

$$\sigma_i = \delta(x_*, y_*, z_*) \cdot \delta(x_i, y_i, z_i).$$

Тогда, если $\sigma_i > 0$, то точка (x_*, y_*, z_*) лежит с той же стороны от плоскости раздела фаз, что и центр i -го шара (x_i, y_i, z_i) , и, следовательно, принадлежит соответствующей фазе. В случае $\sigma_i < 0$ эта точка принадлежит другой фазе. При $\sigma_i = 0$ данную точку можно отнести произвольно к любой из двух фаз.

Векторная имитационная модель дает возможность существенно снизить требования к вычислительным ресурсам системы за счет:

1. Векторного представления геометрических объектов, которое позволяет хранить в памяти только уравнения поверхностей, ограничивающих эти объекты.

2. Использования модели массового обслуживания, которая дает возможность перейти от дискретизации по времени к дискретизации по событиям, что уменьшает количество пересчетов состояния системы.

СПИСОК ЛИТЕРАТУРЫ

1. Ермаков С. С., Вязников Н. Ф. Порошковые стали и изделия. Л.: Машиностроение, Ленинградское отделение, 1990. С. 319.

2. Гуревич Ю. Г., Ивашко А. Г., Цыганова М. С. Математическое моделирование распада аустенита с целью построения термокинетической диаграммы расчетным путем // Известия вузов. Черная металлургия. 2004. № 9. С. 45-48.

3. Гуляев А. П. Металловедение. М.: Металлургия, 1977. С. 647.

4. Бусленко Н. П. Моделирование сложных систем. М.: Наука, 1978. С. 400.

*Александр Григорьевич ИВАШКО —
профессор кафедры информационных систем,
доктор технических наук*

*Юлия Владимировна БИДУЛЯ —
старший преподаватель
кафедры компьютерных технологий*

УДК 004.94

МОДЕЛИРОВАНИЕ СМЫСЛОВОГО ОПИСАНИЯ КОНТЕНТА

АННОТАЦИЯ. В статье излагаются принципы построения модели понятийного описания контента в целях программной реализации смыслового анализа материала. Предложен алгоритм преобразования текста в семантическую сеть объектов и критерии сравнения полученных объектов для нескольких видов учебного материала.

This article states a construction principles of content term describing model with the purpose of text meaning program realization. A script of text transforming to an object semantic net as well as objects comparing criterions for some kinds of educational materials are offered.

Процесс подготовки учебного контента состоит из довольно большого числа итераций, в которых автор вынужден многократно обращаться к одному и тому же материалу. Учебный контент представляет многоликое отражение материалов, описывающих предметную область, таких, как конспекты лекций, словарь терминов, часто задаваемые вопросы, тесты и т.д. Согласованность в отображении информации требует обработки больших объемов информации,

которую можно реализовать машинным способом. Все вышесказанное определяет востребованность программных средств, позволяющих осуществлять анализ контента на интеллектуальном, понятийном уровне. Результат такого анализа выражается в определении ключевых понятий текста, установления взаимосвязей между ними, выявления общих и различных черт. Полученные данные дают возможность построить семантическую модель или фреймовую модель, которые могут быть применены для проверки согласованности различных отображений учебного контента.

На рынке довольно много программных продуктов, реализующих лексический анализ текста [1]. В данной работе ставится задача формализации описания различных моделей представления контента с целью построения модели преобразования лексического представления текста в смысловое.

Понятийное, смысловое описание опирается на понятия, фигурирующие в тексте, и должно удовлетворять следующим требованиям:

1. Возможность выявления всех видов отношений между понятиями.
2. Исключение повторов описаний отношений между объектами во избежание избыточности хранения смысловой информации.
3. Возможность формализованного представления структуры материала в виде записей в базе данных или разметки текста для дальнейшего анализа программными средствами.

Постановка задачи моделирования смыслового описания материала определяется, в первую очередь, возможностью автоматизации понятийного анализа для использования в информационных системах образовательного направления. Модель описания должна включать в себя три составляющие:

- 1) лексическая модель исходного контента;
- 2) модель семантической сети;
- 3) алгоритм перехода от модели исходного контента к семантической сети.

Исходный контент можно представить в виде совокупности текстовых единиц — лексем $L = \{l_1, l_2, \dots, l_n\}$, каждая из которых относится к определенному частеречному типу из множества типов $T = \{t_1, t_2, \dots, t_k\}$ и обладает характеристиками $V^t = \{b_1^{t_i}, b_2^{t_i}, \dots, b_m^{t_i}\}$, определяемыми для каждого типа t_i . Множество типов определено, ограничено и не зависит от содержимого контента. То же самое можно сказать и о характеристиках текстовых единиц, определенных для каждого типа. Между текстовыми единицами определяются связи r_{ij} типов t_k . Множество всех связей обозначим $R = \{r_{ij}\}$, его мощность определяется исходным контентом. Множество всех типов связей $F = \{f_1, f_2, \dots, f_q\}$ определено независимо от контента.

Таким образом, каждая единица исходного контента в рамках предлагаемой лексической модели представляется в виде собственно содержимого текстовой единицы l_j (цепочки символов естественного языка), типа t_i и набора характеристик $\langle b_1^{t_i}, b_2^{t_i}, \dots, b_m^{t_i} \rangle$.

Контент, структурированный в соответствии с лексической моделью, можно рассматривать как упорядоченный набор $D = \langle L, T, V, R, F \rangle$

Современные программные средства, к примеру, серия программных продуктов, выпускаемых под маркой Russian Context Optimizer, предназначены для поддержки средств поиска и анализа текстовой информации и реализуют следующие возможности:

- 1) выделение именных групп, определение их производности, определение тематического веса (для ранжирования по значимости в тексте);

- 2) определение числа слов в каждой именной группе, главного и зависимых слов;
- 3) определение части речи и грамматических характеристик каждого слова (род, число, падеж и т.д.);
- 4) определение синтаксической роли слова в предложении;
- 5) нахождение синтаксико-семантических отношений между словами и определение их типа.

Полученные данные нельзя трактовать как смысловое описание текста, так как построенная на их основании семантическая сеть дает лишь общее представление о контенте. Смысловое описание должно содержать информацию о каждом ключевом понятии контента, о его связях с другими понятиями в целях сопоставления их свойств и выявления характера отношений между ними в контексте решения перечисленных выше практических задач.

С другой стороны, указанные результаты позволяют реализовать описанную выше модель исходного контента, а именно: выделить лексемы l , определить тип каждой лексемы t , характеристики b и связи g типов f . Взаимосвязи между лексемами устанавливаются на основе их грамматических и синтаксических характеристик.

Перейдем к вопросу модели семантической сети, которая, по сути, и определяет смысловое описание. Ее построение возможно, если каждое ключевое понятие представить в виде объекта, обладающего признаками, определяемыми на основании контента, а также классовой структурой, образующей иерархию. Объект модели — это некоторая сущность, предмет, понятие, фигурирующее в контенте. Соотношения и взаимодействия понятий представляются в виде связей между объектами.

В качестве прообраза объектов предлагается принять понятие *лингвистического фрейма*. Теория фреймов в лингвистике имеет свою историю. Считается, что впервые термин «фрейм» был введен М. Минским [2]. Он интерпретирует фрейм как структуру данных, предназначенную для представления стереотипной ситуации. Фреймы определяются как структуры признаков, характеризующих участников и их действия (роли) в конкретных ситуациях.

Предлагались различные определения и классификации лингвистических фреймов для решения лингвистических задач распознавания и представления смысла текста. К примеру, С. А. Жаботинская [3] использует пять базисных частеречных фреймов. В них представлены концепты, относящиеся к основным понятийным категориям. В предметном фрейме одна и та же сущность характеризуется по своим количественным, качественным, бытийным, локативным и темпоральным параметрам. В акциональном фрейме несколько предметов, являющихся участниками события, наделены аргументными ролями. Эти предметы объединяются связями, которые заданы действием агенса и обозначены глаголом «действует» или «делает» в сопровождении предлогов. Поссесивный фрейм демонстрирует связь обладатель (*possessor*)-обладемое (*possessed*). Таксономический фрейм представляет отношения категоризации, которые проявляются в трех вариантах: вид, род и роль. Компаративный фрейм, примыкающий к таксономическому фрейму, формируется связями тождества, сходства и подобия. Объединенные друг с другом, фреймы образуют межфреймовую сеть, которая организует информацию о предметах, их свойствах и отношениях.

Описание контента в виде сети лингвистических фреймов позволяет оперировать ключевыми понятиями. Недостатком такого описания является его недоста-

точная формализованность, что не позволяет воспользоваться им для построения алгоритмов и дальнейшего программирования. С другой стороны, выступая основой для объекта семантической модели, лингвистический фрейм позволяет отобразить все те свойства, которые необходимы для смыслового описания контента.

В целях отражения в модели таксономических и родо-видовых отношений все понятия удобно представить в виде иерархической структуры. Каждая единица e_i смыслового описания представляет либо экземпляр, либо класс или подкласс. Степень «вложенности» классов теоретически ничем не ограничивается, на практике же определяется содержанием исходного контента. Экземпляр является самой нижней ступенью в иерархии объектов. Имеет смысл ввести понятие уровня иерархии для понятийного анализа.

Таким образом, смысловое описание контента представляется в виде совокупности единиц $E = \{e_1, e_2, \dots, e_s\}$, где s — общее число понятий контента.

Каждая единица e_i характеризуется принадлежностью классу c_j с уровнем иерархии $I = (0, 1, 2, \dots, r)$. Экземпляр — это e_i такое, что $I = 0$. Для всех единиц с $I > 0$ определен класс.

В семантической сети каждый элемент связан с другим элементом. В данной модели элемент, представляющий экземпляр, может быть связан отношением s_k как с другим экземпляром, так и с классом. Множество всех связей $S = \{s_1, s_2, \dots, s_k\}$ определяется исходным контентом. Каждая связь имеет тип k из множества типов $K = \{k_1, k_2, \dots, k_n\}$. Отношения могут иметь носитель (к примеру, предикатное), а могут и не иметь (пример: отношение принадлежности). В зависимости от типа связи имеет характеристики $P^k = \{p_1^k, p_2^k, \dots, p_m^k\}$, определяемые для каждого типа k . Контент, структурированный в соответствии с моделью семантической сети, можно рассматривать как упорядоченный набор $G = \langle E, S, P, K \rangle$

Рассмотрим переход от лексической модели к семантической модели контента. Формально он представляется в виде функции $TR: D \rightarrow G$, где $D = \langle L, T, V, R, F \rangle$ и $G = \langle E, S, P, K \rangle$ — упорядоченные наборы, представляющие структурированный контент. Функция определяется в виде алгоритма преобразования множества лексем и связей исходного контента во множество экземпляров, классов и связей семантической модели.

Этап 1. Каждую лексему l_j исходной модели необходимо отнести к одному или нескольким элементам e_i семантической модели.

1. К классу объектов [кл.], если частеречный тип $t_i =$ «имя существительное нарицательное»;

2. К экземпляру класса объектов [экз.], если частеречный тип $t_i =$ «имя собственное». При этом:

а) Если из текста можно сделать вывод о классе, класс также определяется, например:

Вася Иванов [экз.] — студент [кл.].

б) Один экземпляр может принадлежать одновременно двум и более классам, например:

Мой друг [кл.1] Вася Иванов [экз.] — студент [кл.2].

в) Класс может быть не определен из текста. В таком случае он подразумевается и считается неопределенным. Пример:

Вася Иванов [экз.] вошел в аудиторию [кл.].

3. К признаку, определяющему подкласс объектов, если частеречный тип $t_i =$ {«имя прилагательное», «числительное», «местоимение»}. Пример:

мой друг: друг — класс, мой друг — подкласс.

Подкласс образуется в том случае, если в исходном контенте лексема-признак связана атрибутивным или именным атрибутивным отношением с лексемой типа «существительное нарицательное».

4. К носителю предикативного отношения [пред.] между экземплярами или классами объектов.

Носителем может выступать лексема типа а) «глагол» б) «причастие» в) «деепричастие»

Студент [кл.1] вошел [пред.] в аудиторию [кл.2].

5. К признаку предикативного отношения [призн.пред.] между экземплярами или классами объектов.

Студент [кл.1] быстро [призн.пред.] вошел [пред.] в аудиторию [кл.2].

6. К коннектору предикативного отношения [конн.] между экземплярами или классами объектов (предлог).

Студент [кл.1] вошел [пред.] в [конн.] аудиторию [кл.2].

В результате на этапе 1 идентифицируются объекты и связи между ними.

Этап 2. Данные по каждому объекту приводятся к единому стандарту описания. К примеру, определяется порядок перечисления: класс, признаки класса, связи класса, подкласс, признаки подкласса, связи подкласса, экземпляр, признаки экземпляра, связи экземпляра.

Этап 3. Производится сравнение объектов на предмет совпадения названий (цепочек символов). Каждое совпадение e_i и $e_{i,1}$ при наличии некоторых условий означает отнесение их к одному классу или экземпляру, а все связи e_i и $e_{i,1}$ с другими объектами суммируются.

Образованная таким образом сеть иерархических объектов отражает смысловую направленность контента, так как опирается на частеречные фреймы, по определению предназначенные для описания смысла контента.

Охват смысловых характеристик контента семантической моделью G можно оценить при помощи табл. 1.

Таблица 1

Охват смысловых характеристик семантической сетью

Факторы частеречного фрейма Жаботинской	Представление в семантической сети	Примечания
1	2	3
1. Предметный фрейм, параметры:		
а. Количественный (сколько?)	Кванторность класса/подкласса как «несколько» Число экземпляров класса/подкласса	Имя числительное сопоставляется числовой переменной, значение которой определяет число экземпляров класса или подкласса.
б. Качественный (какой?)	Признак экземпляра или класса	Определяются атрибутивные связи с существительными или именами собственными, их тематический вес, в соответствии с которым присваиваются идентификаторы классов, экземпляров.

Продолжение табл. 1

1	2	3
с. Бытийный (как?)	Носитель и признак предикативного отношения	Для каждого предиката определяются наречные связи
d. Локативный (где?)	Носитель и экземпляр предикативного отношения	Данное определение переключается с представлением акционального фрейма в роли «Цель».
е. Темпоральный (когда?)	Носитель и признак предикативного отношения	
2. Акциональный фрейм	Носитель и экземпляр предикативного отношения	Набор носителей предикативного отношения может определять подмножество объектов одного класса.
3. Посессивный фрейм	Связь между двумя экземплярами и классами	
4. Таксономический фрейм	Связь между классом и экземпляром	
5. Компаративный фрейм	Носитель и экземпляр предикативного отношения	

Построение сети иерархических объектов для нескольких видов учебного контента позволяет провести сравнение их содержания. Описанная модель позволяет сформулировать критерии сравнения:

1. Наличие идентичных объектов.
2. «Глубина» иерархии этих объектов.
3. Наличие у этих объектов идентичных связей.
4. Наличие противоречивых связей объектов.
5. Сходства и различия признаков объектов.

Подобное сравнение контента в значительной степени превосходит обычное реферирование, позволяя избежать повторов, тавтологий, неточностей формулировок, синонимичность высказываний.

Выводы.

1. Формализовано описание лексической модели и модели частеречных фреймов, дающее возможность реализовать алгоритм преобразования одной модели в другую.

2. Разработан алгоритм преобразования лексической модели в модель частеречных фреймов.

3. Показано, что модель частеречных фреймов и семантические модели дают возможность производить сравнение различных отображений образовательного контента.

СПИСОК ЛИТЕРАТУРЫ

1. Ермаков А. Е., Плешко В. В. Синтаксический разбор в системах статистического анализа текста // Информационные технологии. 2002. № 7. С. 30-34.
2. Минский М. Структура для представления знаний // Психология машинного зрения. М.: Мир, 1978. С. 250-338.
3. Жаботинская С. А. Концептуальный анализ: Типы фреймов // Вестник Черкасского университета. Вып. 11. 1999. С. 3-20.