

На правах рукописи

СЕМИКИН Виктор Алексеевич

**СЕМАНТИЧЕСКАЯ МОДЕЛЬ КОНТЕНТА
ОБРАЗОВАТЕЛЬНЫХ ЭЛЕКТРОННЫХ ИЗДАНИЙ**

**05.13.18 — математическое моделирование, численные
методы и комплексы программ**

**Автореферат диссертации на соискание ученой
степени кандидата технических наук**

Тюмень 2004

Работа выполнена на кафедре программного обеспечения Тюменского государственного университета

Научный руководитель кандидат физико-математических наук,
доцент **Деревнина Анна Юрьевна**

**Официальные
оппоненты:** доктор технических наук, доцент
Захаров Александр Анатольевич,
кандидат физико-математических наук,
доцент **Захаров Сергей Дмитриевич**

Ведущая организация **Томский государственный университет**

Защита диссертации состоится 4 июня 2004 г. в 15 часов на заседании диссертационного совета К 212.274.01 при Тюменском государственном университете по адресу: 625003, г. Тюмень, ул. Перекопская, 15а, аудитория 217.

С диссертацией можно ознакомиться в библиотеке Тюменского государственного университета.

Автореферат разослан 26 апреля 2004 года.

Ученый секретарь
диссертационного совета

Бутакова Н.Н.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Широкое распространение образовательных электронных изданий (ОЭИ) и, в частности, наполнение образовательных интернет порталов, существенно повышает требования к функциональным возможностям и, следовательно, к технологиям разработки ОЭИ. Для создания ОЭИ, как правило, применяется технология «контент+браузер», предполагающая явное отделение содержания (контента) от программ его визуализации (браузера) и редактирования. Таким образом, решающее влияние на функциональные возможности создаваемых ОЭИ оказывает модель данных, используемая для представления контента, и функциональные возможности браузера.

В практике создания ОЭИ наибольшее распространение получила гипертекстовая модель (В. Буш, Д. Энгельбарт, Т. Нельсон, Дж. Конклин, Н.В. Агеев, Л.Г. Ованесбеков). Данная модель позволяет успешно решать задачи навигации и визуализации контента. Однако, слабая структурированность данных, представленных в гипертекстовой модели (в частности, отсутствие типизации данных), существенно ограничивает возможности их программной обработки, что в свою очередь отрицательно сказывается на функциональных возможностях ОЭИ. Для преодоления указанного недостатка необходима более сложная модель.

Данная проблема частично решена в семантических моделях, к которым относятся семантические сети, фреймовая модель, модель сущность-связь и др. Несмотря на то, что некоторые из этих моделей и, в первую очередь, семантические сети были использованы в ряде успешных проектов (Дж. Карбонелло, П.Л. Брусиловский, М.И. Зырянов), они не получили широкого распространения в практике создания ОЭИ, так как не поддерживают визуализацию информации.

Кроме того, иногда применяется реляционная модель, предложенная Е.Ф. Кодом. Однако, контент, как правило, имеет сетевую структуру, плохо соответствующую данной модели. Обычно реляционная модель используется как базовая для хранения гипертекстовых документов.

Таким образом, в настоящее время не существует модели данных, позволяющей адекватно представлять семантическую структуру контента и поддерживающей его визуализацию. Разработка такой модели имеет важное значение для совершенствования технологий создания ОЭИ.

Цель работы состоит в совершенствовании технологий создания образовательных электронных изданий путем разработки модели данных для представления контента, позволяющей адекватно отражать семантическую структуру дидактической информации и поддерживающей визуализацию контента в форме, ориентированной на его изучение.

Для достижения поставленной цели в работе решаются следующие **задачи**:

- исследование принципов и технологий создания ОЭИ;
- построение модели представления контента ОЭИ, позволяющей отражать семантическую структуру учебной информации;
- разработка методов и алгоритмов количественной оценки структурных характеристик контента;
- создание технологии разработки образовательных электронных изданий, позволяющей учитывать семантику контента при реализации основных функций ОЭИ;
- разработка инструментального программного комплекса для создания ОЭИ с широким спектром функциональных возможностей;
- апробация предложенных моделей и технологий в процессе создания ОЭИ.

Методы исследования. При построении и исследовании модели контента использовались методы теории графов, общей алгебры, теории алгоритмов, теории множеств, теории формальных грамматик. Для разработки технологии и инструментального средства для создания ОЭИ применялись методы объектно-ориентированного анализа и программирования.

Научная новизна исследования заключается в разработке семантической модели представления контента ОЭИ. Для данной модели впервые предложены методы и алгоритмы количественной оценки структурных характеристик контента. Предлагаемые характеристики позволяют исследовать соответствие структуры контента семантике дидактических материалов и оценить возможности его программной обработки.

Научно-практическая значимость работы состоит в разработке на основе предложенной модели технологии создания ОЭИ, позволяющей учитывать семантику контента при реализации основных функций ОЭИ. На базе данной технологии разработан инструментальный программный комплекс, поддерживающий создание сетевых, локальных и печатных версий ОЭИ.

Апробация работы:

- Основное содержание диссертации опубликовано в 11 печатных работах.
- Материалы диссертации докладывались на следующих конференциях и семинарах: Всероссийский семинар «Электронные учебники и учебно-методические разработки в открытом образовании» (Москва, 2000); Всероссийская школа-семинар «Информационные технологии в управлении качеством образования и развитии образовательного пространства» (Москва, 2000); Всероссийская научно-методическая конференция «Телематика'2002» (Санкт-Петербург, 2002); Всероссийская научно-методическая конференция «Качество высшего профессионального образования в начале XXI века» (Туапсе, 2002); Всероссийская научно-методическая конференция «Телематика'2003» (Санкт-Петербург, 2003).
- Работа выполнена при поддержке гранта №2.7.5.(43.1).228.059 программы Минобразования РФ «Научное, научно-методическое, материально-техническое и информационное обеспечение системы образования», 2001 - 2002 гг.
- С использованием инструментального программного комплекса созданы ОЭИ для федерального комплекта электронных средств поддержки обучения по общим гуманитарным и социально-экономическим дисциплинам.
- Разработанные ОЭИ изданы на компакт-дисках промышленным тиражом 16500 экземпляров.
- ОЭИ зарегистрированы в Отраслевом фонде алгоритмов и программ Минобразования РФ, часть из них имеет гриф «Рекомендовано Минобразования РФ в качестве учебного пособия для использования в учебном процессе высших учебных заведений».
- Разработанные ОЭИ внедрены в учебный процесс ряда высших учебных заведений Российской Федерации.

На защиту выносятся:

- Модель представления контента ОЭИ, позволяющая отражать семантическую структуру учебной информации и поддерживающая визуализацию контента в форме, ориентированной на его изучение.
- Методы и алгоритмы количественной оценки структурных характеристик контента. Предлагаемые характеристики позволяют исследовать соответствие структуры контента семантике дидактических материалов и оценить возможности его программной обработки.

- Технология разработки образовательных электронных изданий, позволяющая учитывать семантику контента при реализации основных функций ОЭИ.
- Инструментальный программный комплекс, поддерживающий создание сетевых, локальных и печатных версий ОЭИ.

Структура и объем работы. Диссертация состоит из введения, трех глав, заключения и списка литературы, включающего 103 наименования. Общий объем работы составляет 122 страницы.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационной работы, сформулированы цели и задачи исследования, показаны научная новизна и практическая значимость работы.

Первая глава «Принципы и технологии создания ОЭИ» носит обзорный характер. В главе рассмотрены различные типы ОЭИ, приведена их классификация, используемая в отечественной и зарубежной литературе. Исследованы основные функции ОЭИ, а также предъявляемые к ним научно-методические и программно-технические требования.

Рассмотрены различные технологии создания ОЭИ, наибольшее внимание уделяется технологии «контент+браузер», при которой программа, реализующая функции ОЭИ (браузер), отделена как от контента, так и от программ создания и редактирования контента. Проведен анализ основных типов инструментальных программных средств, применяемых при создании ОЭИ, раскрываются технологические особенности и возможные области применения каждого типа средств.

Особое внимание уделяется исследованию моделей данных, используемых для представления контента ОЭИ. Приводится характеристика классических моделей, кратко описаны иерархическая, сетевая и реляционная модели. Подробно рассмотрена гипертекстовая модель, как получившая наибольшее распространение в практике создания ОЭИ. Кроме того, описаны семантические модели данных, среди которых наиболее подробно освещены семантические сети. Проведен сравнительный анализ различных моделей данных с точки зрения целесообразности их использования для представления контента ОЭИ.

На основании проведенного анализа используемых технологий создания ОЭИ и моделей данных, применяемых для представления контента, сформулированы основные задачи диссертационной работы.

Вторая глава «Модель контента образовательных электронных изданий» посвящена описанию предлагаемой семантической модели и исследованию численных характеристик структуры контента.

Описание модели. Семантическая модель использует следующие принципы структурирования контента:

1. Разбиение контента на множество структурных элементов $E = \{e_1, \dots, e_n\}$.
2. Типизация структурных элементов $G : E \rightarrow T$, где $T = SUCUL$ — множество типов. S — множество типов семантических элементов, соответствующих дидактическим единицам контента, например «лекция», «определение», «теорема», «термин»; C — множество типов мультимедиа элементов, например «текст», «графика», «анимация», «видео»; L — множество типов элементов семантических связей, используемых для установления логических зависимостей, например «следует из», «определяется», «иллюстрирует».
3. Иерархическое упорядочение контента, отражающее отношение «целое-часть» между структурными элементами. Иерархия определяется отображением $F : E \rightarrow 2^E$, ставящим в соответствие каждому структурному элементу e_i множество его дочерних элементов (рис. 1).
4. Задание семантических связей, отражающих логические зависимости между структурными элементами. Семантические связи определяются отображением $H : \{e : G(e) \in L\} \rightarrow \{e : G(e) \in S\}$ множества элементов семантической связи во множество семантических элементов.

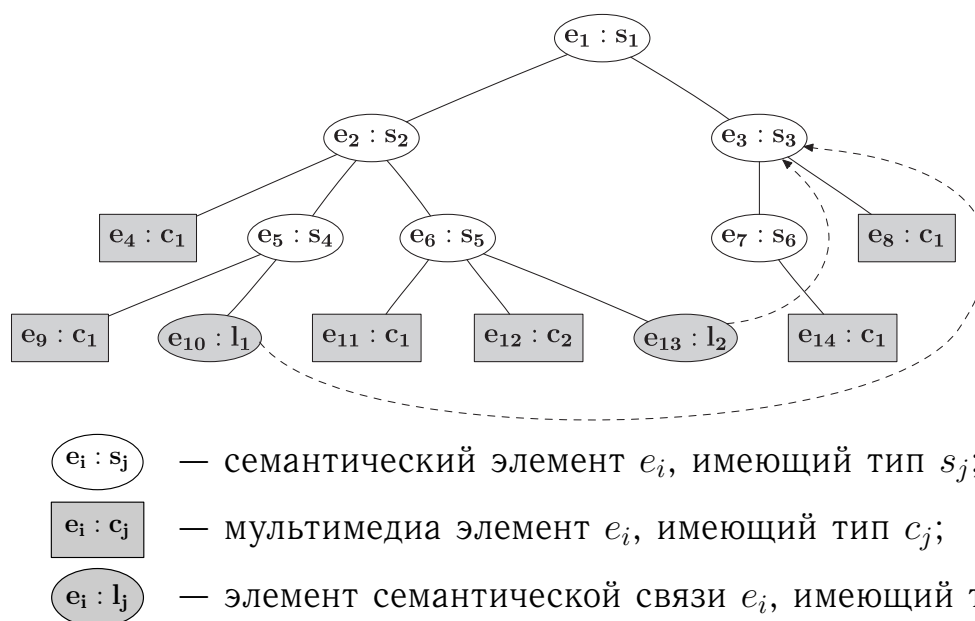


Рис. 1. Пример структуры контента

Спецификация структуры контента. При создании ОЭИ такие факторы как предметная область, целевая аудитория и многие другие накладывают определенные ограничения на типы используемых элементов и возможные связи между ними. Для контроля корректности структуры контента необходим механизм формальной спецификации таких ограничений. Предлагаемый метод позволяет задавать ограничения трех видов.

1. Определение множества типов структурных элементов:

- $S = \{s_1, s_2, \dots, s_{n_1}\}$ — множество используемых типов семантических элементов;
- $L = \{l_1, l_2, \dots, l_{n_2}\}$ — множество используемых типов элементов семантической связи;
- $C = \{c_1, c_2, \dots, c_{n_3}\}$ — множество используемых типов мультимедиа элементов.

2. Спецификация иерархической структуры контента, состоящая из выражений вида $s \rightarrow exp$, где exp — шаблон, определяющий типы дочерних элементов для s и содержащий следующие выражения:

- t — обязательный элемент типа t ;
- $t?$ — необязательный элемент типа t ;
- t^* — любое количество элементов типа t ;
- $t_1|t_2$ — элемент типа t_1 или t_2 ;
- t_1, t_2 — элемент типа t_1 , за которым следует элемент типа t_2 ;
- (\dots) — группа элементов.

3. Спецификация семантических связей, определяющая для каждого типа связи множество типов элементов, на которые он может ссылаться и состоящая из выражений: $l_i \rightarrow \{s_{j_1}, \dots, s_{j_k}\}$.

Несмотря на то, что данный метод спецификации структуры контента позволяет задавать большинство практически важных ограничений, он не является универсальным. Более того, в работе доказана следующая теорема.

Теорема 1. *Не существует нотации, позволяющей задать любую спецификацию структуры контента и для всякой структуры эффективно (с помощью некоторого алгоритма) проверять ее соответствие спецификации.*

Характеристики структуры контента. В работе предлагается ряд численных характеристик, предназначенных для оценки особенностей семантической структуры контента. Основными требованиями к представлению контента являются адекватное отражение семантики дидактических материалов и предоставление широких возможностей

по обработке контента. Предлагаемые характеристики нацелены на оценку соответствия структуры контента данным требованиям. Исходя из используемых в модели механизмов структурирования выделяются три группы характеристик: *типизации, семантической связи, иерархической структуры*.

Определение 1. Пусть $A \subseteq E$. Для получения характеристик структуры контента будем использовать следующие операторы:

- $TP_{t_i}(A) = \{e : e \in A \ \& \ G(e) = t_i\}$ — множество элементов типа t_i ;
- $PR(A) = \{e : F(e) \cap A \neq \emptyset\}$ — множество родительских элементов;
- $CH(A) = \{e : e \in F(a) \text{ для некоторого } a \in A\}$ — множество дочерних элементов;
- $LT(A) = \{e : H(a) = e \text{ для } a \in A \text{ такого, что } G(a) \in L\}$ — множество элементов, на которые ссылаются элементы семантической связи из A ;
- $LS(A) = \{e : G(e) \in L \ \& \ H(e) = a \text{ для некоторого } a \in A\}$ — множество элементов семантической связи, ссылающихся на элементы из A ;

Очевидно, что перечисленные операции вместе со стандартными множественными операциями (\cup , \cap , \setminus) задают алгебру на 2^E — семействе всех подмножеств структурных элементов.

Характеристики типизации. Основной характеристикой структуры контента является *количество структурных элементов*. Как правило, чем больше количество структурных элементов, используемых для представления данного контента, тем шире возможности его программной обработки.

Следующие **характеристики типизации** позволяют оценить распределение структурных элементов по типам и классам:

- $|TP_{t_i}(E)|$ — количество элементов типа t_i ;
- $|\bigcup_{t_i \in S} TP_{t_i}(E)|$ — количество семантических элементов;
- $|\bigcup_{t_i \in C} TP_{t_i}(E)|$ — количество мультимедиа элементов;
- $|\bigcup_{t_i \in L} TP_{t_i}(E)|$ — количество элементов семантической связи.

Характеристики семантической связи. Для исследования структуры семантических связей определим отношение *ссылается* на множестве семантических элементов. Если семантический элемент s содержит

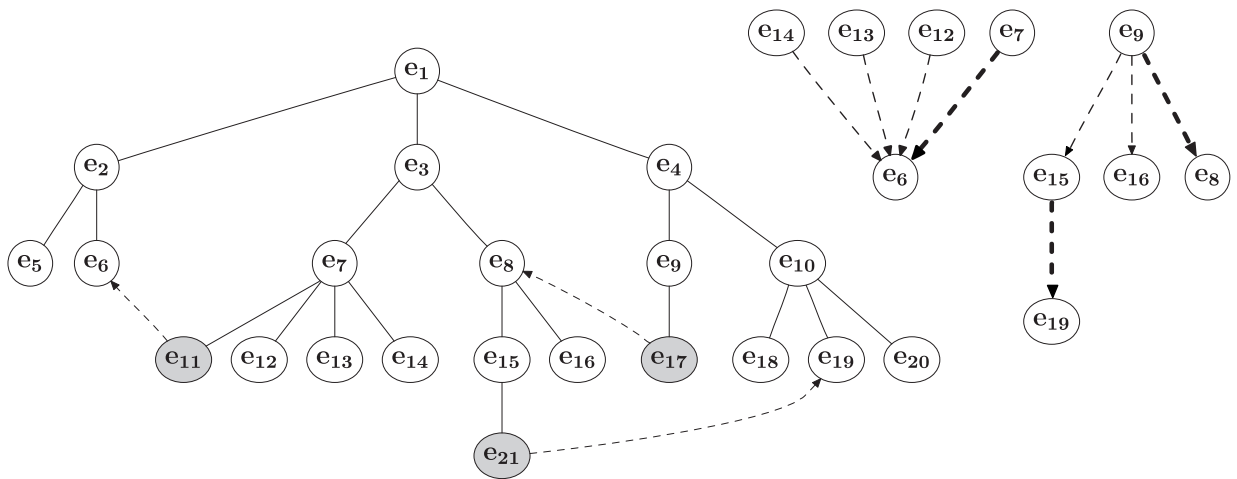


Рис. 2. Структура контента и задаваемое ей отношение *ссылается*

дочерний элемент семантической связи, ссылающийся на s' , — будем считать, что имеет место непосредственная ссылка. Однако, иногда важно рассматривать и опосредованные ссылки, при этом предлагается учитывать ссылки через третьи элементы и ссылки через родительские элементы.

Определение 2. Пусть s, s' — семантические элементы. Будем говорить что s *ссылается* на s' и писать $s \mapsto s'$, если выполнено одно из следующих условий:

- $s' \in LT(CH(\{s\}))$ (*непосредственная ссылка*);
- существует $x \in PR(\{s'\})$ такой, что $s \mapsto x$;
- существует $x \in PR(\{s\})$ такой, что $x \mapsto s'$;
- существует x такой, что $s \mapsto x$ и $x \mapsto s'$.

Следовательно, для каждого семантического элемента s можно определить следующие **характеристики семантической связи**:

- $|LT(CH(\{s\}))|$ — количество семантических элементов, на которые непосредственно *ссылается* s ;
- $|PR(LS(\{s\}))|$ — количество семантических элементов, непосредственно *ссылающихся* на s ;
- $|\{e : s \mapsto e\}|$ — количество семантических элементов, на которые *ссылается* s ;
- $|\{e : e \mapsto s\}|$ — количество семантических элементов, *ссылающихся* на s .

Пример структуры контента и задаваемое ей отношение *ссылается* приведены на рис. 2, жирными линиями выделены непосредственные ссылки.

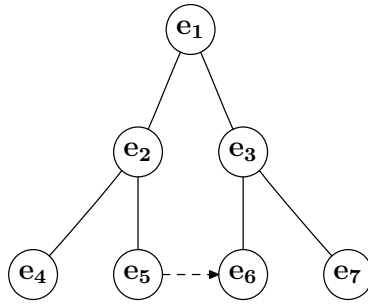


Рис. 3. Структурные элементы e_2, e_3 , связанные через дочерние элементы

Однако, отношение *ссылается* не охватывает все виды семантических зависимостей между элементами, в частности, не учитываются связи через дочерние элементы (рис. 3). Например, если параграф одной лекции ссылается на параграф другой, естественно считать такие лекции в некотором роде связанными. Для анализа подобных зависимостей введем отношение *семантически связаны*. Будем считать элементы семантически связанными, если выполнено одно из следующих условий: элементы совпадают; элементы или их потомки связаны отношением *ссылается*; элементы связаны рекурсивно. Так как каждый элемент оказывается семантически связанным со своими потомками, предлагается рассматривать данное отношение на множествах, не содержащих таких пар элементов.

Определение 3. Обозначим через Z_e — множество элементов поддерева с корнем e . Пусть $A \subset E$ — множество семантических элементов, такое, что для любых $e, e' \in A$ выполнено $e \notin Z_{e'}$ и $e' \notin Z_e$. Для $s, s' \in A$ будем говорить, что s семантически связан с s' ($s \sim s'$), если выполнено одно из следующих условий:

- $s = s'$;
- существуют $x \in Z_s$ и $x' \in Z_{s'}$ такие, что $x \mapsto x'$ или $x' \mapsto x$;
- существует $x \in A$ такой, что $s \sim x$ и $x \sim s'$.

Теорема 2. *Отношение семантической связи (\sim) является отношением эквивалентности на множестве семантических элементов.*

Таким образом, отношение эквивалентности \sim разбивает множество A на семейство подмножеств Y_1, Y_2, \dots, Y_m , содержащих семантически связанные элементы. Например, пусть $A = \{e_5, e_6, e_7, e_8, e_9, e_{10}\}$ (рис. 2), тогда отношение \sim разбивает данное множество на следующие классы эквивалентности: $\{e_5\}$, $\{e_6, e_7\}$, $\{e_8, e_9, e_{10}\}$.

Следовательно, для каждого множества семантических элементов A , удовлетворяющего условиям предыдущего определения, можно рас-

считать следующие **характеристики семантической связи**:

- m — количество классов эквивалентности;
- $|Y_i|$ — количество элементов в данном классе;
- n/m — среднее количество элементов в одном классе.

Характеристики иерархической структуры. Для оценки иерархической структуры контента необходимо рассмотреть ее влияние на возможности обработки данных и, в первую очередь, на поиск информации. Для этого предлагается оценить как много различных подмножеств можно выделить из множества структурных элементов с помощью операций, результат выполнения которых непосредственно определяется иерархической структурой.

Определение 4. Алгеброй \mathcal{A} будем называть кортеж $\langle U, O \rangle$, где:
 O — множество операций $CH, PR, TP_{t_i}, \cup, \cap, \setminus$;
 U — семейство всех подмножеств E , которые можно получить из множеств E применением конечного числа операций из O .

Тогда основной характеристикой иерархической структуры контента является $|U|$ — мощность алгебры \mathcal{A} . Для вычисления данной характеристики введем понятие *отделимых элементов*.

Определение 5. Структурные элементы $e, e' \in E$ назовем *отделимыми*, если существует такое множество $B \in U$, что $e \in B$ и $e' \notin B$. Если такого множества B не существует, e, e' назовем *неотделимыми*. Очевидно, отношение неотделимости является транзитивным и образует классы эквивалентности. Множества, состоящие из всех попарно неотделимых элементов, будем называть *неделимыми*.

Теорема 3. Пусть M_1, M_2, \dots, M_k — семейство всех неделимых множеств, тогда справедливы следующие утверждения:

- $M_i \in U$ для всех i ;
- $M_i \cap M_j = \emptyset$ для всех $i \neq j$;
- $E = \cup_{i=1 \dots k} M_i$;
- $B = \cup_{i \in I} M_i$ для всех $B \in U$, где $I \subseteq \{1, \dots, k\}$.

Следствие 1. Если M_1, M_2, \dots, M_k — семейство всех неделимых множеств, то мощность алгебры \mathcal{A} равна $|U| = 2^k$.

Таким образом, в качестве **характеристик иерархической структуры** предлагается использовать:

- k — количество неделимых множеств;
- 2^k — мощность алгебры;
- $k/|E| \in (0, 1]$ — отношение количества неделимых множеств к максимально возможному количеству неделимых множеств;
- $2^k/2^{|E|} \in (0, 1]$ — отношение мощности алгебры к максимально возможной мощности алгебры.

Для получения разбиения множества структурных элементов на неделимые подмножества необходим критерий принадлежности структурных элементов одному неделимому множеству. С целью построения такого критерия введем понятие *изоморфных деревьев*.

Определение 6. Деревья с корнями e, e' *изоморфны* тогда и только тогда, когда выполнено каждое из следующих условий:

- элементы e, e' имеют один и тот же тип;
- для каждого $x \in CH(\{e\})$ существует $x' \in CH(\{e'\})$ такой, что поддеревья с корнями x, x' *изоморфны*;
- для каждого $x' \in CH(\{e'\})$ существует $x \in CH(\{e\})$ такой, что поддеревья с корнями x, x' *изоморфны*.

Теорема 4. *Если поддеревья с корнями $e, e' \in E$ не изоморфны, то элементы e, e' отделимы.*

Теорема 5 (Критерий делимости). Пусть $e', e'' \in E$. Через e обозначим элемент, являющийся ближайшим общим предком элементов e', e'' . То есть, e является предком каждого из e', e'' , и всякий потомок e не является предком одного из e', e'' . Пусть $e, e'_1, e'_2, \dots, e'_n, e'$ — последовательность дочерних элементов от e к e' , аналогично $e, e''_1, e''_2, \dots, e''_k, e''$ — последовательность дочерних элементов от e к e'' . Элементы e', e'' неотделимы тогда и только тогда, когда выполнены следующие условия:

- поддеревья с корнями e', e'' *изоморфны*;
- $n = k$;
- поддеревья с корнями e'_i, e''_i *изоморфны* для всех i ($1 \leq i \leq n$).

Рассмотрим алгоритм нахождения неделимых множеств, основанный на критерии делимости. Для описания алгоритмов используется псевдоязык по синтаксису схожий с языком SETL2.

Так как критерий делимости базируется на изоморфности поддеревьев, проанализируем вначале вспомогательный алгоритм нахождения изоморфных поддеревьев. Будем рассматривать изоморфизм как бинарное отношение на множестве E , имеющее место для e, e' тогда и только тогда, когда поддеревья с корнями e, e' изоморфны. Очевидно, данное отношение является отношением эквивалентности и разбивает множество E на классы эквивалентности. Процедура `MakeISO(e)`, где e — корень дерева, (рис. 4), разбивает исходное множество на данные классы. В основу алгоритма положены следующие свойства изоморфных поддеревьев. Во-первых, все изоморфные поддеревья имеют одинаковую высоту. Кроме того, пусть семейство множеств ISO содержит классы эквивалентности для всех $e \in E$, таких, что высота поддерева с корнем e не превышает i и высота поддеревьев с корнями

Procedure MakeISO(e)

```
ISO  $\leftarrow \emptyset$ ;  
C  $\leftarrow E$ ;  
for  $i \leftarrow 1$  to  $TreeHeight(e)$  do  
  B  $\leftarrow \{x : x \in C \ \& \ TreeHeight(x) = i\}$ ;  
  for  $e1 \in B$  do  
    if  $e1 \in C$  then  
      A  $\leftarrow \{e2 : e2 \in B \ \& \ IsIsoTree(ISO, e1, e2)\}$ ;  
      C  $\leftarrow C \setminus A$ ;  
      ISO  $\leftarrow ISO \cup \{A\}$ ;  
    end  
  end  
end
```

Function IsIsoTree($ISO, e1, e2$)

```
lsoe1  $\leftarrow \bigcup_{I \in ISO \ \& \ CH(\{e1\}) \cap I \neq \emptyset} I$ ;  
lsoe2  $\leftarrow \bigcup_{I \in ISO \ \& \ CH(\{e2\}) \cap I \neq \emptyset} I$ ;  
return  $G(e1) = G(e2)$  and  $lsoe1 = lsoe2$ ;
```

Function TreeHeight(e)

```
h  $\leftarrow 0$ ;  
for  $x \in CH(\{e\})$  do  
  k  $\leftarrow TreeHeight(x)$ ;  
  if  $k > h$  then  $h \leftarrow k$ ;  
end  
return  $h + 1$ ;
```

Рис. 4. Алгоритм нахождения изоморфных поддеревьев

$e_1, e_2 \in E$ не превышает $i + 1$. При этом поддеревья с корнями e_1, e_2 изоморфны тогда и только тогда, когда выполнены следующие два условия:

1. Типы элементов e_1, e_2 совпадают, т.е. $G(e_1) = G(e_2)$.
2. Дочерний элемент $x_1 \in CH(\{e_1\})$ принадлежит некоторому $I \in ISO$ тогда и только тогда, когда существует $x_2 \in CH(\{e_2\})$ такой, что $x_2 \in I$. То есть, $\bigcup_{I \in ISO \ \& \ CH(\{e_1\}) \cap I \neq \emptyset} I = \bigcup_{I \in ISO \ \& \ CH(\{e_2\}) \cap I \neq \emptyset} I$.

Для реализации алгоритма используются следующие вспомогательные функции: $IsIsoTree(ISO, e1, e2)$ — проверяет являются ли поддеревья с корнями $e1, e2$ изоморфными и $TreeHeight(e)$ — вычисляет высоту поддерева с корнем e .

Рекурсивная процедура $MakeND(e)$, где e — корень дерева, разбивающая структурные элементы на неделимые множества, приведена

Procedure NotDiv(e)

```
MakeISO( $e$  );  
ND  $\leftarrow \emptyset$ ;  
MakeND( $\{e\}$ );
```

Procedure MakeND(M)

```
ND  $\leftarrow$  ND  $\cup$   $\{M\}$ ;  
C  $\leftarrow$  CH( $M$ );  
S  $\leftarrow$   $\{C \cap A : A \in \text{ISO} \ \& \ C \cup A \neq \emptyset\}$ ;  
for  $x \in$  S do MakeND( $x$ );
```

Рис. 5. Алгоритм нахождения неделимых множеств

на рис. 5. Алгоритм основан на следующем свойстве, непосредственно вытекающем из критерия отделимости: если множество M является неделимым, то $e, e' \in CH(M)$ неотделимы тогда и только тогда, когда поддеревья с корнями e, e' изоморфны. Главной процедурой для получения неделимых множеств является $\text{NotDiv}(e)$, результатом ее работы является семейство неделимых множеств ND.

Основными требованиями к структуре контента являются: адекватное представление семантики дидактических материалов и предоставление широких возможностей по обработке контента. Предлагаемые характеристики нацелены на оценку соответствия структуры контента данным требованиям. Характеристики типизации и семантической связи позволяют специалистам-предметникам и персоналу, занимающемуся подготовкой контента к публикации, исследовать его семантическую структуру и оценивать ее соответствие семантике дидактических материалов и закономерностям предметной области. Характеристики иерархической структуры предназначены для оценки возможностей программной обработки контента и являются в некотором роде мерой его структурированности.

Третья глава «Технология создания ОЭИ» посвящена описанию предлагаемой технологии и инструментального программного средства для создания ОЭИ.

Так как в рамках модели непосредственно выделяются структурные элементы, соответствующие дидактическим единицам учебной информации, в разработанной на ее основе технологии стало возможным отделить семантическую структуру контента от механизмов его визуализации и организации диалога с пользователем. Описанный ранее метод спецификации структуры контента позволил автоматизировать

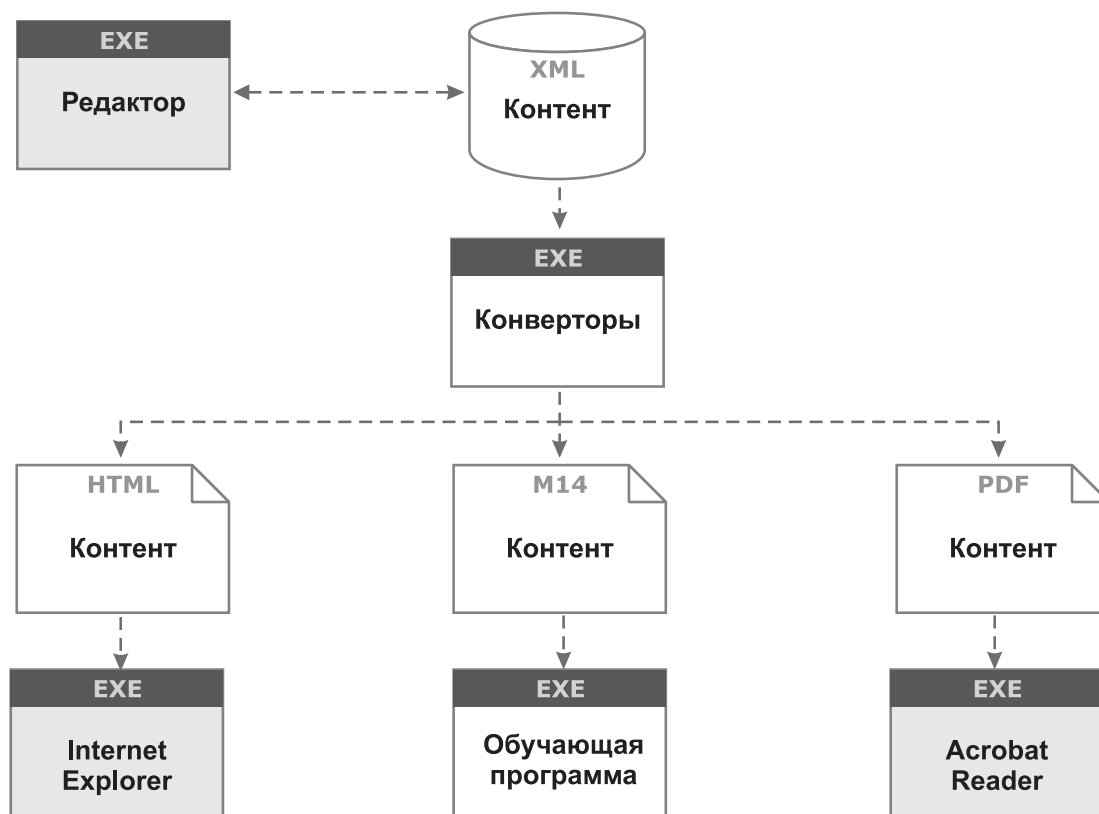


Рис. 6. Структура программного комплекса

контроль его корректности в процессе создания ОЭИ. Предложенные характеристики контента позволили разработать алгоритмы и программы численной оценки его параметров. Важной особенностью предлагаемых технологий является учет семантики контента в процессе его обработки, что позволяет существенно расширить функциональные возможности создаваемых ОЭИ.

Структура программного комплекса. В работе реализован инструментальный программный комплекс для создания ОЭИ, структура программного комплекса показана на рис. 6 (на схемах стандартное программное обеспечение отмечено серым цветом). Для хранения контента используется язык разметки XML, для его редактирования стандартный XML редактор.

Контент может конвертироваться в различные специализированные форматы, для его визуализации и организации диалога используются соответствующие браузеры. Такая технология позволяет создавать различные версии ОЭИ. Инструментальный программный комплекс поддерживает создание сетевых версий, в которых для распространения контента используется Интернет, локальных версий, распространяемых на компакт дисках, и печатных версий. При этом, вследствие учета

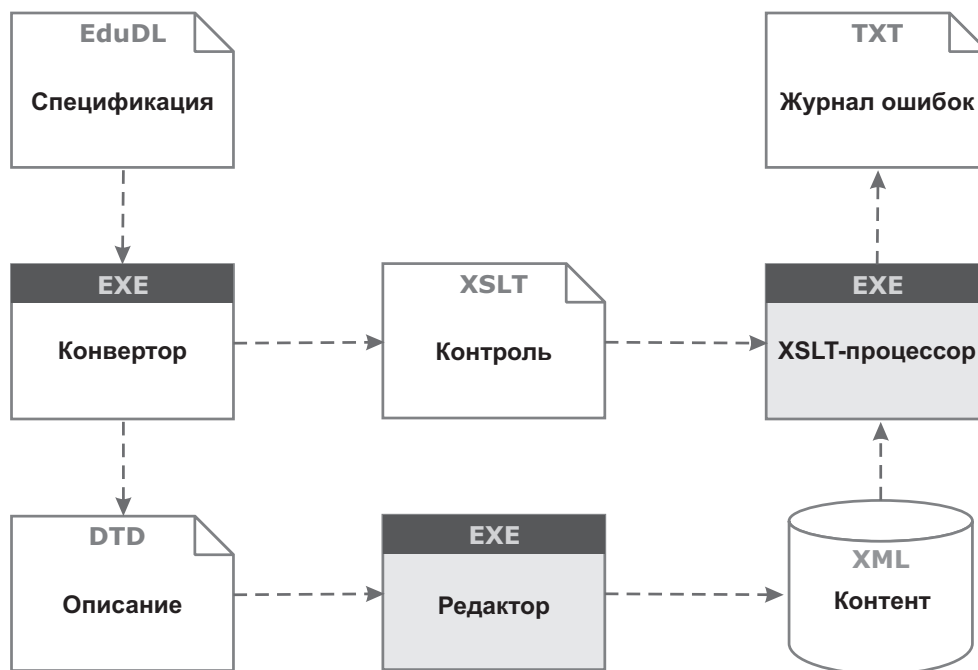


Рис. 7. Технология редактирование и контроль корректности структуры контента

семантики контента при конвертации появляется возможность использовать преимущества каждой из технологий распространения.

Редактирование и контроль. На рис. 7 представлена технология редактирования и контроля корректности контента. Обычно для описания структуры XML документов используется язык DTD, однако он не позволяет накладывать ограничения на семантические связи. Для решения данной проблемы был разработан язык EduDL, основанный на рассмотренном ранее методе спецификации структуры контента. Описание на EduDL преобразуется в DTD описание для контроля в процессе редактирования и в XSLT-преобразование для дополнительного контроля.

Технология создания локальных версий ОЭИ. Технология создания локальных версий ОЭИ базируется на использовании пакета Microsoft MediaView (рис. 8). Контент из XML преобразуется в формат RTF со специфической разметкой. Далее с помощью компилятора MediaView происходит преобразование контента в базу собственного формата. Для работы с данной базой пакет MediaView предоставляет специализированную библиотеку нижнего уровня, на основе которой создана библиотека высокого уровня EduLib. Локальные версии создаются на базе данной библиотеки с использованием языка C++.

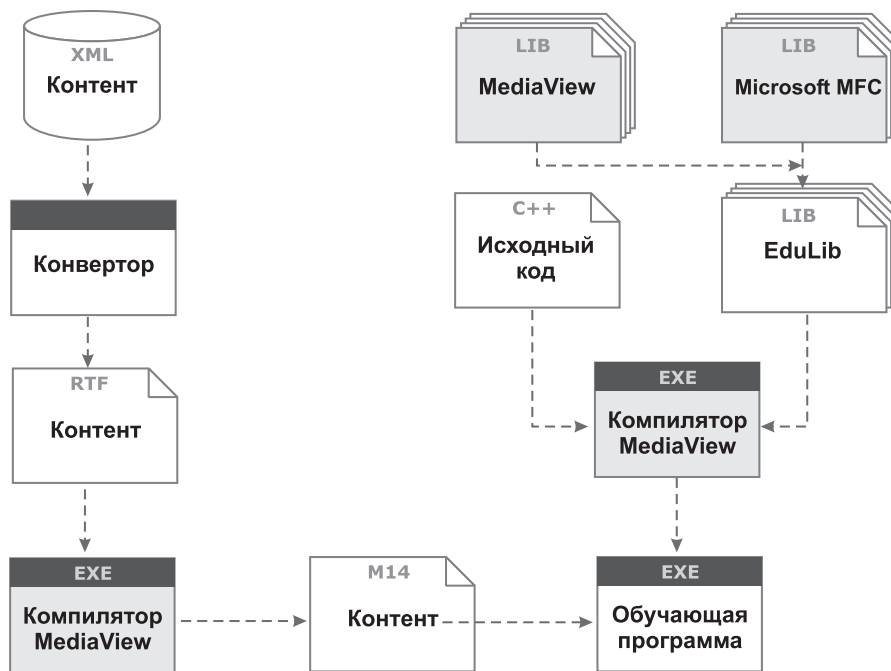


Рис. 8. Технология создания локальных версий ОЭИ

Технология создания сетевых версий ОЭИ. В сетевой версии контент по запросу пользователя с помощью PHP-скриптов и XSLT-преобразований конвертируется в формат HTML и пересылается на клиентскую машину (рис. 9). Для хранения служебной информации (в частности, информации о обучаемых) используется реляционная база.

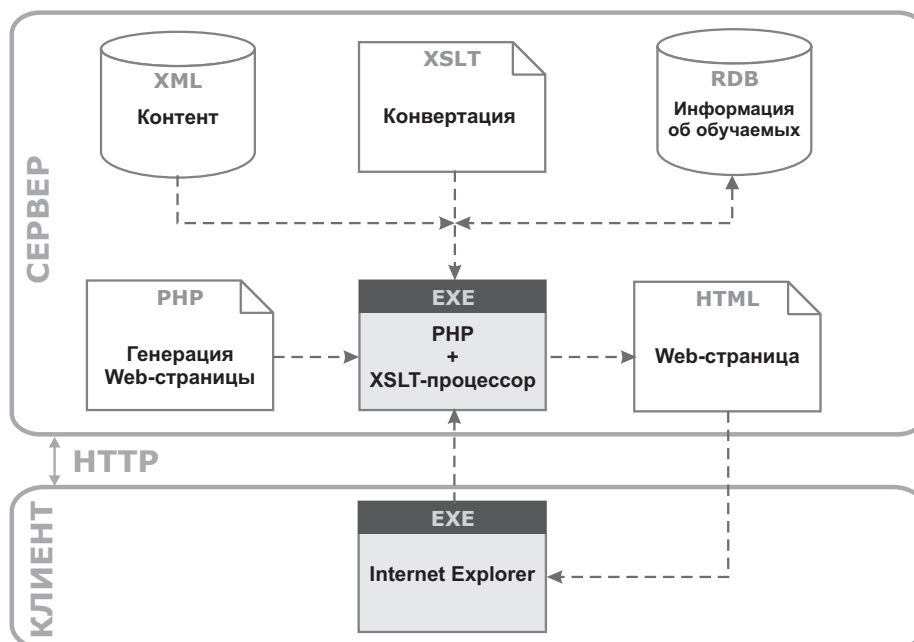


Рис. 9. Технология создания сетевых версий ОЭИ

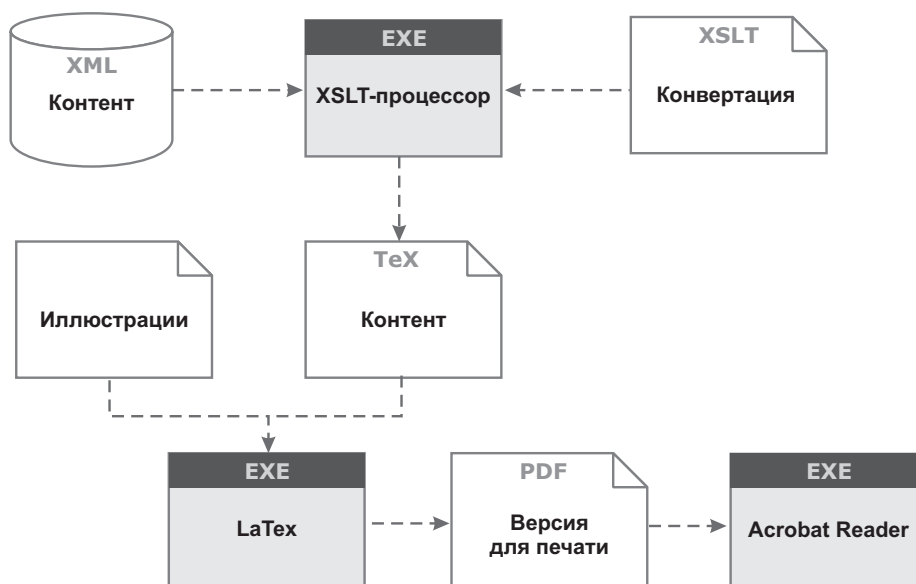


Рис. 10. Технология создания печатных версий ОЭИ

Подготовка печатной версии ОЭИ. Для подготовки печатной версии ОЭИ контент преобразуется в формат TeX, из которого с помощью пакета LaTeX получаем PDF документ, пригодный для печати (рис. 10).

Язык запросов. Основным недостатком современных поисковых систем является невозможность наложения ограничений на семантику информации при задании критериев поиска. Для решения данной проблемы в работе предлагается технология поиска, в основе которой лежит специализированный язык запросов EduQuery (рис. 11), разработанный на базе операторов рассмотренной ранее алгебры. EduQuery позволяет осуществлять выбор элементов контента в соответствии с их типом и содержащейся в них информацией. Кроме того, можно накладывать условия на родительские и дочерние элементы, а также на семантические связи.

```

select
  type = "... "
  val = "... "
  child = (select ...)
  parent = (select ...)
  refin type = "... "(select ...)
  refout type = "... "(select ...)
...
select
  ...
  
```

Рис. 11. Язык запросов EduQuery

В заключении приведены основные результаты диссертационной работы:

1. Построена модель представления контента ОЭИ, позволяющая отражать семантическую структуру учебной информации и поддерживающая визуализацию контента в форме, ориентированной на его изучение.
2. Разработаны методы и алгоритмы количественной оценки структурных характеристик контента. Предлагаемые характеристики позволяют исследовать соответствие структуры контента семантике дидактических материалов и оценить возможности его программной обработки.
3. Создана технология разработки образовательных электронных изданий, позволяющая учитывать семантику контента при реализации основных функций ОЭИ.
4. Разработан инструментальный программный комплекс, поддерживающий создание сетевых, локальных и печатных версий ОЭИ.
5. С использованием предложенных моделей и технологий созданы ОЭИ, внедренные в учебный процесс ряда высших учебных заведений Российской Федерации.

Основное содержание диссертации опубликовано в следующих работах:

1. *Захарова И. Г., Семикин В. А.* О некоторых общих принципах разработки обучающих систем // Математическое и информационное моделирование: Сборник статей. — Тюмень: Изд-во ТюмГУ, 2000. — С. 171–177.
2. *Деревнина А. Ю., Семикин В. А.* Об одном подходе к созданию электронных учебников гуманитарного направления // Тезисы докладов семинара «Электронные учебники и учебно-методические разработки в открытом образовании». — М.: Изд-во МЭСИ, 2000. — С. 65–68.
3. *Деревнина А. Ю., Семикин В. А.* Разработка интегрированной системы создания и демонстрации электронных учебников гуманитарного направления // Совершенствование образовательной деятельности. Сборник докладов часть II: Серия материалов Всероссийской школы семинара «Информационные технологии в управлении качеством образования и развитии образовательного пространства». — М.: Исследовательский центр проблем качества подготовки специалистов, 2000. — С. 73–75.

4. Деревнина А. Ю., Семикин В. А., Кошелев М. Б. Принципы создания электронных учебников // *Открытое образование*. — 2001. — № 2. — С. 14–17.
5. Деревнина А. Ю., Семикин В. А., Кошелев М. Б. Системы тестирования в электронных учебниках // *Информационные технологии*. — 2002. — № 5. — С. 39–44.
6. Деревнина А. Ю., Семикин В. А. Подготовка контента электронных учебных изданий на основе XML-технологий // Труды IX Всероссийской научно-методической конференции «Телематика'2002». — СПб.: 2002. — С. 246–247.
7. Деревнина А. Ю., Семикин В. А. Технология создания электронных учебников: контент+браузер // *Индустрия образования: Сборник статей*. — М.: МГИУ, 2002. — Т. 2. — С. 431–438.
8. Деревнина А. Ю., Семикин В. А. Об оценке качества электронных учебников // *Качество высшего профессионального образования в начале XXI века: Материалы всероссийской научно-методической конференции*. — Тюмень: Изд-во ТюмГУ, 2002. — С. 82–86.
9. Разработка комплекса электронных средств поддержки обучения по 11 общим гуманитарным и социально-экономическим дисциплинам для системы ВПО, предусмотренным государственным образовательным стандартом для всех непрофильных специальностей: Отчет о НИР Министерства образования РФ № 2.7.5.(43.1) 228.059, № ГР 01.20.00 09209, Инв. № 02.20.02 04563. Тюмень, 2002. / А. Ю. Деревнина, В. А. Семикин, Г. Ф. Куцев, Г. М. Заболотная. — Тюмень: Тюменский госуниверситет, 2002. — 146 с.
10. Деревнина А. Ю., Семикин В. А. Об организации данных в интернет порталах // Труды X Всероссийской научно-методической конференции «Телематика'2003». — СПб.: 2003. — С. 218–219.
11. Семикин В. А. Семантическая модель контента образовательных электронных изданий // *Математическое и информационное моделирование: сборник статей*. — Тюмень: Изд-во ТюмГУ, 2004.