носильными, необходимо «уравновешивать» качество доставляемого пользователю учебного контента в рамках всего курса. Кроме того, руководствуясь представленными в статье подходами, можно столкнуться с «измельчением» учебных объектов, что не всегда хорошо: с одной стороны, чем меньше размер учебного объекта (в смысле информации), тем больше вероятность того, что объект будет полезным и будет использоваться многократно; с другой стороны, если учебный объект слишком мал, тогда есть опасность, что он станет бессодержательным и непригодным для многовариантного использования. Однако, несмотря на указанные недостатки, подобные подходы имеют свои преимущества: они создают большие возможности для варьирования и при создании учебного контента, и при его освоении, а значит, и для повышения эффективности образовательного процесса.

## СПИСОК ЛИТЕРАТУРЫ

1. Hodgins, Wayne. Into the future // http://www.learnativity.com/download/MP7.PDF.

2. Материалы комитета стандартов образовательных технологий IEEE // http://ieeeltsc.org.

3. Норенков И.П. Технологии разделяемых единиц контента для создания и сопровождения информационно-образовательных сред // Информационные технологии. 2003. № 8. С. 34-39

4. Гильманов А.С. Информационное моделирование обучающей системы, использующей гибридные технологии доставки контента // Современные проблемы математическ го и информационного моделирования. Перспективы разработки и внедрения информац онных IT-решений. Сб. науч. тр. под ред. Кутрунова В.Н. Тюмень. Изд-во ТюмГУ, 2008

5. Информационно-аналитические материалы Российского портала открытого обра зования // http://www.openet.ru

***Юлия Владимировна БИДУЛЯ —***
***ст. преподаватель кафедры***
***компьютерных технологий***
***Института математики и компьютерных наук***
***Тюменского государственного университета***

**УДК 004.421**

# *АЛГОРИТМИЗАЦИЯ СМЫСЛОВОГО ОПИСАНИЯ КОНТЕНТА*

*АННОТАЦИЯ. В статье излагаются принципы построения модели описания контента в целях программной реализации смыслового анализа учебно-методических материалов. Показана взаимосвязь синтаксических характеристик единиц предложения с семантической структурой всего контента. Предложен алгоритм преобразования синтаксического описания предложений текста в семантическую сеть ключевых понятий.*

*This article states a construction principles of content term describing model with the purpose of text meaning program realization. A correlation between syntactic characteristics of sentence units and whole content semantic structure is represented. A script of syntactic describing text sentences transforming to an keywords semantic net is offered.*

Подготовка учебного материала для дисциплин как дистанционного, так и очного обучения подразумевает работу преподавателя с различными источника-

ми информации, будь то учебники или научные статьи, материалы Интернета и собственные исследования.

Автоматизация обработки контента позволяет значительно сократить временные затраты на подготовку учебно-методических материалов. Это определяет востребованность программных средств, позволяющих осуществлять анализ контента на смысловом, понятийном уровне. 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 следующих практические задач:

1. Возможность анализа контента с точки зрения полноты описания значимых понятий.

2. Анализ и сравнение двух текстов родственного содержания.

3. Возможность формирования контрольных вопросов по значимым объектам.

Рассмотрим процесс построения смысловой модели контента. Очевидно, что моделирование проводится в два этапа:

1. Синтаксический анализ на уровне каждого предложения. Результатом является список его семантически значимых текстовых единиц и различных типов отношений между ними.

2. Семантический анализ всего контента. Результат: семантическая сеть, концептами которой выступают ключевые понятия, являющиеся с точки зрения грамматики и синтаксиса русского языка именными субтантивными словосочетаниями или *именными группами* [1]. Следующий шаг в построении семантической сети — определение отношений между именными группами. Определяющими для смыслового описания являются *предикативные* отношения [2].

Таким образом, на первом этапе единицей синтаксического анализа является отдельное предложение текста, для которого строится сеть синтаксических отношений между составляющими его текстовыми единицами — словоформами. Семантический анализ текста базируется на результатах синтаксического анализа. Переход от синтаксического анализа к смысловому означает преобразование модели структуры предложения в модель структуры всего контента, отражающей его смысловой портрет.

Синтаксическая структура предложения — это совокупность сведений о связях между его словами и словосочетаниями. Синтаксическая структура предложения может быть представлена деревом синтаксического подчинения или просто деревом подчинения, заданным на множестве словоформ предложения [3]. Обозначим это множество $T=\{t_i^{(s)}\}$ в виде совокупности словоформ $t_i^{(s)}$, определяемых для каждого предложения с уникальным идентификатором $s$.

Каждой текстовой единице ставится в соответствие лексема $l \in L$, где $L$ — множество всех лексем контента. Каждую лексему отнесем к частеречной категории $k$ из множества категорий $K$.

Каждой лексеме соответствует конечное число $m^{(k)}$ возможных словоформ, каждая из которых в свою очередь характеризуется набором грамматических характеристик $F^{(k)}=\{fi(k)\}$. Число возможных словоформ $m^{(k)}$ определяется категорией лексемы $k$. К примеру, каждой словоформе лексемы категории «существительное» соответствуют такие характеристики, как $F^{(k)}=\{$«род», «число», «падеж»$\}$. В таком случае, число $m^{(k)}$ определяется количеством сочетаний этих трех характеристик.

В итоге каждая структурная единица t сети s-того предложения описывается в виде

$$t^{(s)}\ (l,\ k,\ f_1^{(k)},\ f_2^{(k)},\ ...,\ f\ m^{(k)}),$$

где:
$l$ — лексема, представителем которой является словоформа t,
$f_1^{(k)}$, $f_2^{(k)}$, ..., $fm^{(k)}$ — характеристики словоформы t.

Словоформы, составляющие словосочетания, находятся в определенных синтаксических отношениях, которые строятся на основе взаимодействия лексических значений этих слов и их грамматических форм. Классификация таких отношений обобщенно сводится к основным типам: атрибутивное, предикатное, аппозитивное, определительное, обстоятельственное, отпредложное. В каждом отношении словоформ одна из них является главной, а другая — зависимой [3].

Определим множество $B^{(s)}=\{b_{ij}^{(s)}\}$, каждый элемент которого $b_{ij}$ представляет синтаксическое отношение между i-той и j-той текстовыми единицами s-того предложения. Кроме того, на основании классификации отношений определим множество типов отношений $C=\{c_n\}$. Каждое отношение определяется упорядоченной парой, т.е. $b_{ij}^{(s)}=<t_i^{(s)}, t_j^{(s)}, c_n>$. Примем за правило ставить на первое место текстовую единицу, которая является в данном отношении главным словом, а на второе место, соответственно — зависимое.

Таким образом, структуру s-того предложения в соответствии с описанной моделью можно рассматривать как упорядоченный набор $D^{(s)}=<T,B>$, где T — множество словоформ s-того предложения, B — множество возможных отношений, определяемых упорядоченными парами $b_{ij}^{(s)}$ словоформ s-того предложени

Описание структуры контента представим следующим образом. Обозначи U — множество всех $u_i$ — именных групп контента. Исходя из постанов задачи и определения семантической сети, можно утверждать, что каждая имен ная группа участвует в отношении $r_{ij}\in R$ по крайней мере с одной именной группой. Поскольку речь идет о предикативных связях, то каждая именная группа участвует в конкретном семантическом падеже [1]. Следовательно, отношение между $u_i$ и $p_j$ несимметрично. Кроме того, предикативная связь по определению выражается сказуемым (чаще всего глаголом) и может иметь признаки действия. Таким образом, каждое отношение между ключевыми понятиями можно представить в виде

$$r_{ij}=<u_i, u_j, p_{ij}, a_{ij}, v_p>,$$

где $u_i$, $u_j$ — именные группы контента, $p_{ij}$ — носитель предикативного отношения, $a_{ij}$ — атрибут предикативного отношения, $v_p$ — семантический падеж. Контент, структурированный в соответствии с моделью семантической сети, можно рассматривать как упорядоченный набор $Q=<U, R>$.

Переход к семантической структуре контента определим в виде функции преобразования $F$:

$$F: \{D^{(s)}\rightarrow Q,$$

где $D^{(s)}$ — упорядоченный набор, представляющий синтаксическую структуру s-того предложения.
Q — упорядоченный набор, представляющий структурированный контент.

Функция преобразования по сути представляет алгоритм перехода от синтаксического описания предложений контента к описанию в терминах ключевых понятий и связей между ними. Этот алгоритм можно охарактеризовать следующими утверждениями:

1. Входная информация для алгоритма формируется в результате синтаксического анализа каждого s-того предложения и представляется в виде набора пар отношений $b_{ij}^{(s)}=<t_i^{(s)}, t_j^{(s)}, c_n>$.

2. В каждом предложении выделяются пары текстовых единиц с общим словом, связанные определительным, аппозитивным или атрибутивным отношением. Эти пары объединяются в именные группы.

3. В каждом предложении выделяются пары текстовых единиц, связанные предикатным отношением. Главная текстовая единица относится к носителю $p_{ij}$ отношения $r_{ij}$. Для зависимой единицы устанавливается ее участие в именной группе.

4. В каждом предложении выделяются пары текстовых единиц, связанные обстоятельственным отношением. Для главной текстовой единицы устанавливается ее участие в предикатном отношении $r_{ij}$. Зависимая текстовая единица относится к атрибуту $a_{ij}$ отношения $r_{ij}$, если она относится к категории «наречие». В противном случае ее участие устанавливается в именной группе, а семантический падеж $v_p$ на основании грамматической характеристики $f_m^{(k)}$.

5. На выходе алгоритма формируется набор пар отношений $r_{ij}=<u_i, u_j, p_{ij}, a_{ij}, v_p>$, характеризующих весь контент.

Выводы:

1. Предложено формальное описание синтаксической модели предложения и семантической структуры контента.

2. Разработан алгоритм перехода от синтаксической модели к семантической.

3. Показана взаимосвязь между синтаксическими характеристиками единиц предложения и структурными единицами предложенного смыслового описания 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. Гладкий А.В., Мельчук И.А. Грамматики деревьев I. Опыт формализации преобразований синтаксических структур естественного языка. Сб. Инф. вопросы семиотики, лингвистики и автоматического перевода. М.: Наука, 1971.

3. Гладкий А.В. Формальные грамматики и языки. М.: Наука, 1973.

***Сергей Александрович БАРДАСОВ —***
***доцент кафедры экономики***
***и управления собственностью***
***Тюменского государственного университета,***
***кандидат физико-математических наук***

**УДК 519.224.2**

# *ПРИМЕНЕНИЕ ИНФОРМАЦИОННОГО КРИТЕРИЯ АКАЙКА ДЛЯ ОПРЕДЕЛЕНИЯ ОПТИМАЛЬНОГО ЧИСЛА ГРУПП ВАРИАЦИОННОГО РЯДА*

*АННОТАЦИЯ. Информационный критерий Акайка применяется для определения оптимального числа групп при построении гистограммы. Рассматриваются случаи интервалов равной длины и равной вероятности. Для распределения Гаусса получены формулы, позволяющие рассчитать число групп по заданному объему выборки.*