

ПРОЦЕСС ПОСТРОЕНИЯ МОДЕЛЕЙ «СТРУКТУРА-СВОЙСТВО» ДЛЯ МОЛЕКУЛЯРНЫХ ГРАФОВ И ЕГО UML-ПРЕДСТАВЛЕНИЕ

М.И.Кумсков, Е.Н.Сенкова, Л.А.Пономарева, К.Ю.Богачев

В работе описан процесс построения модели идентификации и верификации зависимостей «структура-свойство» (QSPR) и «структура-биологическая активность» (QSAR). Процесс представлен в визуальной нотации UML (Унифицированного Языка Моделирования) и может быть «опубликован» в виде веб-сайта как справочник работ по конкретному QSAR-проекту.

При описании процесса использовалась среда IBM Rational Method Composer - инструмента настройки и публикации веб-сайтов на основе Rational Unified Process с использованием UML.

Основные элементы представления процесса на UML: роль, задача, артефакт. В результате появляется возможность настраивать процесс решения задачи на особенности конкретного проекта и «публиковать» соответствующие регламенты работ для исполнителей QSAR-проекта.

Для обеспечения эффективных коммуникаций исполнителей проекта создан веб-сайт, содержащий основную документацию по проекту, текущие планы работ, включая состояния работы и её исполнителей, запросы на изменения и их состояния, основные публикации по проекту (<http://qsar-msu.ru>)

Одним из интенсивно развивающихся направлений использования математических методов в химии является поиск зависимостей между структурами химических соединений и их свойствами. Это направление получило название QSAR (Quantitative Structure Activity Relationships). QSAR-модели используются для анализа состава молекулярных баз данных и поиска в них потенциально активных веществ. Применение методов QSAR при создании новых соединений с заданными свойствами позволяет значительно сократить затраты на их поиск и осуществлять более целенаправленный синтез только соединений, потенциально обладающих заданным набором свойств.

Целью работы является разработка методики визуального представления процесса решения QSAR-задач на UML для автоматической генерации регламентов работ исполнителей проекта (решения QSAR-задачи) в виде методического веб-сайта. Было проведено:

- 1) создание информационной системы для поддержки QSAR-проекта в виде веб-сайта на основе CMS (Control Management System);
- 2) описание и UML-визуализация процесса решения задачи «структура-свойство» (QSAR-задачи);
- 3) построение моделей в нотации UML (Унифицированного языка моделирования);
- 4) использование системы Rational Method Composer для представления регламентов работ;

1. Задача «структура-свойство» (QSAR-задача) для молекулярных графов

Пусть $G = \{E, V\}$ – помеченный граф, вершины которого интерпретируются как атомы молекулы, а ребра – как валентные связи между парами атомов. Метки вершин и ребер (числа или символы) кодируют атомы и связи различной химической природы. В качестве меток вершин могут быть использованы любые характеристики соответствующих атомов (например, трехмерные координаты, символ химического элемента, заряд ядра, поляризуемость, атомный вес, атомный радиус и др.), а в качестве меток ребер – любые характеристики соответствующих связей (кратность, длины, порядки связей, полученные из квантово-химических расчетов, и т.д. [10]).

Пусть задана обучающая (или эталонная) выборка - база данных из N химических соединений, где:

1) i -ое соединение представлено меченым молекулярным графом G_i , имеющим укладку в трехмерном пространстве (т.е., для каждой вершины в качестве меток заданы ее трехмерные координаты);

2) либо i -ое соединение отнесено к C_i - одному из K классов активности (например, «активных», «слабоактивных», «неактивных» веществ) согласно исследуемому свойству, либо для него задано численное значение исследуемого свойства A_i .

Необходимо построить классифицирующую функцию F , получающую в качестве аргумента произвольный молекулярный граф с метками того же типа, и «наилучшим образом» относящую это соединение к одному из классов активности, либо «наилучшим образом» предсказывающую численное значение исследуемого свойства.

Какая из классифицирующих функций «лучше», позволяет определить функционал качества $\varphi(F)$. Например, в качестве функционала качества можно использовать процент верно классифицированных функцией F молекул из обучающей выборки:

$$\varphi(F) = 1 - \frac{\sum_{i=1}^N \varepsilon_i}{N}, \text{ где } \varepsilon_i = \begin{cases} 0, & \text{если } F(G_i) = C_i \\ 1, & \text{в противном случае} \end{cases}, \quad (1)$$

или, в случае, когда функция должна предсказывать численное значение свойства,

$$\varphi(F) = 1 - \frac{\sum_{i=1}^N (F(G_i) - A_i)^2}{\sum_{i=1}^N A_i^2}. \quad (2)$$

Поставленную таким образом задачу поиска классифицирующей функции будем называть **задачей «структура-свойство»** или **QSAR-задачей**.

Дескриптором будем называть какое-либо свойство, численное значение которого может быть вычислено для произвольного молекулярного графа G .

Алфавитом дескрипторов будем называть множество всех дескрипторов, используемых для анализа обучающей выборки, обозначенных различными символьными метками.

Пусть алфавит дескрипторов состоит из M элементов. **Вектором признаков** молекулярного графа G будем называть вектор $\bar{x} = (x_1, \dots, x_M) \in R^M$, где x_j - значение j -ого дескриптора, вычисленное для G .

Матрицей «молекула-признак» (матрицей признаков) для рассматриваемой обучающей выборки будем называть матрицу размера $N \times M$, в i -ой строке которой стоит вектор признаков $\bar{x}_i = (x_{i1}, \dots, x_{iM})$ i -ого соединения.

Основные этапы решения QSAR-задачи

Задача «структура-свойство» естественно разбивается на два этапа: этап описания и этап поиска модели функциональной зависимости

Этап описания:

Исходя из формата молекулярных графов (типа меток вершин и ребер) выбирается алфавит дескрипторов A . На основе этого алфавита строится отображение из множества молекулярных графов в признаковое пространство R^M и формируется матрица «молекула-признак» для обучающей выборки.

Этап поиска модели функциональной зависимости:

В результате анализа матрицы «структура-свойство» на признаковом пространстве строится модель функциональной зависимости - классифицирующая функция F с наилучшей прогностической способностью, т.е. с наибольшим значением функционала качества.

После выделения фрагментов каждому фрагменту сопоставляется **структурный дескриптор**, значение которого соответствует либо наличию или отсутствию данного фрагмента в молекулярном графе, либо количеству повторений фрагмента. В первом случае получаем дескриптор, принимающий логические значения, во втором - целые неотрицательные.

В общем виде метод состоит из следующих этапов:

Проводится дополнительная классификация атомов (вершин молекулярного графа) на основе их локальных свойств (заряда, эксцентриситета вершины, каких-либо топологических свойств). В результате этого метка каждой вершины заменяется на другую, содержащую информацию о локальных свойствах.

В молекулах выбираются структурные фрагменты (атомы, цепочки связанных атомов, группы атомов).

Каждому структурному фрагменту сопоставляется символьное имя - тип фрагмента (например, если рассматриваются цепочки атомов, то «именем» цепочки может служить объединение символьных меток входящих в нее атомов).

Множества фрагментов для всех молекулярных графов выборки объединяются.

Для каждого молекулярного графа и каждого фрагмента находим значение соответствующего структурного дескриптора (либо количество повторений, либо наличие/отсутствие в молекулярном графе). В итоге получаем матрицу «молекула-признак»

Этап поиска функциональной зависимости

После формирования матрицы «молекула-дескриптор» для обучающей выборки необходимо построить классифицирующую функцию $F(x_1, x_2, \dots, x_M)$, где (x_1, x_2, \dots, x_M) – вектор признаков молекулярного графа. Причем, построенная функция должна обеспечивать лучшее значение функционала качества.

Обычно вид классифицирующей функции F заранее задается (например, функция может быть линейной, квадратичной и др.) и зависит от ряда параметров, которые определяются по обучающей выборке соединений. Чаще всего в качестве F используется линейная функция. Получаемое уравнение называют **линейной регрессионной моделью**.

Следует отметить, что в задаче «структура-свойство» число дескрипторов M , как правило, значительно превышает число молекул в обучающей выборке ($M \gg N$), что затрудняет анализ матрицы «молекула-признак». Для того чтобы сократить число дескрипторов, необходимо рассматривать лишь наиболее **информативные** из них, т.е. те, которые потенциально будут значимы при построении классифицирующей функции на признаковом пространстве. Это можно сделать как на этапе описания (например, при эволюционном формировании дескрипторов), так и на этапе анализа матрицы признаков. В данной работе предлагается отбирать наиболее информативные дескрипторы путем взаимодействия этих двух этапов – использования результатов этапа анализа на этапе описания.

Исходной информацией являются описания объектов, ситуаций, предметов, явлений или процессов S в виде векторов значений признаков $S = (x_1(S), x_2(S), \dots, x_n(S))$, где признаки $x_i, i = 1, \dots, n$, характеризуют различные стороны-свойства S . У объектов S существует "основное свойство" $y(S)$, которое для части объектов S_1, S_2, \dots, S_m предполагается известным, а для части объектов нет. Задача распознавания (прогноза, идентификации, "классификации с учителем") состоит в определении значения свойства $y(S)$ по информации $S_1, S_2, \dots, S_m, y(S_1), y(S_2), \dots, y(S_m)$ (*обучающей или эталонной выборке*).

Признаки могут быть числовыми (задающими степень выраженности какого-либо свойства), бинарными ("есть" или "нет" свойство), номинальными (обозначающими наличие различных свойств без числовой оценки - пол, цвет, и т.д.).

2. Процесс построения моделей «структура-свойство»

Бизнес-процесс – это описание последовательности работ, направленной на достижение определенной цели, обладающее следующими признаками:

- 1) интегрированное описание функций, документов и организационных подразделений;
- 2) иерархический характер описания.

Характеристики процесса:

- может быть представлен в графической нотации (в виде визуальной схемы);
- может быть улучшен путем выявления соотношений между его компонентами и последующим изменением этих компонентов и их взаимосвязей.

Модель бизнес-процесса должна давать ответы на вопросы:

- 1) Какие процедуры (функции, работы) необходимо выполнить для получения заданного конечного результата?
- 2) В какой последовательности выполняются эти процедуры?
- 3) Какие механизмы контроля и управления существуют в рамках рассматриваемого бизнес-процесса?
- 4) Какие входящие документы/информацию использует каждая процедура процесса?
- 5) Какие исходящие документы/информацию генерирует процедура процесса?
- 6) Какие ресурсы необходимы для выполнения каждой процедуры процесса?
- 7) Какая документация/условия регламентирует выполнение процедуры?
- 8) Какие параметры характеризуют выполнение процедур и процесса в целом?

В процессе решения задачи «структура-свойство» возникла необходимость реализовать следующие возможности:

- успешное взаимодействие нескольких участников коллектива, каждый из которых выполняет свой участок работы, но обязан представлять ситуацию в целом
- экономия времени и ресурсов коллектива при введении в курс дел новых участников
- повышение эффективности опытных участников коллектива (особенно в условиях большой скорости обновления участников)

Необходимо создание «базы знаний» элементов процесса, которую можно просматривать, управлять и «публиковать» на веб-сайте проекта. Эта база должна включать информацию из базы знаний процессов организации, включающей в себя руководства, шаблоны, инструкции, процедуры, учебные материалы и другие компоненты. Эта база служит основой для построения описания процессов. IBM Rational Method Composer RMC разрабатывался как система управления содержанием (content management system), предоставляющая единую структуру управления и отображения всей базы знаний процессов организации. Все содержание баз знаний процессов может быть опубликовано в виде html-файлов и выложено на веб-сервера для распределенного использования.

RMC дает возможность управления процессной методологией руководителю проекта посредством выбора, адаптации и быстрой сборки Процессов для конкретного проекта разработки. RMC предоставляет предопределенный набор Процессов для типичных проектных ситуаций, которые могут быть адаптированы под конкретные нужды. Он также предоставляет блоки построения процессов разработки, называемые Процессы Ключевых Областей разработки (capability patterns), которые представляют собой лучшие практики разработки для конкретных дисциплин, технологий или стилей управления проектами. Эти строительные блоки формируют инструментарий для быстрого создания Процесса разработки, основываясь на специфических нуждах проекта. RMC также позволяет настраивать специфические для организации Процессы Ключевых Областей разработки. Наконец, Процессы разработки, созданные с помощью RMC, могут быть опубликованы в виде веб-сайтов.

Выделение статических элементов

В работе на первом этапе построения модели «структура-свойство» предлагается построить бизнес use-case диаграмму, на которой отразить основные роли и бизнес use-cases, а также сделать первичную декомпозицию (разбиение на дисциплины) нашего бизнес-процесса, проводя аналогию с дисциплинами в RUP. Этот шаг позволит структурировать процесс, проследить связи между отдельными частями процесса, более четко организовать дальнейшую работу по моделированию. Далее выделяются:

- роли (Business Worker) и группы ролей;
- артефакты или work products (реальные входные/выходные элементы для процесса, которые могут быть представлены как документами, так и настройкой конфигурации в инструменте);
- задания (tasks) (мы будем следовать подходу, близкому к анализу вариантов использования в объектно-ориентированном подходе, то есть выявлять задания, выполняемые ролями, и их оценивать)

Затем определяются взаимоотношения между выделенными выше элементами. Сначала, как основной элемент, используются роли и описываются все взаимосвязи по отношению к ним. Далее можно оценить связи ролей, артефактов между собой и т.д.

Выделение динамического процесса

Теперь приступаем к созданию диаграмм (breakdown structure), которые определяют последовательность, синхронность активностей, а также всевозможные зависимости. Иногда, этот шаг может быть выполнен при выделении заданий, но, как мы увидим при работе с новым инструментом IBM Rational Method Composer (RMC), такая последовательность действий себя оправдывает.

Не следует также забывать про элементы, которые облегчают адаптацию процесса, его использование и управление (guidelines, checklists, tool mentors и т.д.).

Генерация сайта

Наконец, последний этап: конкретное воплощение результата (генерация сайта наподобие RUP). Таким образом, мы получаем подход, близкий по идее к метамодели, лежащей в основе RUP – это метамодель SPEM (Software Process Engineering Metamodel). Более наглядно метамодель SPEM для RUP представлена на следующем рисунке.

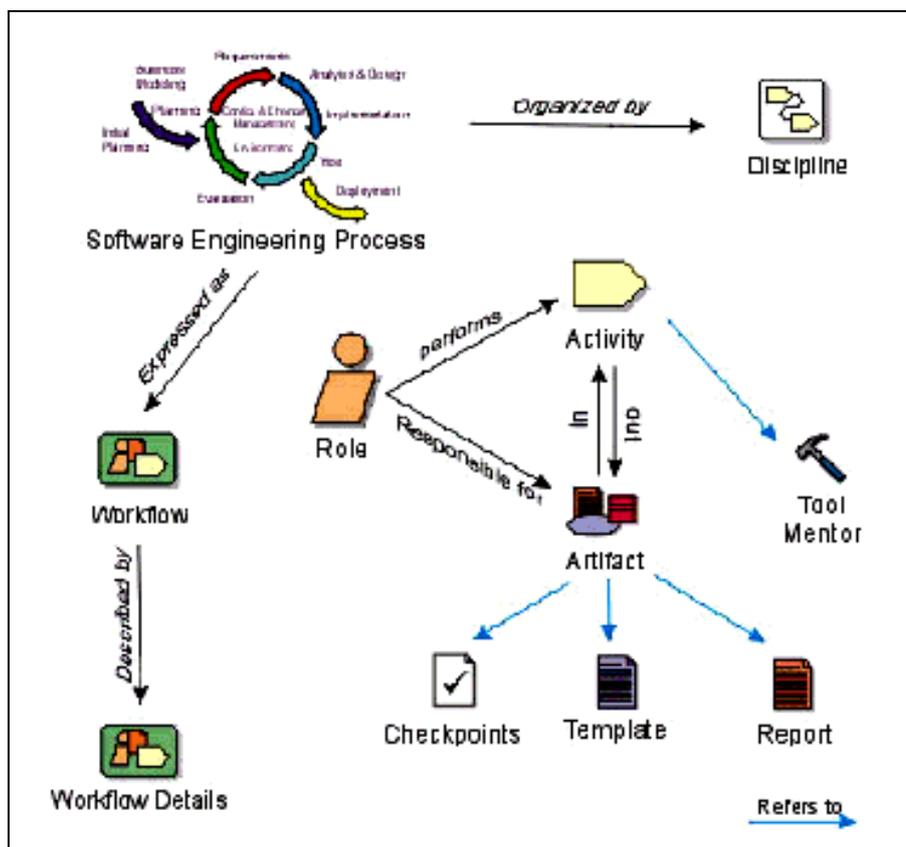


Рисунок 2.1. Компоненты модели (основа: метамодель SPEM).

3. Описание процесса решения QSAR-задачи

Существует несколько подходов к бизнес-моделированию, описанных в предыдущей главе данной работы. А также предлагается подход к решению задачи построения модели бизнес-процессов, поддерживаемую инструментальными средствами, позволяющими легко ее модифицировать при изменении (улучшении) бизнес-процессов. П

Выделение элементов построения модели «структура-свойство»

Прежде чем приступить к реализации цели данной дипломной работы, необходимо построить на UML объектно-ориентированную модель бизнес-процессов решения задачи «структура-свойство» таким образом, чтобы можно было сгенерировать сайт с регламентом работ, определить требования к информационной системе, автоматизирующей бизнес-процесс, легко модифицировать с целью улучшения процесса.

Выделены все статические элементы процесса решения задачи «структура-свойство» (Роли, Задачи, Артефакты) и далее в таблице представлена их взаимосвязь.

Роли процесса:

- **Заказчик** - Ответственный за определение цели, выделение ресурсов и постановку задачи;

- **Математик** - Ответственный за построение и верификацию модели;

- **Библиотекарь** - Ответственный за содержимое и состав ОВМ, а также за предварительный анализ ОВМ при вводе ее в репозиторий

- **Тестировщик** - Ответственный за проверку качества моделей, включая прогнозирование на тестовых выборках;

- **Химик** - Ответственный за расчет конформаций молекул ОВМ с помощью специальных программ, включая программы квантово-химических расчетов.

Описание артефактов:

- **ОВМ** – обучающая выборка молекул, служит для построения модели «структура-свойство» (модели «структура-свойство»);

- **ТВМ** – тестовая выборка молекул, используемая для тестирования модели;

- **Модель** (модель «структура-свойство», QSAR-модель) – функциональное представление зависимости «структура-свойство» на дескрипторах молекул из ОВМ;

- **Параметры качества модели** – числовые значения, агрегирующие значения ошибки модели и характеризующие качество аппроксимации целевого свойства ОВМ для заданного типа зависимости. Определяются функцией качества, заданной для данного типа зависимости «структура-свойство» (Например, коэффициент множественной корреляции для линейной регрессионной зависимости);

- **ОВМ-ТП** – обучающая выборка молекул в топологическом представлении (т.е. координаты атомов не заданы, или имеют условное значение с целью визуального представления рисунка);

- **ОВМ-ЗД** – обучающая выборка молекул в геометрическом представлении (т.е. заданы пространственные координаты атомов молекул ОВМ);

- **ТВМ-ТП** – тестовая выборка молекул в топологическом представлении;

- **ТВМ-ЗД** – тестовая выборка молекул в геометрическом представлении;

- **Конформация** (конформация молекулы) – пространственное представление молекулы из ОВМ, молекула может быть представлена в ОВМ-ЗД как в одной, так и несколькими конформациями;

- **Профиль** – двоичный вектор, задающий выбор молекул из ОВМ;

- **К-профиль** (или Профиль кластера) – профиль задающий кластер ОВМ;

- **МД-матрица** – матрица «молекула-дескриптор», где строки соответствуют молекулам, а столбцы – дескрипторам молекул, как топологическим, так и геометрическим. Состоит из матрицы, списка имен столбцов и списка имен молекул из ОВМ;

- **Кластерная структура** (К-структура ОВМ) – ОВМ, разбитая на кластеры, которые представлены списком К-профилей. Для К-структуры известна метрика, определяющая расстояния между молекулами.

| | <i>Задача</i> | <i>Вход</i> | <i>Выход</i> | <i>Роль</i> |
|--|---|---|--|--------------------|
| | «Постановка задачи» | Заказ на работу | Цели построения моделей, Определение доступных ресурсов для моделирования | Заказчик |
| | «Предоставление данных» | Заказ на работу | Литературные данные по ОВМ и ТВМ | Заказчик |
| | | | | |
| | «Расчет пространственного представления молекул» | ОВМ-ТП, Параметры расчета конформаций | ОВМ-3Д | Математик |
| | «Оценка "кластерной структуры" обучающей выборки» | МД-матрица, Метрика | Кластерная структура ОВМ, Профиль, задающий «используемые» молекулы в ОВМ | Математик |
| | «Определение "области действия" модели» | Обучающая выборка молекул, К-профили | К-структура ОВМ | Математик |
| | «Построение модели на кластере» | Кластерная структура ОВМ, К-профиль, Тип модели | Модель, Параметры качества | Математик |
| | «Построение кластеров» | Профиль, МД-матрица | К-структура, т.е. К-профиль, Центр кластера, Условие принадлежности к кластеру для каждого найденного кластера | Математик |
| | Подготовка отчётов по результатам расчётов» | Модель | Отчет | Математик |
| | | | | |
| | «Ввод структурных молекул ОВМ» | Литературные источники | ОВМ-ТП, статус «введено» | Библиотекарь |
| | «Проверка правильности данных о свойствах» | Литературные источники | ОВМ-ТП, статус «проверено» | Библиотекарь |
| | Проведение первичного расчета топологических дескрипторов | ОВМ-ТП, список топологических дескрипторов и топологических индексов (ТИ) | МД-матрица | Библиотекарь |
| | «Формирование стандартных пространственных представлений молекул» | ОВМ-ТП, статус «проверено», параметры расчета конформаций | ОВМ-3Д | Библиотекарь |
| | «Формирование пространственных поверхностей» | ОВМ-3Д, параметры расчета поверхностей | 3Д-поверхности молекул | Библиотекарь |

| <i>Задача</i> | <i>Вход</i> | <i>Выход</i> | <i>Роль</i> |
|---|---|--|-------------|
| <i>молекул»</i> | | | |
| «Расчёт устойчивых построенных конформаций» | ОВМ-ТП, статус «проверено», параметры расчета конформаций | ОВМ-3Д | Химик |
| «Настройка проверки» | Определение целей тестирования | Параметры проверки модели | Тестировщик |
| «Проведение проверки модели» | Модель, параметры проверки модели | Параметры качества модели, сводные результаты тестирования | Тестировщик |

Динамика процесса решения QSAR задачи

Для наших задач в первую очередь наиболее важными являются диаграммы вариантов использования (Business Use-Case диаграммы) и диаграммы активностей UML. Диаграмма активностей отражает последовательность видов деятельности, связанных с определенным объектом, включая переходы, точки ветвления, объединения, распараллеливание и слияние. Этот вид диаграмм часто бывает необходим для отображения динамического аспекта моделируемого процесса.

Диаграммы сценариев использования

На первом этапе необходимо решить, как сделать декомпозицию организации бизнес-процессов решения задачи «структура-свойство» так, чтобы можно было удобно организовать предстоящую работу и управлять отношениями подчинения. Здесь же можно построить Business Use-Case диаграмму, чтобы отобразить Business Actors (это некоторые роли, внешние по отношению к бизнес-процессам организации) и Business Use-Cases (вариант использования с точки зрения бизнес-процессов), которые определяются как описания последовательностей действий (потоков событий) в рамках некоторого бизнес-процесса, приносящих ощутимый результат конкретному действующему лицу.

На следующем рисунке представлена Business Use-Case диаграмма для бизнес-процесса функционирования работы QSAR-коллектива, выполненная с помощью IBM Rational Rose:

Диаграммы активностей

Для представления динамического аспекта работы над моделируемым процессом, построения самого процесса и его конфигурации удобно воспользоваться возможностью используемого инструментального средства определять work breakdown structure, или получать диаграммы активностей (workflows). Диаграмма активностей отражает последовательность видов деятельности, связанных с определенным объектом.

Сначала была выявлена общая последовательность выполняемых Activity для решения задачи «структура-свойство»:

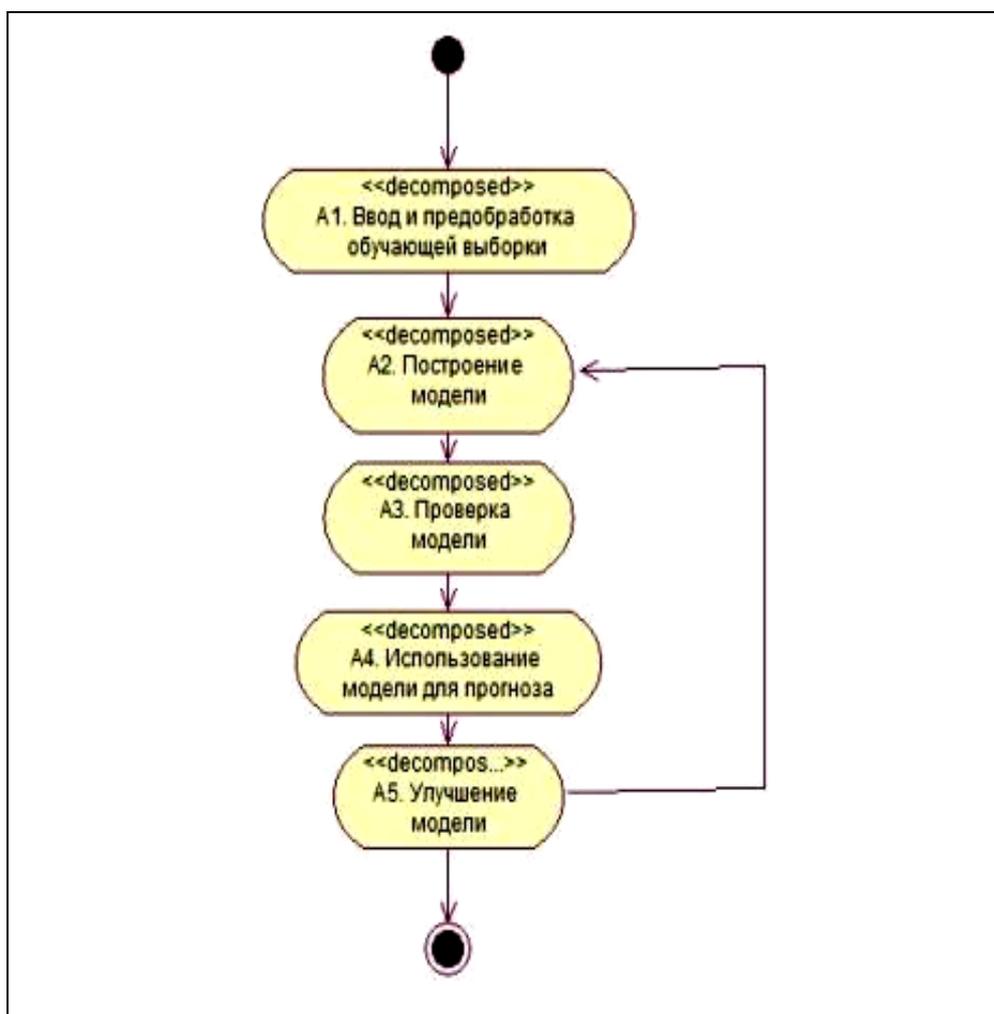


Рис.3.2. Диаграмма активностей для процесса жизненного цикла QSAR-моделей
Каждая задача (Activity) раскрыта ниже в виде последовательности подзадач:

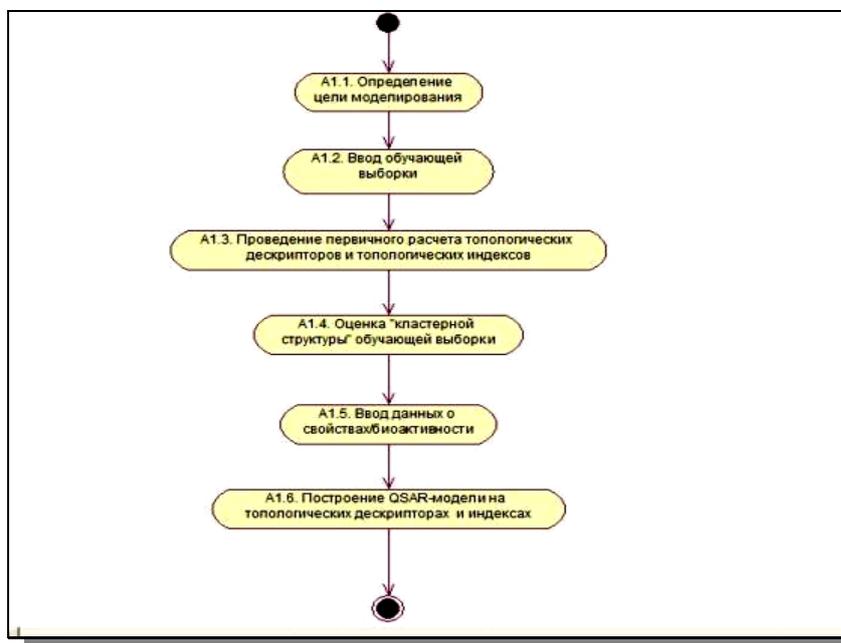


Рис.3.3. Ввод и предобработка обучающей выборки

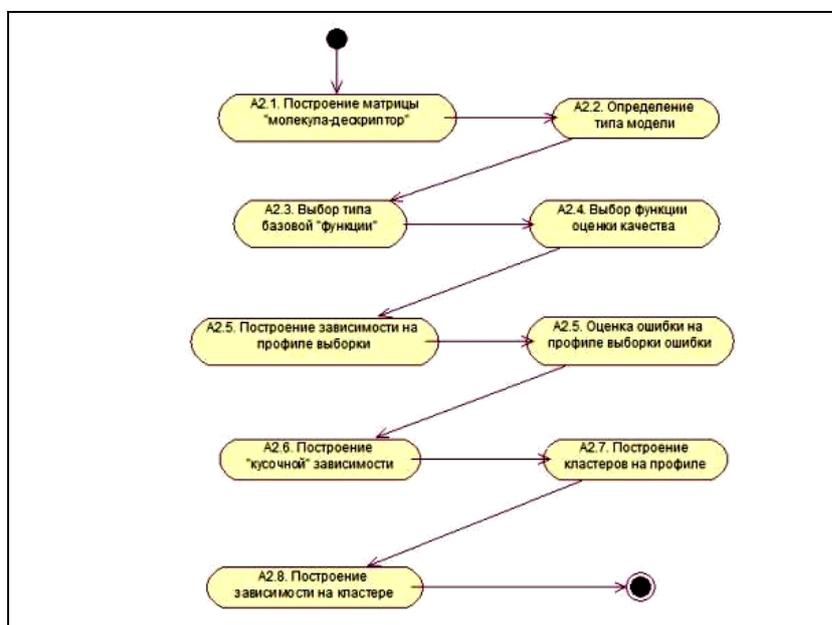


Рис.3.4. Построение модели

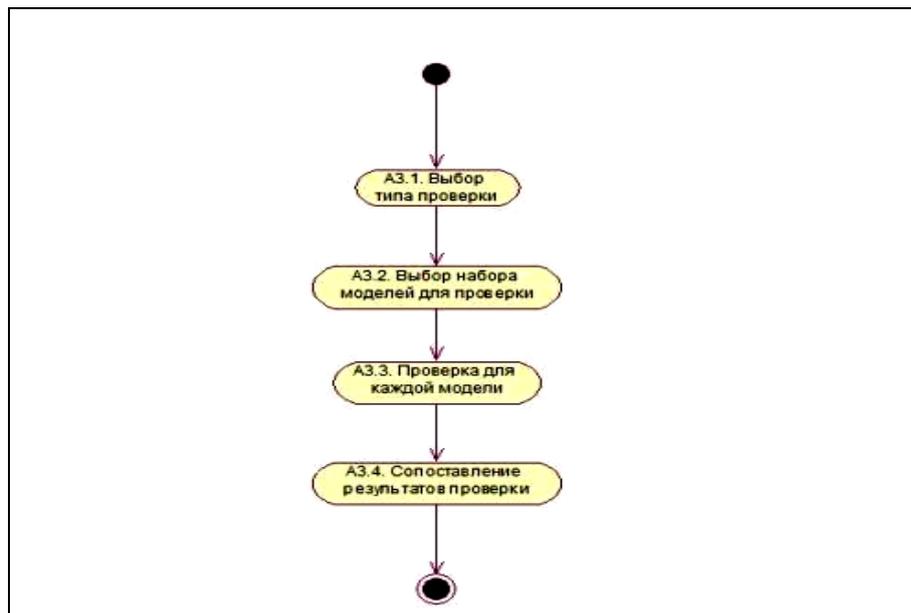


Рис.3.5. Проверка модели

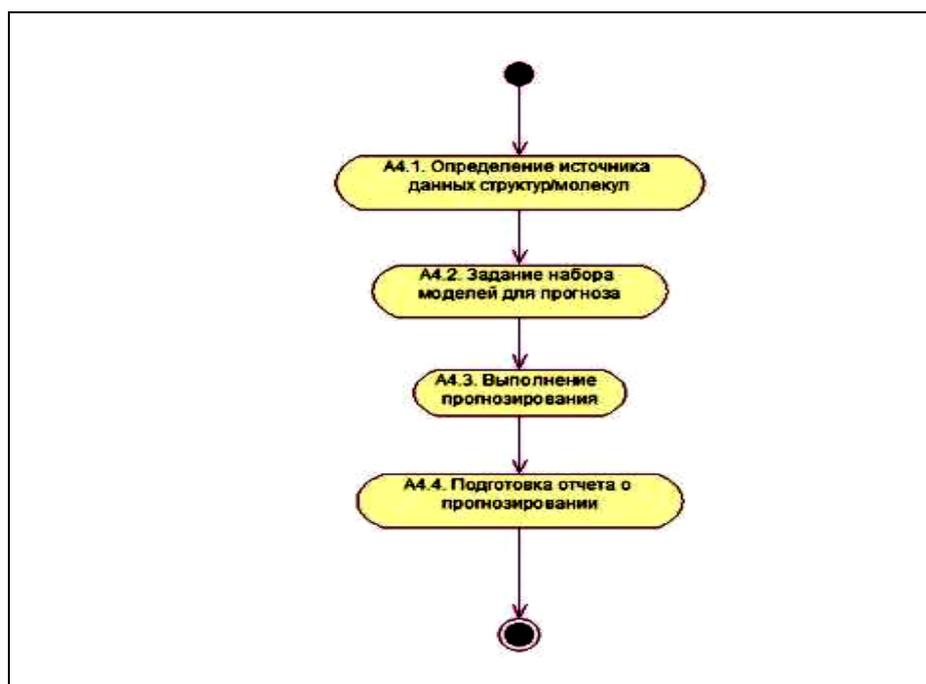


Рис.3.6. Использование модели для прогноза

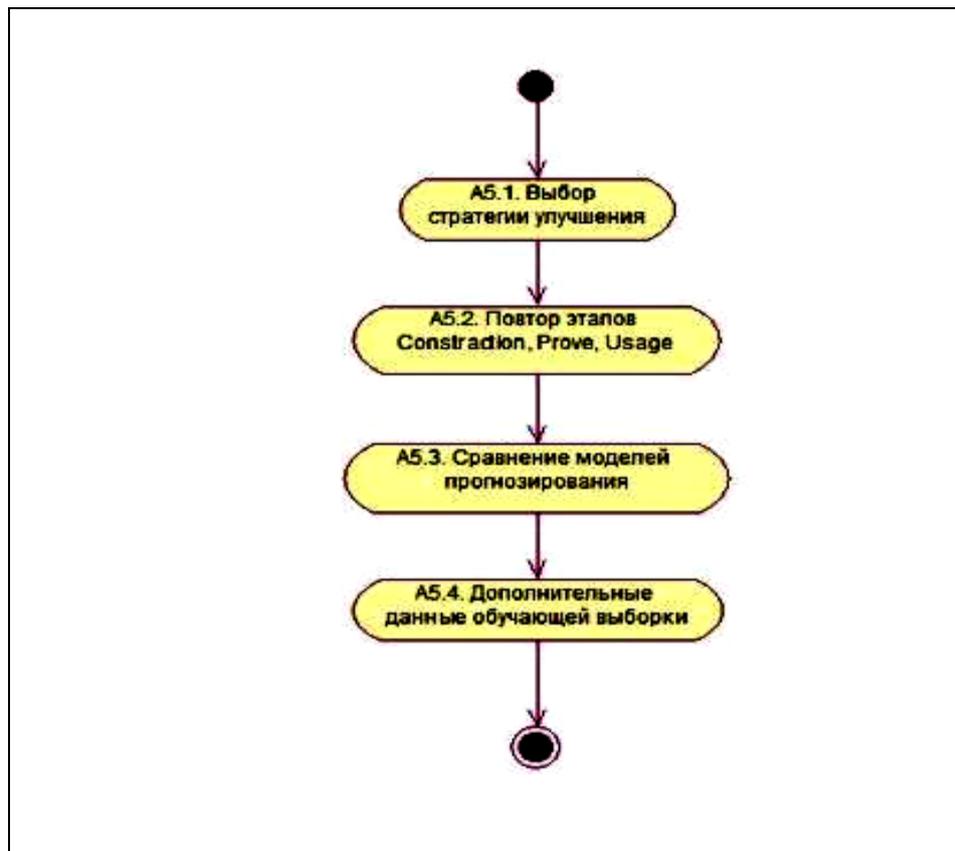


Рис.3.7. Улучшение модели

Моделирование процесса в среде RMC

IBM Rational Method Composer представляет собой платформу с гибкими процессами, содержащую средства для управления жизненным циклом информационных технологий, которая призвана обеспечить настраиваемую, но единообразную стратегию для процессов, осуществляемых проектными группами. Это инструмент настройки и публикации веб-сайтов на основе RUP. RMC предназначен для тех, кому необходимо внести значительные изменения в RUP, сделав их основой для дальнейшего использования в проекте.

Для создания процесса разработки требуется понимание способа применения выполняемых методов в течении всего цикла разработки. Необходимо также четко осознавать, как различные задания, входящие в методы, связаны друг с другом.

Назначение RMC

1) Создание базы знаний процессов, которую можно просматривать, управлять и публиковать. Содержание этой базы знаний может включать как внешне разработанные методы, так и собственные, включающие руководства, шаблоны, принципы, лучшие практики, внутренние процедуры и упорядочивание, тренинговые материалы или любые другие описания собственных методов. Эта база может быть использована для обучения, либо как справочная информация. RMC предназначен быть системой управления контентом, которая предоставляет общую структуру и вид для всего контента, это не система управления документами, которая является труднодоступным хранилищем документов разных видов и форматов. Любой управляемый в RMC контент может быть опубликован в виде html-страничек и использован на различных компьютерах

2) Предоставление возможности создания индивидуального процесса путем выбора, приспособления и установления процессов для конкретных проектов разработки. RMC содержит каталоги предопределенных процессов для типичных ситуаций (эти процессы могут быть приспособлены под индивидуальные требования). Он также дает в распоряжение структурные блоки, называемые *capability patterns*, которые являются лучшими практиками по разработке специфических дисциплин, технологий или стилей управления. Эти блоки образуют инструментарий для быстрой сборки процессов, основанных на особых требованиях проекта. RMC также позволяет настроить собственные библиотеки, содержащие *capability patterns*. Наконец, процессы, созданные с помощью RMC могут быть опубликованы в виде веб-сайтов.

RMC обеспечивает интуитивно понятными текстовыми редакторами для создания иллюстрированных описаний контента. Эти редакторы разрешают использование стилей, изображений, таблиц, гиперссылок, а также непосредственное html-редактирование. RMC также позволяет создавать диаграммы потоков работ (*workflows*). Не зависит от RUP и позволяет генерировать собственные модели.

Для того, чтобы эффективно работать с RMC, необходимо понимать несколько основных концепций, которые используются для организации контента. Самый фундаментальный принцип в RMC – это разделение *method content* и этапов процесса. Почти все концепции RMC разбиваются на категории в виду этого разделения. *Method content* описывает то, что должно быть произведено, необходимые навыки, требуемые для этого и пошаговое объяснение того, как достичь поставленных целей. Это описание не связано с циклом разработки. Процессы, в свою очередь, описывают цикл разработки, используя элементы *method content*, создавая из них полуупорядоченные последовательности, которые подстраиваются под специфические нужды проекта.

RMC является лучшим средством отображения процессов, в котором содержатся описания всех ролей и видов деятельности. Полученный сайт также служит для централизованного хранения всех шаблонов рабочих документов.

Преимущество этого подхода к анализу, проектированию и реализации решения состоит в том, что он обеспечивает полный цикл разработки модели – от формулировки требований до программной реализации.

С учетом всех требований на решение, создан и размещен в сети веб-проект QSAR-коллектива:

The screenshot shows the QSAR Russia website. At the top, there is a navigation menu with tabs for 'Главная', 'Контакты', 'Новости', and 'Ссылки'. Below this is a search bar and a section for 'КРАТКИЕ НОВОСТИ' (Brief News) with a message about the website's development. A 'ОПРОСЫ' (Polls) section asks 'Как Вам наш сайт?' (How do you like our site?) with options from 'Супер!' to 'Ужасно'. The main content area is titled 'Участники' (Participants) and lists several faculty members with their titles and affiliations. On the left, there is a 'ГЛАВНОЕ МЕНЮ' (Main Menu) with links to 'Главная', 'Новости', 'Проекты', 'Поиск', 'Ссылки', 'Контакты', 'Форум', 'Download', and 'Участники'. Below that is a 'МЕНЮ ПОЛЬЗОВАТЕЛЯ' (User Menu) with options like 'Ваши данные', 'Добавить новость', 'Добавить ссылку', 'Проверить материалы', and 'Выход'. There is also an 'АВТОРИЗАЦИЯ' (Authorization) section with a login form for 'Здравствуйте, Сенкова Татьяна' and a 'Выйти' (Logout) button. At the bottom left, there are 'ЭКСПОРТ НОВОСТЕЙ' (Export News) options for RSS, ATOM, and OPML.

<http://qsar-msu.ru/>

Заключение

В работе решена задача генерации рабочих пошаговых инструкций в проектах поиска зависимости «структура-свойство». Для генерации был использован инструмент IBM Rational Method Composer (версия 7.2). Подробно описан процесс решения задачи «структура-свойство». Выделены все статические элементы: Роли, Задачи, Артефакты. На их основе прописаны процессы, отражающие последовательность построения моделей «структура-свойство» таким, как он должен быть.

Предложены настройки под текущий проект и на их основе сгенерирован учебник в виде веб-сайта. Под каждый новый проект появилась возможность генерировать свою последовательность шагов для исполнителей разных квалификаций с подробным описанием, кто за что и в какой момент отвечает и что получает на входе и на выходе в результате своей деятельности.

Также для обеспечения эффективных коммуникаций исполнителей проекта создан специальный веб-сайт, который будет содержать основную документацию по проекту, текущие

планы работ, включая состояния работы и её исполнителей, запросы на изменения и их состояния, основные публикации по проекту.

Таким образом, появляется возможность сравнивать текущее состояние проекта с тем, каким оно должно быть, и обнаружить возможные проблемы сразу по мере их возникновения (соответственно, вовремя принять решение по их устранению).

Можно выделить следующие основные направления развития данной работы:

- 1) сопровождение полученной пошаговой инструкции подробным техническим описанием;
- 2) расширение контента созданного веб-сайта (например, добавление каталога молекулярных выборок, включая поддерживаемые представления по каждой выборке и перекрестные ссылки, имеющиеся QSAR/QSPR-модели).

Список литературы

1. Журавлев Ю.И., Рязанов В.В., Сенько О.В. *«Распознавание». Математические методы. Программная система. Практические применения.* - М.: ФАЗИС, 2006.
2. Скворцова М.И., Станкевич И.В., Палюлин В.А., Зефирова Н.С. *Концепция молекулярного подобия и ее использование для прогнозирования свойств химических соединений.* - *Успехи химии, 2006, Том 75, ном.11, стр. 1074-1093.*
3. Журавлев Ю. И. *Об алгебраическом подходе к решению задач распознавания и классификации.* - М.: Наука, 1978, вып. 33.
4. Рязанов В.В. *О построении оптимальных алгоритмов распознавания и таксономии (классификации) при решении прикладных задач.* – В кн.: Распознавание, классификация, прогноз: Математические методы и их применение – М.:Наука, 1998, вып. 1.
5. Скотт К. *«UML. Основные концепции»*, М.: Вильямс, 2002.
6. Крачтен Ф. *«Введение в Rational Unified Process»*, М.: Вильямс, 2002.