейся в ИС и получаемой извне.

Необходимо максимально автоматизировать решение этих задач, снизить трудоёмкость работы с документами на всех этапах жизненного цикла ИС. Структура динамически адаптируемой документ-ориентированной системы, управляемой моделями, показана на рис. 1.

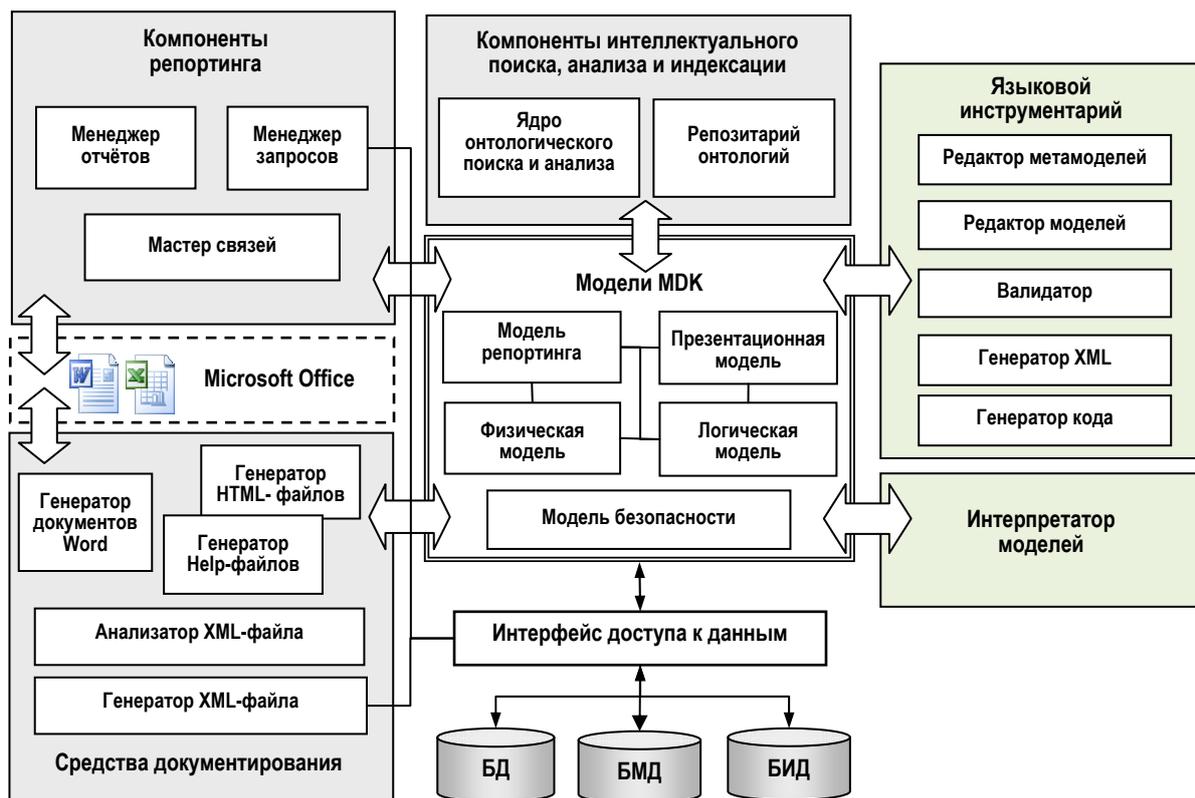


Рис. 1. Упрощённая структура динамически адаптируемой системы, ориентированной на документы, управляемой моделями

Включение средств BI (Business Intelligence), в частности, средств подготовки отчётов (подсистем репортинга) в состав ИС – требование к любой современной системе. Именно эти средства позволяют представить информацию, получаемую пользователями системы, в удобном для них виде,

визуализировать данные, результаты их обработки. Однако адаптируемая система должна обеспечить возможность настройки средств репортинга на любые изменения, которые могут произойти в ходе эксплуатации системы (дать возможность пользователям получать новые отчёты, менять формы документов и пр.). Желательно при этом снизить трудоёмкость этой работы, минимизировать необходимость вмешательства разработчиков – позволить пользователям самим адаптировать ИС к своим потребностям. Современные предметно-ориентированные средства создания ИС позволяют решить эти задачи, дать возможность принимать участие в создании системы и её настройке в процессе эксплуатации пользователям-непрограммистам, являющимся специалистами в тех предметных областях, для которых разрабатываются ИС. Для настройки средств репортинга используются высокоуровневые средства, предметно-ориентированные модели [Lanin, 2008].

В отличие от «традиционных систем», условия и правила функционирования которых не меняются в процессе эксплуатации, для систем, допускающих динамическую настройку, задачи анализа предметной области и разработки документации необходимо выполнять не только при создании системы, но и в ходе её функционирования – каждый раз, как только возникает необходимость внесения изменений в «поведение» системы, адаптации её к новым условиям и требованиям. Таким образом, анализ документов и подготовка документации – задачи, которые выполняются в течение всего жизненного цикла ИС. Решение этих задач требует автоматизации, использования средств, которые позволили бы снизить трудоёмкость настройки системы, анализа изменений условий функционирования и потребностей пользователей и внесения изменений в правила функционирования ИС и их документирования.

Все перечисленные средства управления документами используют для настройки и выполнения своих операций модели различных уровней. Модели создаются при разработке системы с помощью специального инструментария. Обычно строится несколько моделей, описывающих систему с различных точек зрения, на разных уровнях абстракции. Базовые уровни моделей, реализованные в CASE-системе METAS, – это логический (представляет объекты предметной области и связи между ними), физический (описывает представление данных об объектах в базе данных (БД) системы в терминах таблиц и связей) и презентационный (описание пользовательского интерфейса) [Лядова, 2007; Лядова, 2008]. Для их хранения используется база метаданных (БМД). Однако возможность использования предметно-ориентированных средств (языков) требует интеграции в систему языкового инструментария [Лядова, Сухов, 2010]. Работа по созданию метамodelей (предметно-ориентированных языков) основана на использовании средств анализа предметной области, описывающих её документов. Результаты анализа формализуются и сохраняются в системе. Основа средств автоматизации анализа документов – онтологии предметной области и онтологии документов.

Информационный поиск, анализ, индексация и каталогизация документов

Основная часть документов, с которыми работает современное предприятие, – это неструктурированные документы. Документами такого типа являются документы, представляющие нормативную и справочную информацию, распорядительные документы. Таким образом, основные знания о деятельности бизнес-системы содержатся именно в неструктурированных документах. Если принять во внимание экспоненциальный рост количества документов, проблема обработки неструктурированных документов, их семантического индексирования становится чрезвычайно актуальной.

Задачи обработки текстовой информации могут получить более эффективное решение за счёт применения подходов, основанных на явном представлении знаний.

Задача информационного поиска решается на основе онтологического подхода [Ланин, 2008; Ланин, 2009]. Онтологии являются не только средством интеллектуального поиска и анализа документов, но и основой организации их эффективного хранения, создания системы взаимосвязанных документов и средств навигации по ним. Предусматривается описание с помощью онтологий как структуры документа, так и его содержания, оформления. Вся эта информация используется для поиска и анализа документов, их аннотирования, классификации и каталогизации в системе в форме, удобной для решения различных задач, связанных с управлением документами. Извлекаемая из документов информация служит основой для анализа изменений предметной области ИС, потребностей её пользователей и условий функционирования, внесения соответствующих изменений в модели, снижая трудоёмкость работы системного аналитика при создании и адаптации ИС.

Предлагаемый подход предполагает расширение понятия традиционного документа: с *документом необходимо связать метаданные*, позволяющие интерпретировать и обрабатывать хранящуюся в этом документе информацию, т.е. включить в документ информацию, описывающую его структуру и семантику его содержания с помощью онтологии.

Согласно общепринятому определению под *онтологией* (в широком смысле) понимается база знаний специального типа, которая может «читаться» и пониматься, отчуждаться от разработчика и/или физически разделяться её пользователями. Учитывая специфику решаемых в данной работе задач и предлагаемый для их решения подход, можно конкретизировать понятие онтологии и рассмотреть различные типы используемых онтологий:

- онтология предметной области конкретной информационной системы;
- онтология как база знаний (БЗ) интеллектуального агента;
- онтология как описание документа.

Онтологии предметной области имеют наиболее типичное применение, они используются для описания понятий предметной области ИС (например, образование, государственные услуги или инновационное развитие регионов). В этой онтологии описывается связь понятий, языковые единицы для их выражения, аксиомы предметной области. Онтология предметной области используется для семантического индексирования и анализа всех документов системы.

Для анализа документов (реализации ядра онтологического поиска и анализа, рис. 1) предлагается использовать *мультиагентный подход*. Интеллектуальные агенты, руководствуясь *онтологией как базой знаний* (второй тип онтологий), производят поиск и анализ конкретных понятий документа. Каждая из вершин такой онтологии имеет определённый прототип, интерпретация которого известна агенту. Таким образом, агент использует онтологию как определённую программу своих действий. Вершинами онтологии данного типа могут являться понятия из онтологии предметной области.

Третий тип онтологий используется для *описания структуры и содержания документов*. Этот тип онтологий включает в себя два класса (две «плоскости») вершин. К первому классу относятся вершины, описывающие *структуру документа*, его реквизитный состав (например: таблица, дата, должность и т.д.); они представляют собой общие понятия, не зависящие от конкретной предметной области. К другому типу относятся вершины, содержащие *понятия документа*, эти вершины отражают специфику

предметной области, к которой относится документ, его семантику. Первый тип вершин в рамках предложенного подхода будем называть *структурными вершинами*, а второй тип – *семантическими вершинами*. Благодаря такому подходу из документа можно получить требуемую информацию, т.к. известно, где искать данные и как они могут быть интерпретированы.

Если представлять документ с использованием онтологий, то задача сопоставления онтологии и анализируемого документа сводится к задаче поиска понятий онтологии в документе. Как следствие, системе необходимо ответить на вопрос: описывает ли данная онтология документ или нет. На последний вопрос можно ответить утвердительно, если в процессе сопоставления в документе были найдены все понятия, включённые в онтологию. Таким образом, исходная задача сводится к задаче поиска в тексте документа общих понятий на основе формальных описаний. На основе онтологии может быть получен фрейм, слоты которого заполняются в процессе анализа документа. В качестве слотов фрейма выступают понятия онтологии, а значения этих фреймов заполняются данными анализируемого документа. Таким образом из неструктурированного документа может быть получен структурированный документ-фрейм.

Онтологии располагаются на трёх уровнях *репозитария (библиотеки) онтологий* (рис. 1). На первом уровне расположены онтологии, описывающие объекты, используемые в конкретной системе и учитывающие её особенности. На втором уровне описываются объекты, инвариантные к предметной области. Объекты третьего уровня описывают наиболее общие понятия и аксиомы, с помощью которых описываются объекты нижележащих уровней.

Одна из задач, решаемых подсистемой интеллектуального поиска и анализа документов, – поиск зависимостей и установление связей между документами, регламентирующими деятельность бизнес-системы. В результате анализа должна быть построена *система взаимосвязанных документов*:

- относящихся к определённым направлениям деятельности бизнес-системы (к определённым понятиям, объектам предметной области ИС);
- отражающих связи между этими понятиями (с каждым понятием может быть связан документ или совокупность документов, связи между документами отражают связи между понятиями);
- содержащих нормативную информацию, которая также может быть выделена на основе анализа содержания документов.

На основе построенной системы взаимосвязанных документов можно частично автоматизировать процесс анализа изменений предметной области и внесения изменений в модель предметной области ИС (т.е. реализовать поддержку процесса разработки и адаптации ИС). Таким образом, система управления документами становится не только «надстройкой» над ИС, позволяющей получать результаты обработки данных, хранящихся в базе данных (БД) ИС, в удобной для пользователей форме (подсистема репортинга), но и становится основой средств разработки ИС.

Результатом анализа документов должно стать автоматическое построение онтологии, вершинами которой будут сами анализируемые документы и их понятия. Схематически процесс создания онтологии показан на рис. 2. Процесс анализа текста и извлечения из него информации включает шаги, показанные на рис. 3. Документы с включённой в них семантической разметкой могут быть сохранены в базе индексированных документов (БИД, рис. 1).

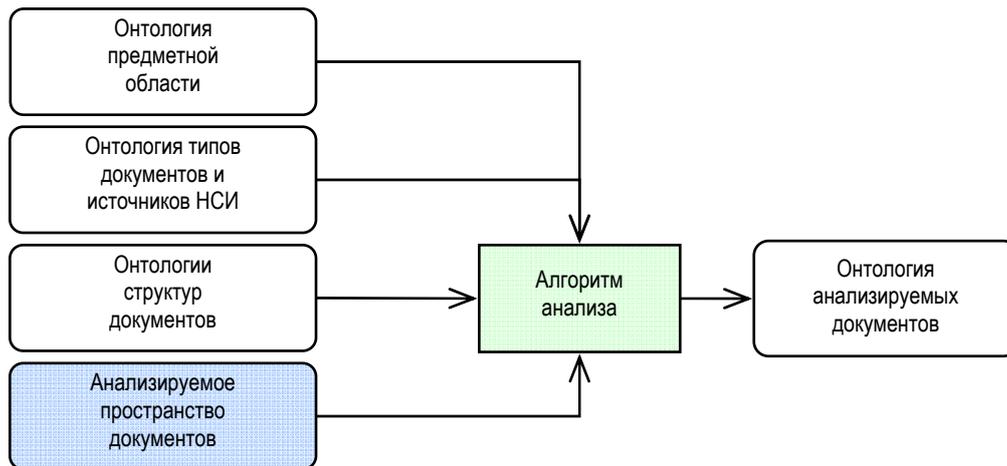


Рис. 2. Схема построения системы взаимосвязанных документов

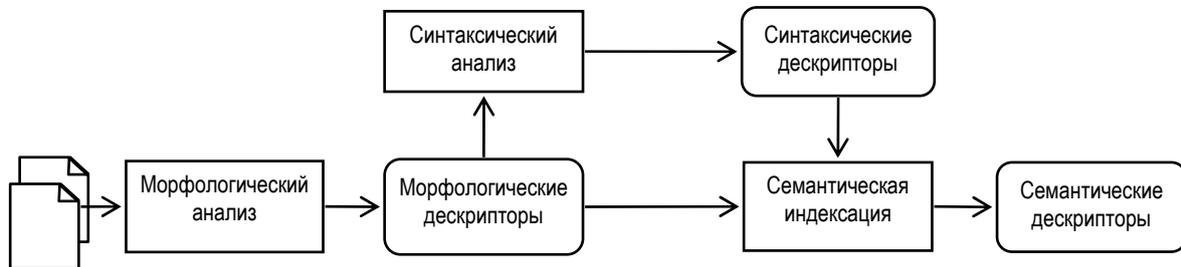


Рис. 3. Последовательность шагов анализа документа

Дескрипторы (морфологические, синтаксические и семантические) – это множества тегов. Морфологические и синтаксические дескрипторы могут быть помещены в реляционные таблицы. Результат семантического анализа (семантического индексирования) – семантические дескрипторы документов. Специальная разметка (метаданные) могут быть включены в текст электронных документов в различных форматах. Разметка позволяет связать документы и конкретные их фрагменты с онтологиями предметных областей, что в дальнейшем может быть использовано для поиска нужной информации, для генерации документов (например, пользовательской документации) с учётом установленных связей. Для семантического индексирования предлагается реализовать *агентный подход* (упрощённая структура агентной платформы показана на рис. 4), [Ланин, 2008; Ланин, 2009].

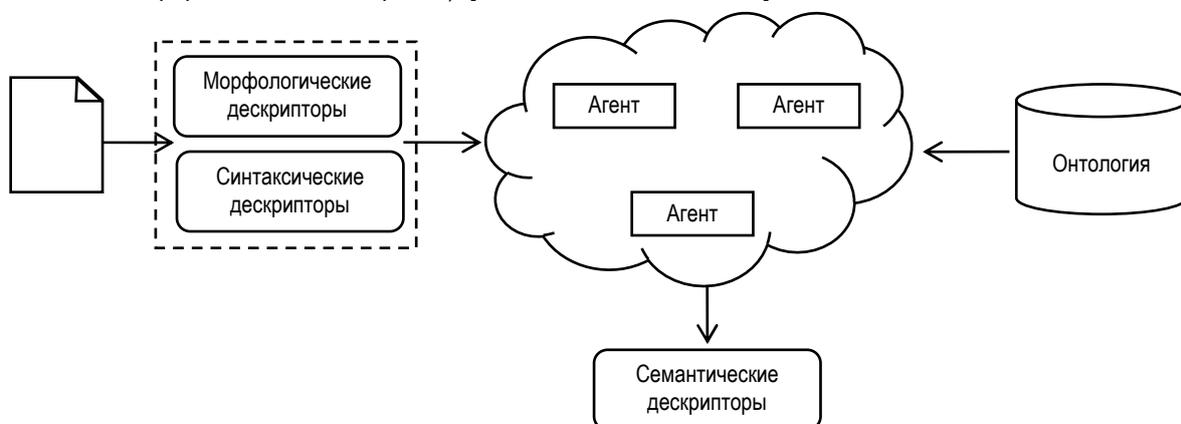


Рис. 4. Компоненты агентной платформы для решения задачи семантического индексирования документов

Каждый агент имеет доступ к общей онтологии, синтаксическим и морфологическим дескрипторам и к индексированным документам. Агенты играют различные роли в процессе анализа и индексации документов (рабочие агенты, агенты-лидеры), формируя команды в начале анализа документа. Число агентов в команде зависит от структуры и содержания документа. Агенты взаимодействуют друг с другом в процессе анализа.

Преставление *знаний агента* с помощью онтологии – наиболее выразительный способ, использующий все преимущества явного представления знаний. Достоинством данного способа является то, что для «доказательства» вершины онтологии мы можем применить различные средства. Например, это может быть простое совпадение ключевой фразы или обращение к базе данных ИС. Онтологии позволяют описать различные ситуации в случае, если не удаётся найти точное соответствие: мы можем найти обобщающее или конкретизирующее понятие и т.п.

Кроме автоматического построения системы взаимосвязанных документов, возможно «ручное» создание схемы документов пользователем. В офисные приложения встраиваются специальные модули расширения, которые позволяют связать фрагменты документа с понятиями и отслеживать связи между ними.

Для работы с документами используется *формальная модель документа*, являющаяся обобщением нескольких различных моделей документов [Ланин, 2010; Ланин, 2011]. Модель создаёт основу для описания документа, позволяющую подойти к решению задачи семантической индексации неструктурированных документов, являющейся ключевой в интеллектуальном управлении документами.

Электронный документ представляет собой набор структурных элементов, называемых в данной работе *фрагментами*. Примерами фрагментов могут служить таблица, заголовок, реквизиты углового бланка и т.д. Таким образом, документ может быть представлен тройкой вида:

$$d = (S(F, R), C),$$

где $S(F, R)$ – ориентированный гиперграф, вершинам которого сопоставлены элементы множества F (множество фрагментов документа), а R – это множество рёбер графа, соответствующее связям между фрагментами; элементы множества C представляют информационное содержание документа (его контент). Гиперграф $S(F, R)$ задаёт отношение между фрагментами документа. Ориентированность графа необходима, например, для отслеживания связей «часть-целое» между фрагментами. Вершины, входящие в ребро, пронумерованы, что позволяет установить порядок следования фрагментов в тексте документа. Очевидно, что ребро, включающее все вершины, соответствует документу в целом. Фрагменты могут быть двух видов: элементарные фрагменты представляют простейшие неделимые элементы (например, реквизиты документа (дата составления и пр.) или такие элементы оформления, структурирования как заголовки и т.п.), а составные фрагменты содержат в себе другие фрагменты.

На данной модели определены операции, которые могут быть выполнены над документами в ИС. Представленная модель позволяет формализовать алгоритмы интеллектуальной обработки электронных документов в информационных системах. Модель предоставляет широкие возможности для интеграции документов с онтологическими ресурсами. Практическая значимость модели находит подтверждение при работе с документами в современных форматах, таких как OpenXML и OpenDocument Format (ODF).

Генерация документации

Этап создания документации является необходимым при разработке любой информационной системы. Руководство программиста необходимо для обеспечения сопровождения системы. Пользовательская документация необходима для эффективного обучения пользователей работе с новой ИС. Однако очень часто разработчики программных систем и комплексов игнорируют данный процесс в связи с большим количеством времени, необходимого для создания качественной документации. Причём чем больше сложность системы, тем сложнее создание документации, поэтому возникают проблемы при эксплуатации ИС, её сопровождении. Особенно эти проблемы обостряются для систем, допускающих динамическую адаптацию, т.к. при настройке системы появляются расхождения её «поведения» с описаниями, данными в документации. Однако при этом, если система работает в режиме интерпретации, программный код интерпретатора не изменяется, следовательно, модифицировать при настройке системы в ходе её эксплуатации необходимо только пользовательскую документацию.

При создании компонента генерации документации пользователя решаются следующие задачи [Цыбин, 2007; Tsybin, 2008]:

- вывод в документ информации об объектах ИС и связях между ними;
- вывод в документ информации о нормативной основе автоматизации бизнес-процессов в системе, выполнения отдельных операций, полученной из документов, регламентирующих работу системы;
- вывод изображений экранных форм, с которыми работает пользователь ИС, с необходимыми пояснениями (описание пользовательского интерфейса системы);
- возможность управлять структурой и содержанием документации;
- возможность представления документа в различном формате;
- автоматическая генерация содержания и расстановка ссылок внутри документа.

На сегодняшний день на рынке программного обеспечения (ПО) наибольшее распространение получили программы автоматического документирования исходного кода проектов. В частности, это программы, работающие на основе фиксированных грамматик (программа NDoc для документирования исходного кода проектов на платформе dotNET; система JavaDoc для построения программной документации проектов на языке Java; система LPdoc создаёт документацию для языков Lisp и Prolog). Использование изменяемых грамматик – ещё один подход к реализации системы документирования программных проектов. В частности, существуют системы, полностью построенные на основе конечных автоматов.

В данном проекте создание документации реализовано на основе *шаблонов*, заранее заданных пользователем. Шаблоны содержат информацию о том, какую структуру должен иметь полученный документ. В данном случае грамматика – это набор правил извлечения информации из файла, структурированного определённым образом. Отличие от подхода, основанного на фиксированных грамматиках, состоит в том, что наборы правил по выделению информации из различных файлов вынесены в XML-файл и хранятся вне системы документирования.

Для решения задачи подготовки пользовательской документации использованы метаданные логической и презентационной моделей MDK (рис. 1). Логическая модель позволяет получить информацию о свойствах и связях сущностей. Эта информация используется при выводе в документ описаний объектов предметной области ИС и связей между ними. Презентационная модель хранит информацию обо всех

визуальных элементах интерфейса. ИС имеет настраиваемые формы ввода-редактирования данных для каждой сущности. Информация не только о свойствах и связях сущности, но и о внешнем виде форм также хранится в базе метаданных. Более того, презентационная модель связана с логической, что позволяет связывать описания элементов пользовательского интерфейса с соответствующими объектами, их описанием. Результаты семантического индексирования документов, полученные системой анализа документов, могут также использоваться при генерации пользовательской документации: в документы внесена семантическая разметка, которая позволяет установить связь нормативных и справочных документов и их фрагментов с соответствующими объектами предметной области.

Генерация документации разбита на несколько этапов. В процессе исследовательской работы была разработана и реализована двухслойная структура компонента документирования. Разработанные средства документирования работают в соответствии со схемой, показанной на рис. 5. На основе метаданных сначала генерируется *XML-документ* в соответствии с заданным пользователем шаблоном, а затем создаются конечные документы нужных форматов.

Для создания схемы шаблона XML-файла создан *интерфейс разработчика документации*, обеспечивающий удобный инструментарий для разметки документа. Путём сочетания вложенности различных элементов разработчик с помощью этого инструментария может задать *структуру документации* – создать *дерево содержания* документа. Определить структуру можно как визуально, так и с помощью текстового описания. Текстовое описание синхронизируется с построенным деревом. Кроме того, в интерфейсе реализована функция проверки синтаксических и семантических ошибок описания структуры документа. Интерфейс передаёт информацию о структуре документа в виде XML-файла структуры генератору XML-файла.

Генератор обходит *списки метаданных* в базе метаданных системы (БМД) в порядке, заданном структурой документа. Каждая вершина XML-файла описывает определённый объект системы, информация о котором выбирается из БМД. В генераторе XML-файла реализован также алгоритм поиска путей к сущности в *рекурсивном дереве объектов*. Задача разработки этого алгоритма возникла в связи с тем, что CASE-система METAS, в рамках которой был реализован исследовательский прототип средств документирования, позволяет на главной форме настроить дерево объектов (сущностей) ИС для удобного доступа пользователей к сущностям (для навигации по объектам системы). При этом может потребоваться описать путь к вершине, представляющей сущность в дереве, показать, как пользователь может до неё «добраться», используя связи, имеющиеся между объектами системы. Текущая настройка средств навигации, таким образом, отображается в генерируемом XML-файле. Для описания *экранных форм* ввода/редактирования данных в генератор добавлена функция «фотографирования» форм. Полученное изображение сохраняется в отдельном файле, а затем в XML-файл добавляется ссылка на данное изображение.

Нормативные и справочные документы с соответствующей семантической разметкой, позволяющей связать документы с описанием в БД объектов, работу с которыми они регламентируют, находятся в базе документов (БЖД). В создаваемый XML-файл заносится ссылка на соответствующий документ (его фрагмент).

Построенный XML-файл служит основой для генерации документации в различных форматах (документы MS Word, файлы HTML, Help-файлы). Анализатор выполняет разбор созданных ранее XML-файлов

описаний ИС и передаёт эти описания в специализированные генераторы конечных документов, поддерживающие определённый интерфейс. Анализатор использует технологию событийного разбора документов SAX.

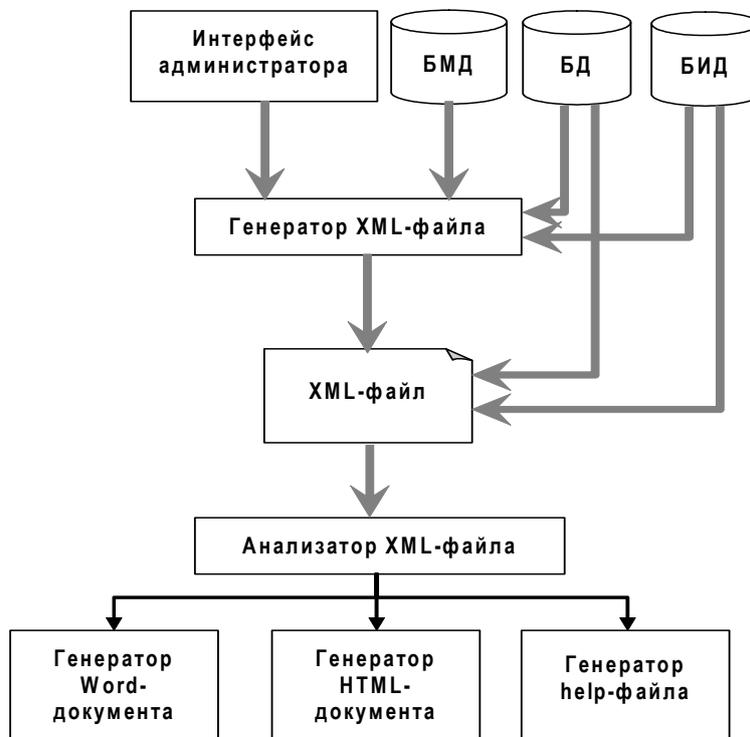


Рис. 5. Общая схема работы компонента документирования

Использование XML-файла документации позволяет хранить данные об информационной системе, её описание, независимо от представления документа, осуществлять настройку описания на нужный формат. Список генераторов конечных документов (т.е. форматов, в которых генерируется документация) может быть расширен.

Перспективной является возможность применения в генераторе шаблонов *изменяемой грамматики*. Алгоритм обработки элементов метаданных при построении документа меняется в зависимости от смысла отношений элементов метаданных. Была поставлена задача представления на декларативном уровне правил обработки сочетаний элементов метаданных, причём необходимо предоставить разработчику возможность изменять эти правила. При этом информация о связях элементов метаданных представляет собой метаметаданные, или метаданные второго уровня. Такой подход позволяет реализовать компонент документирования в виде отдельного модуля с возможностью подключения к любой информационной системе, основанной на списках. Данный подход предоставляет также более широкие возможности для приведения структуры документации к требуемому виду.

В ходе исследования разработана математическая модель, являющаяся основой применения изменяемых грамматик в алгоритме документирования, формально доказана обоснованность принятых проектных решений.

Подсистема генерации отчётов

Средства репортинга, реализованные в проекте создания CASE-технологии METAS, предназначенной для поддержания жизненного цикла динамически адаптируемых систем, включают два основных компонента: «Менеджер запросов» и «Менеджер отчётов» [Lanin, 2008].

Одно из основных требований к подсистемам создания запросов и отчётов – это возможность их разработки пользователями-непрограммистами. Такое требование может быть выполнено только за счёт введения дополнительного семантического слоя, основой которого могут быть метаданные, уже присутствующие в системе. Это даёт пользователю возможность работы с данными в соответствии с терминологией, принятой в конкретной предметной области ИС, позволяет абстрагироваться от физической структуры данных в БД.

Семантика предметной области учитывается при создании специальных предметно-ориентированных языков (DSL), для разработки которых в CASE-систему интегрируются языковой инструментарий [Лядова, Сухов, 2010].

Согласно предложенной концепции при создании запросов к БД пользователь выбирает сущности (объекты предметной области), участвующие в запросе, и связи между ними, которые должны быть учтены. Выбирает интересующие его атрибуты сущностей и другие параметры, влияющие на сортировку и группировку данных. Для разработки правил, задания условий в запросах применяются визуальные и текстовые DSL. В результате учёта требований пользователя и интерпретации метаданных построитель запросов автоматически генерирует SQL-запрос к базе данных информационной системы.

В качестве шаблонов (форм) отчётов могут быть использованы документы MS Word и рабочие книги MS Excel. Для обеспечения обмена с другими подсистемами распределённой ИС построитель имеет функции экспорта и импорта шаблонов отчётов.

В соответствии с предлагаемым подходом последовательность действий пользователя для создания нового отчёта выглядит следующим образом:

- 1) подготовка необходимых запросов с помощью «Менеджера запросов»;
- 2) подготовка и разметка шаблона офисного документа (включение в шаблон информации о диапазонах, куда будут помещаться данные при генерации документа на базе данного шаблона, элементов оформления и пр.), добавление формул для вычислений, диаграмм;
- 3) связывание запросов и соответствующих диапазонов документа, в которые должны быть помещены их результаты;
- 4) сохранение полученного шаблона в базе метаданных системы (БМД, рис. 1).

Предложенный подход обладает рядом преимуществ. Во-первых, для создания нового отчёта не требуется программирование и написание запросов на языке SQL. При необходимости аналитическая обработка информации может быть произведена в Microsoft Excel с помощью всех доступных средств пакета. Во-вторых, хранение шаблона отчёта в БМД делает этот отчёт частью метаданных, что позволяет тиражировать шаблоны отчётов вместе с запросами, на базе которых строятся отчёты между узлами распределённой информационной системы, обеспечивая единство системы документации.

Процедура генерации отчётов включает следующие шаги:

- 1) из базы метаданных извлекаются шаблоны отчёта, на их базе создаются документы;
- 2) выполняются запросы к БД ИС, связанные с каждым из шаблонов;
- 3) происходит вставка результатов выполнения запросов в размеченные диапазоны отчёта, при этом в нем могут производиться дополнительные вычисления, если в шаблон были включены соответствующие средства (формулы Excel, поля Word);
- 4) созданный отчёт может быть сохранён в БД как документ, он может быть также распечатан или передан по сети.

Благодаря реализованной возможности хранения электронных документов в БД (создан специальный тип данных), сгенерированные отчёты становятся частью данных ИС, над ними можно выполнять разрешённые в системе операции.

Заключение

Описанные выше средства реализованы в рамках создания CASE-технологии METAS, они прошли апробацию при реализации нескольких проектов. Возможности технологии значительно расширяются при интеграции в CASE-систему языкового инструментария MetaLanguage, предназначенного для разработки предметно-ориентированных языков, играющих роль метамоделей при создании ИС.

Структура компонента документирования обеспечивает дальнейшее расширение возможностей. В ходе исследований появились также новые идеи построения логики работы компонента документирования, основанные на использовании семантической разметки.

Средства анализа документов могут быть использованы как для снижения трудоёмкости работы пользователей с документами, так и для поддержки решения задачи анализа предметной области разработчиками, задачи документирования проекта на всех этапах.

В данном случае предлагается глубокая интеграция функциональных подсистем ИС, включающих как средства разработки, так и средства, с которыми работают «конечные пользователи». Это даёт возможность создания интеллектуальной CASE-технологии, предназначенной для создания динамически настраиваемых ИС, обладающих уникальными возможностями адаптации к меняющимся условиям эксплуатации на основе средств поддержки «обратной связи».

Благодарности

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Библиографический список

- [Арушанян, 2010] Арушанян О.Б., Богомолов Н.А., Волченскова Н.И., Ковалев А.Д. Средства автоматизации документирования больших комплексов программ // Вычислительные методы и программирование. 2010. Т.11. С.26-30.

-
- [Заболеева-Зотова, 2008] Заболеева-Зотова А.В., Орлова Ю.А. Автоматизация семантического анализа документации технического задания // Вестник компьютерных и информационных технологий. 2008. №9. С.26-34.
- [Заболеева-Зотова, 2009] Заболеева-Зотова А.В., Орлова Ю.А. Computer Support of Semantic Text Analysis of a Technical Specification on Designing Software // Intelligent Processing: suppl. to Int. Journal "Information Technologies and Knowledge" Vol. 3. - 2009. - Int. Book Series "Information Science & Computing", №9. С. 29-35.
- [Климов, 2010] Климов Б.А., Романов Д.Ю. Анализ систем документооборота в проектах по разработке программного обеспечения // Бизнес-информатика, №2(12), 2010. С. 15-23.
- [Ланин, 2008] Ланин В. Использование многоуровневого репозитория онтологий для анализа электронных документов // Труды международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т. 1. – М.: Физматлит, 2008. С. 202-206
- [Ланин, 2009] Ланин В. Онтологии как основа функционирования систем обработки электронных документов // Материалы Всероссийской конференции с международным участием «Знания-Онтологии-Теории». Новосибирск, 2009, Т. 2. С. 173-177.
- [Ланин, 2010] Ланин В.В. Модель документов, основанная на гиперграфах, в системах интеллектуального управления документами // Современные проблемы математики и её прикладные аспекты: сб. ст. / Пермский университет: Пермь. 2010. С. 81-86.
- [Ланин, Лядова, 2010] Ланин В.В., Лядова Л.Н. Управление документами в динамически адаптируемых системах, основанных на метамоделировании // Труды Конгресса по интеллектуальным системам и информационным технологиям «AIS-IT'10» Научное издание в 4-х томах. Т. 1. – М.: Физматлит, 2010. С. 510-518.
- [Ланин, 2011] Ланин В.В. Модель документа в адаптируемых информационных системах // Информатизация и связь. № 3, 2011. С. 29-31.
- [Лядова, 2007] Лядова Л.Н. Технология создания динамически адаптируемых информационных систем // Труды междунар. науч.-техн. конф. «Интеллектуальные системы» (AIS'07). Т. 2. – М.: Физматлит, 2007. С. 350-357.
- [Лядова, 2008] Лядова Л.Н. Метамоделирование и многоуровневые метаданные как основа технологии создания адаптируемых информационных систем // Advanced Studies in Software and Knowledge Engineering / International Book Series "Information Science & Computing", Number 4. Supplement to the International Journal "Information Technologies & Knowledge. Volume 2, 2008. Institute of Information Theories and Applications FOI ITHEA, Sofia, 2008. P. 125-132.
- [Лядова, 2010] Лядова Л.Н. О создании DSM-платформы на основе метамоделирования и онтологий // Сборник трудов Четвертой международной научно-технической конференции «Инфокоммуникационные технологии в науке, производстве и образовании» (Инфоком-4): Часть I / Ставрополь, 28-30 июня 2010 г. / Северо-Кавказский государственный технический университет. С. 222-227.
- [Лядова, Сухов, 2010] Лядова Л.Н., Сухов А.О. Визуальные языки и языковые инструментари: методы и средства реализации // Труды Конгресса по интеллектуальным системам и информационным технологиям «AIS-IT'10» Научное издание в 4-х томах. Т. 1. – М.: Физматлит, 2010. С. 374-382.
- [Мамаев, 2008] Мамаев А.С., Прошин А.А., Флитман Е.В. Создание системы документирования и контроля распределённых информационных систем // Современные проблемы дистанционного зондирования земли из космоса: Сб. научн. статей. Вып. 5, 2008. М.: ООО «Азбука-2000». С. 557-560.
- [Смирнов, 2012] Смирнов М.Н., Соколов Н.Е., Романовский К.Ю. DocLineFM: среда разработки повторно-используемой документации семейств программных продуктов на базе пакета Adobe FrameMaker // Системное программирование. Вып. 6. 2012. С. 80-98.
- [Сонис, 2006] Сонис Р.Г. Совершенствование элементов системы управления электронным документооборотом на основе методов функциональной стандартизации и технологии открытых систем: Автореферат диссертации на соискание учёной степени кандидата технических наук по специальности 05.13.05 и 05.13.15. Москва, 2006.

- [Суясов, 2007] Суясов Д.И., Шалыто А.А. Автоматическое документирование программных проектов на основе автоматного подхода: [Электронный документ] (<http://is.ifmo.ru>).
- [Тарасенко, 2009] Тарасенко А.В. Разработка и исследование методов и моделей автоматической проверки текстов на соответствие требованиям технической документации: Автореферат диссертации на соискание учёной степени доктора технических наук по специальности 05.13.17. Таганрог, 2009.
- [Цыбин, 2007] Цыбин А.В. Автоматическая генерация документации пользователя в информационных системах, управляемых метаданными // Сб. тезисов конференции-конкурса «Технологии Microsoft в теории и практике программирования» / Новосибирск: НГУ, 2007. С. 78-80.
- [Черников, 2010] Черников Б.В. Методология формирования документационного обеспечения деятельности организации на основе лексикологического синтеза: Автореферат диссертации на соискание учёной степени доктора технических наук по специальности 05.25.05. Москва, 2010.
- [Fulford, 2001] Fulford H. Developing Document Analysis and Data Extraction Tools for Entity Modelling // NLDB 2000, LNCS 1959, 2001. P. 265-275.
- [Kuo, 2004] Kuo J.-Y. A document-driven agent-based approach for business processes management // Information and Software Technology, 46 (2004). P. 373-382.
- [Lanin, 2008] Lanin V. Architecture and Implementation of Reporting Means in Adaptive Dynamically Extended Information Systems // International Journal "Information Technologies & Knowledge" / Sofia (Bulgaria) – Vol. 2/2008, Number 3, P. 273-277.
- [Sierra, 2004] Sierra J.L., Fernández-Manjón B., Fernández-Valmayor A., Navarro A. A Document-Oriented Approach to the Development of Knowledge Based Systems // LNAI 3040, 2004. P. 16-25.
- [Song, 2007] Song D., Lau R.Y.K., Bruza P.D., Wong K.-F., Chen D.-Y. An intelligent information agent for document title classification and filtering in document-intensive domains // Decision Support Systems, 44 (2007). P. 251-265.
- [Tsybin, 2008] Tsybin A., Lyadova L. Software Testing and Documenting Automation // International Journal "Information Technologies and Knowledge" / Sofia (Bulgaria) – Vol. 2/2008, Number 3. P. 267-272.

Информация об авторах



Вячеслав Ланин – старший преподаватель кафедры математического обеспечения вычислительных систем ПГНИУ; Россия, 614990, г. Пермь, ул. Букирева, 15; Lanin@Perm.ru.

Область научных исследований: интеллектуальный поиск и анализ документов; средства разработки программного обеспечения, онтологическое моделирование.



Людмила Лядова – доцент кафедры информационных технологий в бизнесе НИУ ВШЭ – Пермь; Россия, г. Пермь, 614070, ул. Студенческая, 38; LNLyadova@gmail.com.

Область научных исследований: инструментальные средства разработки программного обеспечения; метамоделирование.



Антон Цыбин – ведущий программист ООО «Парагон», соискатель ПГНИУ; Россия, 614990, г. Пермь, ул. Букирева, 15; magicdr@mail.ru.

Область научных исследований: проектирование информационных систем; инструментальные средства разработки программного обеспечения.