МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«МИРЭА – Российский технологический университет»
РТУ МИРЭА

Лекция №2
Основные понятия проектирования

Методы и средства проектирования информационно-аналитических систем
(наименование дисциплины (модуля) в соответствии с учебным планом)

Уровень	специалитет *(бакалавриат, магистратура, специалитет)*
Форма обучения	очная *(очная, очно-заочная, заочная)*
Направление(-я) подготовки	10.05.04 «Информационно-аналитические системы безопасности» *(код(-ы) и наименование(-я))*
Институт	Институт кибербезопасности и цифровых технологий (ИКБ) *(полное и краткое наименование)*
Кафедра	Информационно-аналитические системы кибербезопасности (КБ-2) *(полное и краткое наименование кафедры, реализующей дисциплину (модуль))*

Используются в данной редакции с учебного года 2023/24
(учебный год цифрами)

Проверено и согласовано «____» _____20___г. _____ _____
(подпись директора Института/Филиала с расшифровкой)

Москва 2024 г.

1. Понятия проектирования.

Основные понятия в последние годы не претерпели сильных изменений, формулировки стали более точными и лаконичными, исключающими неоднозначность понятий. Наиболее полные определения представлены в Федеральных законах Российской Федерации и стандартах.

Информация – «сведения (сообщения, данные) независимо от формы их представления» (Федеральный закон Российской Федерации от 27 июля 2006 г. N 149-ФЗ. Об информации, информационных технологиях и о защите информации – ФЗ-149).

Информационные технологии – «процессы, методы поиска, сбора, хранения, обработки, предоставления, распространения информации и способы осуществления таких процессов и методов» (ФЗ-149).

Информационная система – «совокупность содержащейся в базах данных информации и обеспечивающих ее обработку информационных технологий и технических средств» (ФЗ-149).

Проектирование информационных систем – это упорядоченная совокупность методологий и средств создания или модернизации информационных систем.

Управление информационными системами – «применение методов управления процессами планирования, анализа, дизайна, создания, внедрения и эксплуатации информационной системы организации для достижения ее целей» (ГОСТ РВ 51987-2002) или «структура взаимоотношений и процессов выбора вектора развития предприятия и его управления, направленных на увеличение его стоимости при сбалансированном риске в сфере информационных и смежных технологий» (CobiT)

Жизненный цикл информационных систем – «развитие рассматриваемой системы во времени, начиная от замысла и кончая списанием».

Модель жизненного цикла – «структурная основа процессов и действий, относящихся к жизненному циклу, которая также служит в качестве общего эталона для установления связей и понимания (ГОСТ Р 59330—2021 «Защита информации в процессе управления моделью жизненного цикла системы»)

Архитектура информационных систем – это концепция, определяющая модель, структуру, выполняемые функции и взаимосвязь компонентов информационной системы.

Бизнес-процесс – это цепочка взаимосвязанных действий, направленных на создание товарной продукции или услуги.

Регламент бизнес-процесса – это четко определенный порядок выполнения бизнес-процесса, определяющий состав и действия участников.

Модель данных – это система организации данных и управления ими.

Методология проектирования информационных систем – это совокупность принципов проектирования (моделирования), выраженная в определенной концепции.

Средства моделирования – это программы описания и моделирования систем.

Типовое проектное решение (ТПР) – это многократно используемое проектное решение.

Нотации – это определенные способы представления элементов информационной системы.

Реинжиниринг бизнес-процессов – это фундаментальная реорганизация бизнес-процессов с целью повышения их эффективности.

Системный подход – процесс рассмотрения любой системы в качестве совокупности взаимосвязанных элементов.

Процессный подход – представление любой системы в качестве совокупности процессов.

Функциональный подход – предусматривает четкое закрепление за каждой структурной единицей набора функций.

Техническое задание – документ, оформленный в установленном порядке и определяющий цели создания ИАС, требования к ИАС и основные исходные данные, необходимые для ее разработки, а также планграфик создания ИАС (ГОСТ 34.003-90).

2. Типология проектов по созданию ИС

Проект создания ИС может быть индивидуальным или типовым.

Индивидуальный проект – подразумевает разработку ИС, как правило с помощью специалистов самой организации.

Типовое проект ИС предполагает создание системы из готовых типовых проектных решений.

Типовое проектное решение (ТПР) – это тиражируемое (пригодное к многократному использованию) проектное решение. Выделяют следующие классы ТПР:

• элементные ТПР – типовые решения по задаче или по отдельному виду обеспечения задачи (информационному, программному, техническому, математическому, организационному);

• подсистемные ТПР – в качестве элементов типизации выступают отдельные подсистемы, разработанные с учетом функциональной полноты и минимизации внешних информационных связей;

• объектные ТПР – типовые отраслевые решения, которые включают полный набор функциональных и обеспечивающих подсистем ИС.

Каждое типовое решение предполагает наличие, кроме собственно функциональных элементов (программных или аппаратных), документации с детальным описанием ТПР и процедур настройки в соответствии с требованиями разрабатываемой системы.

Для реализации типового проектирования используются два подхода: параметрически-ориентированный и модельно-ориентированный.

Параметрически-ориентированное проектирование включает следующие основные этапы:

декомпозиция проектируемой ИС на множество составляющих компонентов (подсистем, программных модулей и т.д.);

выбор и приобретение из имеющихся на рынке ТПР, необходимых для реализации выделенных компонентов;

настройка (доработка) приобретенного ТПР на особенности конкретного предприятия с помощью обслуживающей организации (либо самостоятельно с помощью штатных IT-специалистов).

Выбор и приобретение ТПР подразумевает выполнение следующих шагов:

определение критериев оценки компонентов ИС с точки зрения решения поставленных задач;

анализ и оценка доступных ТПР по сформулированным критериям;

выбор и закупка наиболее подходящего пакета.

Критерии оценки ТПР делятся на следующие группы:

• назначение и возможности;

• отличительные признаки и свойства;

• требования к техническим и программным средствам;

• документация;

• факторы финансового порядка;
• особенности установки;
• особенности эксплуатации;
• помощь поставщика по внедрению и поддержке;
• оценка качества решения и опыт его использования;
• перспективы развития.

Внутри каждой группы критериев выделяется некоторое подмножество частных показателей, детализирующих каждый из десяти выделенных аспектов анализа выбираемых ТПР. Достаточно полный перечень показателей можно найти в литературе. Числовые значения показателей для конкретных ТПР устанавливаются экспертами по выбранной шкале оценок (например, 10-балльной). На их основе формируются групповые оценки и комплексная оценка пакета (путем вычисления средневзвешенных значений). Нормированные взвешивающие коэффициенты также получаются экспертным путем.

Модельно-ориентированное проектирование заключается в адаптации состава и характеристик типовой ИС к модели объекта автоматизации. Технология проектирования в этом случае должна обеспечивать единые средства для работы как с моделью типовой ИС, так и с моделью конкретного предприятия.

Модельно-ориентированное проектирование ИС предполагает, прежде всего, построение модели объекта автоматизации с использованием специального программного инструментария (например, SAP Business Engineering Workbench (BEW), BAAN Enterprise Modeler).

Возможно также создание ИС на базе типовой модели ИС из репозитория, который поставляется вместе с программным продуктом и содержит как базовую (эталонную) модель ИС, так и конфигурации для определенных отраслей или типов производства.

3. Подходы к проектированию систем

Стихийная («лоскутная») автоматизация (подход «снизу-вверх»)

Индустрия разработки ПО начала зарождаться в середине 50-х годов XX в., однако почти до конца 60-х ей не уделялось серьезного внимания, поскольку ее доля в компьютерном бизнесе была слишком мала. Серьезный рост начался в 70-х годах XX в., начиная с принятого фирмой IBM в 1969 г.

решения о развязывании цен (раздельном назначении цен на аппаратуру, ПО и услуги), и продолжился до появления персонального компьютера.

На первом этапе основным подходом к проектированию ИС был метод «снизу-вверх». ИС создавалась в виде набора приложений, наиболее важных в данный момент для поддержки деятельности организации. Основной целью этих проектов было обслуживание текущих потребностей конкретного предприятия, а не создание тиражируемых продуктов. Такой подход отчасти сохраняется и сегодня.

Основной недостаток метода: возникновение проблем при объединении существующих систем.

Системное проектирование (подход «сверху-вниз»)

Противоположностью стихийной автоматизации является системное проектирование, или автоматизация «сверху-вниз». Смысл системного проектирования – реорганизация управления и перепроектирование всей корпоративной информационной системы, которые наилучшим образом достигают целей управления. Этапы системного проектирования:

определение целей и задач управления организацией;

создание модели организации, главное требование к которой – системная целостность; каждое изменение элемента модели требует перепроверки и согласования как «сверху-вниз», так и «снизу-вверх»;

создание корпоративной ИС на основе этой модели.

Основной недостаток системного подхода к проектированию – трудоемкость поддержания целостности модели.

В настоящий момент большинство организаций уже имеет ИС, в различной степени автоматизирующие процессы в них протекающие. Поэтому типичными в настоящее время являются следующие проекты:

• по разработке новых ИС и их интеграции с существующими ИС;

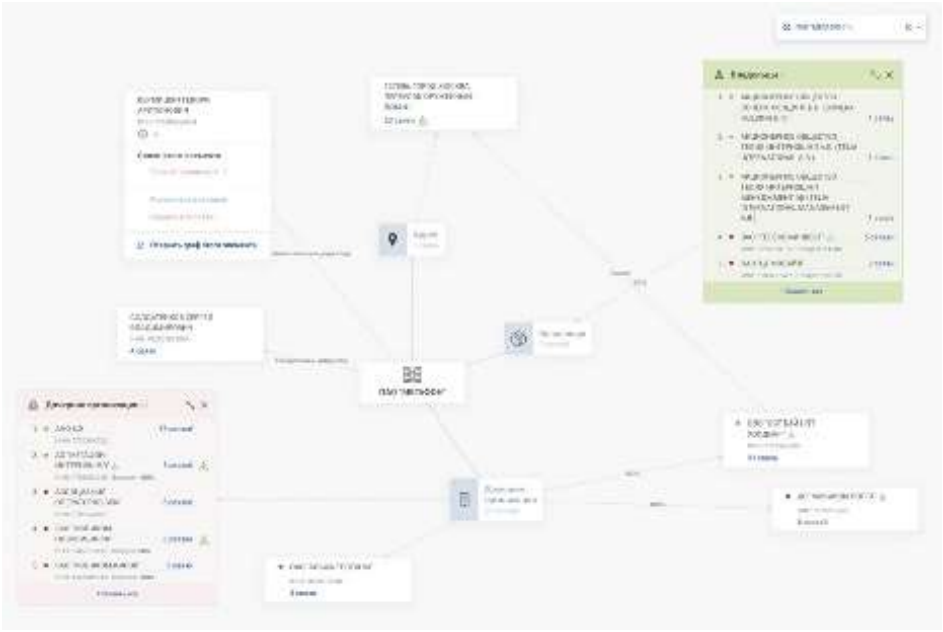
• по разработке новых ИС с целью замены существующих ИС;

• по модернизации (наращиванию функциональности, развитию) существующих ИС.

4. Системы организации знаний

Чем больше становится данных, тем быстрее скорость увеличения их количества. Это очень серьезный тренд – и психологически, и технологически. Такой большой объем новой информации становится все сложнее передавать, обрабатывать, сохранять, несмотря на существенное увеличение

производительности оборудования. Но самая большая проблема заключается не в количестве данных, а в их неструктурированности. Данные появляются из различных источников, в разных форматах, в разное время. Поэтому перед использованием в практических задачах данные упорядочивают, преобразуют, приводят в форму, эффективную для хранения и использования. Традиционным способом хранения и использования данных являются реляционные базы данных, в которых данные хранятся в виде связанных таблиц. Однако не всегда табличные данные эффективны в использовании, поэтому стали появляться альтернативные формы, например, системы организации знаний (Knowledge Organization Systems, KOS), как правило базирующиеся на графах знаний.



Для хранения знаний используются разные структуры:

управляемые словари: обеспечивают способ организации знаний для последующего поиска, используются в схемах предметной индексации, предметных рубриках, тезаурусах, таксономиях и других системах организации знаний,

тезаурусы: объединяют термины в группы по определенному признаку, например, с учетом похожести (синонимы),

таксономии: категорированные слова, упорядоченные по иерархическому признаку,

онтологии: формальное описание знаний из какого-то домена (предметной области) с учетом имеющихся сложных правил и связей между элементами, позволяющим сделать автоматическое извлечение знаний,

датасеты: наборы машиночитаемых данных.

Онтологии служат для систем организации знаний и применяются в тех областях, где требуется обнаружить новые факты, выявить скрытые взаимосвязи между элементами (например, рекомендательные и экспертные системы). Это альтернатива классическим базам данных, которые используют «гипотезу закрытого мира», когда все, чего нет в базе данных – не существует. В противоположность этому онтологии используют «гипотезу открытого мира», то есть если чего-то нет в базе знаний, то это не обязательно не существует, а просто не описано.

Системы организации знаний на основе онтологий уже очень распространены и используются во многих отраслях. Самый яркий пример это knowledge graph для поиска информации в Интернет, благодаря этой технологии качество поиска стало очень высоким. Другие примеры использования онтологий на практике:

банки используют графы знаний для анализа транзакций (fraud detection),

в консалтинге используются графы на основе юридических документов,

в здравоохранении используются накопленные сведения на основе данных о здоровье пациентов, Health Electronic Record,

в промышленности графы знаний используются для анализа цепочек поставщиков (supply-chain management),

в целом для Индустрии 4.0 характерно взаимодействие киберфизических систем между собой, что приводит к автоматизации и необходимости управлять знаниями,

во многих отраслях базы знаний используются для организации работы чат-ботов, в том числе и для обработки сложных запросов на естественном языке (например, сервис asknow),

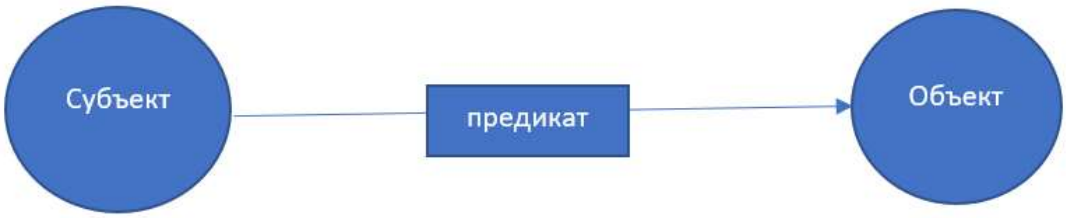
онтологии могут применяться для широкого круга задач обработки естественного языка: аннотирование текстов с помощью онтологий, извлечение знаний, NER, Named Entity Linking, Relation Linking, автоматический вывод новых знаний, ризонинг (reasoning). Направление SemTech переживает в настоящее время бурное развитие, в том числе в России, но информации на русском языке по данной тематике по-прежнему очень мало.

В информационной безопасности тоже есть задачи, где онтологии приносят пользу. Уменьшение возможностей для атак злоумышленников за счет установки патчей и обновления ПО никогда не дает 100%-защиту, поскольку остаются уязвимости, связанные с небезопасным пользовательским

поведением, небезопасной конфигурацией инфраструктуры, ошибками в настройках внедренных средств защиты, нестойкими паролями и недостаточным контролем за привилегированным доступом. Противодействовать zero-day атакам очень сложно, так как в защищающихся системах не существует правил для обнаружения таких атак, распознавать атаку и реагировать надо "на лету". Одним из способов распознать атаку при отсутствии готового шаблона распознавания является использование накопленных знаний и «логический вывод» (ризонинг) из этих знаний с учетом имеющейся информации о происходящем событии. Способом хранения таких знаний могут быть онтологии, которые хранят информацию о взаимосвязях различных сущностей между собой.

Что такое онтологии?

В математическом основании онтологии лежит так называемая дескриптивная логика (раздел математики), предполагающая, что любая информация, высказанная на естественном языке, может быть представлена в виде цепочки триплетов:



Отрывок из стихотворения «Дом, который построил Джек…» (пер. С.Я. Маршака).

Описанные в стихах отношения между различными сущностями мы можем представить в виде онтологии.

Онтология представляется в виде графа, вершины которого это сущности, а ребра – отношения между сущностями. Считается, что любое утверждение на естественном языке можно представить в виде простых предложений, из которых можно извлечь сущности и отношения между ними. Есть два основных инструмента: RDF (Resource Description Framework) или OWL (Ontology Web Language). OWL позволяет дополнительно описывать логические правила над данными. Онтологии (в отличие от обычных баз данных) позволяют находить скрытые данные. Обычные БД хорошо подходят для поиска конкретной информации, а базы знаний нужны там, где надо выявлять новые знания, например, в системах поддержки принятия решений (экспертных системах).

Примеры RDF хранилищ – Virtuoso, 4store, stardog.

Сила онтологии проявляется в том случае, если подробно и качественно описаны взаимосвязи между ее элементами, с использованием математического аппарата дескриптивной логики. Например, для отношений можно задать их свойства (функциональное, транзитивное, рефлексивное). И тогда можно автоматически из онтологии извлекать факты, этот процесс называется ризонинг (reasoning), есть типовые алгоритмы ризонинга, основанные на графах. Возможные применения: уточнение характеристик объекта и выделение из набора похожих объектов уникального, поиск похожих объектов, «понимание текста» и отнесение текста к определенному классу, помощь в NLP задачах (NER, Relation Extraction), анализ корневых причин, выявление паттернов в данных. Наиболее популярный редактор онтологий, поддерживающий ризонинг, это Protégé. Существуют и другие инструменты, например: IBM Watson, Wolfram Alpha.

Создание процесса онтологий обычно выполняется вручную, силами экспертов. Однако есть и примеры автоматизированного создания онтологий на основе имеющихся баз знаний. Открытые графы знаний (по состоянию на конец 2021 года, по данным The Linked Open Data Cloud):

DBpedia;

Yago + wordnet.princeton.edu;

WikiData;

Открытые базы знаний, в том числе предметные и отраслевые базы знаний, например, медицина: BioPortal, Bio2RDF (https://www.sciencedirect.com/science/article/pii/S1532046408000415?via%3Dihub), PubMed.

Описанные выше инструменты содержат различные методы работы с онтологиями. Однако наиболее популярным инструментом является Protégé, который мы далее рассмотрим подробнее.