

Александр Александрович Рыбанов, доцент, заведующий кафедрой информатики и технологии программирования Волжского политехнического института (филиал) Волгоградского государственного технического университета, кандидат технических наук

ПРОЦЕДУРА КВАНТОВАНИЯ КАК ОСНОВА ПОВЫШЕНИЯ КАЧЕСТВА УЧЕБНОГО МАТЕРИАЛА

Существующие в настоящее время системы дистанционного обучения Moodle, Pias, Claroline, Atutor и другие не предоставляют разработчикам дистанционных учебных курсов возможность оценки качества учебного контента. Между тем, оценка учебного контента направлена не только на выявление достоинств и недостатков учебной информации, но и на принятие решения о необходимости и оптимальных условиях его использования в процессе дистанционного обучения¹.

Квантование — это разбиение учебной информации на элементарные фрагменты (учебные единицы, шаги, кадры) различного назначения (информационные, тренирующие, контролируемые, управляющие), направленное на уменьшение сложности усвоения смыслов, содержащихся в каждом фрагменте учебной информации². Объём текстовой информации в этих фрагментах должен быть ограничен.

Процесс квантования — это преобразование: $T' = f(T)$, где $T = T_i | i = (\overline{1, n})$ — учебная информация, предназначенная для квантования, T_i — логически законченный фрагмент учебной информации; $T' = (T_i' | i = (\overline{1, n}))$ — представление учебной информации, где T_i' — квант учебной информации, поставленный в соответствие фрагменту T_i .

Принцип системного квантования учебной информации предполагает учёт следующих закономерностей: учебная информация большого объёма запоминается с трудом; учебная информация, представленная компактно в определённой системе, лучше вос-

¹ Аванесов В.С. Применение заданий в тестовой форме и квантованных учебных текстов в новых образовательных технологиях // Педагогические измерения, 2012, № 2, С. 75–91.

² Рыбанов А.А. Автоматизированный анализ качества процесса обучения по результатам тестирования знаний на основе диаграмм Парето. // Дистанционное и виртуальное обучение, 2009, № 8, С. 54–59.

принимается; выделение в учебной информации смысловых единиц способствует эффективному запоминанию.

Учитывая то, что квант T_i' учебной информации должен содержать наиболее информативную часть фрагмента T_i , требования к кванту учебной информации можно формализовать следующим образом: квант T_i' учебной информации должен обладать более низкой избыточностью и большей энтропией, чем T_i ; квант T_i' учебной информации по объёму должен быть меньше соответствующего ему фрагмента T_i учебной информации: $|T_i'| < |T_i|$.

Процесс построения педагогом кванта T_i' для фрагмента T_i учебной информации состоит из следующих этапов:

1) подготовительный (чтение и осмысление фрагмента T_i учебной информации);

2) аналитический (выделение основных смысловых единиц (предложения, слова, словосочетания), построение структуры кванта T_i' для фрагмента T_i учебной информации);

3) непосредственное построение кванта T_i' для фрагмента T_i учебной информации (выделенные ранее единицы располагаются в единый вторичный текст в соответствии со структурой кванта T_i').

В качестве смысловых единиц кванта T_i' для фрагмента T_i учебной информации могут быть использованы: γ_1 — полное (без изменений) ключевое предложение исходного текста T_i ; γ_2 — перефразированное ключевое предложение исходного текста T_i ; γ_3 — предложение из ключевых слов и словосочетаний исходного текста T_i ; γ_4 — предложение, обобщающее несколько предложений исходного текста T_i .

Качественно подготовленные квантованные учебные тексты обеспечивают понимание учебного материала у большинства учащихся, в силу того, что деление материала на части заметно сужает объём непосредственно воспринимаемой информации и смыслов в каждом фрагменте, делая тем самым постигаемым смыслы всего учебного текста. А работа с заданиями в тестовой форме к таким текстам обеспечивает усвоение содержания каждого текста. Поэтому формирование системы количественных критериев для оценки качества квантования учебной информации является актуальной задачей. Одно из направлений решения данной задачи — использование методов количественной лингвистики. Количественные характеристики текста могут составлять основу данной системы критериев.

Количественная лингвистика (quantitative linguistics) — одно из направлений прикладной лингвистики, которое занимается изучением языка с помощью статистических методов. Преимуществом количественных методов изучения текстов является их точность и однозначность результатов.

Важной является задача автоматизированного подсчёта значений количественных характеристик текста. Программная реализация автоматизированного определения ряда количественных характеристик текста возможна на основе библиотеки морфологического анализа phpMorphy (<http://phpmorphy.sourceforge.net>), реализованной на платформе PHP.

Среди множества количественных характеристик текста рассмотрим следующие: количественные характеристики употребления частей речи; количественные характерис-

тики удобочитаемости текста; количественные характеристики разнообразия текста.

С помощью библиотеки phpMorphy могут быть определены следующие количественные характеристики употребления частей речи в тексте:

- *индекс аналитичности* — отношение числа служебных слов к общему числу слов в тексте;
- *индекс глагольности* — отношение числа глаголов к числу слов в тексте;
- *индекс субстантивности* — отношение числа существительных к числу слов в тексте;
- *индекс адъективности* — отношение числа прилагательных к числу слов в тексте;
- *индекс местоименности* — отношение числа местоименных слов к числу слов в тексте;
- *индекс автосемантической* — отношение числа значащих слов;
- *индекс незначительности* — отношение числа незначительных слов к числу слов в тексте;
- *индекс именной лексики* — отношение суммы чисел существительных и прилагательных к числу слов в тексте.

Среди *количественных характеристик удобочитаемости текста* можно выделить: среднюю длину слова в слогах и среднюю длину предложения в словах. Данные характеристики являются статистическими параметрами, которые используются в формулах для оценки удобочитаемости и необходимы для расчёта индекса формальной удобочитаемости. Эти параметры легко поддаются количественному выражению и пригодны для автоматической оценки.

Следует различать формальную удобочитаемость текста $R_{form}(I)$, являющегося функцией только от параметров самого учебного контента I , и его индивидуальную удобочитаемость $R_{ind}(I, u)$, которая зависит как от характеристик учебного контента I , так и от свойств читателя u . Для количественной оценки формальной удобочитаемости можно использовать индекс Ю. Тулдава³, рассчитываемый по формуле:

$$R(\bar{i}, \bar{j}) = \bar{i} \cdot \lg \cdot \bar{j}, \quad (3)$$

где $R(\bar{i}, \bar{j})$ — индекс формальной удобочитаемости (рисунок 1), \bar{i} — средняя длина слова в слогах, \bar{j} — средняя длина предложений в словах. Формула (3) разработана на основе закономерности, наблюдаемой в разных языках. Поэтому формула Ю. Тулдава предназначена для анализа текста на разных языках. Чем меньше значение $R(\bar{i}, \bar{j})$, тем лучше для восприятия текст.

Количественные характеристики разнообразия текста описываются коэффициентами лексического и синтаксического разнообразия.

Коэффициент лексического разнообразия формируется из отношения числа лексем к общему числу слов текста:

$$K_{lex} = \frac{L}{W}, \quad (1)$$

где K_{lex} — коэффициент лексического разнообразия; L — число лексем (словоформ)

³ Tuldava J. About Measurement of Text Difficulties. In: Proc. Of Tartu State University, 1975, P. 102–120.

в тексте; W — число слов в тексте (единицы между пробелами). Чем выше значение K_{lex} , тем выше лексическое разнообразие текста.

Коэффициент синтаксического разнообразия вытекает из отношения числа предложений к числу слов данного текста:

$$K_{syn} = I - \frac{S}{W}, \quad (2)$$

где K_{syn} — коэффициент синтаксического разнообразия; S — число предложений; W — число слов в тексте. Чем больше значение K_{syn} , тем многословнее в целом предложения данного текста, а следовательно, выше возможность разнообразия синтаксических отношений между словами в составе отдельного предложения.

В результате квантования происходит сжатие предложений исходного текста при помощи приёмов: исключение (γ_3), замена (γ_2) и слияние (γ_4), что приводит к изменению количественных характеристик текста. Результаты эксперимента, проведённого на коллекции учебных текстов, позволяют сделать следующие выводы:

1. Значение индекса $R(i, j)$ для квантованного текста меньше, чем для исходного.

2. Значение коэффициента лексического разнообразия для квантованного текста несколько выше, чем для исходного.

3. Значение коэффициента синтаксического разнообразия для квантованного текста несколько ниже, чем для исходного.

4. Сравнительный анализ значений индексов формальной удобочитаемости, а также коэффициентов лексического и синтаксического разнообразия для соответствующих фрагментов исходного и квантованного текстов позволяет определить фрагменты квантованного текста, для которых необходимо повторное выполнение процедуры квантования учебной информации.

Предлагаемая система количественных характеристик (1–3) учебного контента применима к слабоструктурированным текстам. Данная система критериев неприменима к формулам, таблицам, графическим и мультимедийным объектам. Учитывая, что данные объекты, как правило, не подлежат квантованию, система количественных характеристик (1–3) может успешно применяться в составе автоматизированных систем проектирования учебного контента.