

УДК 37.02

DOI: 10.26140/anip-2019-0804-0026

ПРОБЛЕМА ОЦЕНКИ СЛОЖНОСТИ ДИДАКТИЧЕСКИХ ОБЪЕКТОВ И ЕЕ РЕШЕНИЕ

© 2019

AuthorID: 373852

SPIN: 6444-8902

ResearcherID: I-4605-2014

ORCID: 0000-0001-8166-9299

Майер Роберт Валерьевич, доктор педагогических наук, доцент; профессор кафедры
физики и дидактики физики

*Глазовский государственный педагогический институт им. В.Г.Короленко
(427621, Россия, Глазов, Первомайская, 25, ГГПИ, e-mail: robert_maier@mail.ru)*

Аннотация. Предлагается метод оценки семантической сложности дидактических объектов: учебных тестов, рисунков, формул и т.д. Он состоит в следующем: 1) создают текст, который дает полное и одновременно краткое описание исследуемого дидактического объекта, не содержащее смысловых пустот; 2) определяют количество семантической информации в нем, для этого производят автоматизированный подсчет терминов (слов) с учетом их информационной емкости. Для анализа файла с изучаемым текстом используется специальная компьютерная программа, взаимодействующая с другим файлом, содержащим список терминов и их сложность. Информативность термина определяется путем подсчета числа значимых слов в его определении или объяснении, понятном ученику с заданным тезаурусом. В качестве условной единицы смысловой информации взято количество информации, содержащейся в обычных словах, не требующих объяснения. Дидактическая сложность учебного текста находится как произведение его структурной сложности и семантической сложности. Структурная сложность текста может быть охарактеризована произведением средней длины слова (в слогах) на логарифм от среднего числа слов в предложении, которое увеличено на 1. Представлены результаты использования данного метода, заключающиеся в измерении общей информативности и средней плотности информации в различных учебных текстах, параграфах и их фрагментах, рисунках и формулах.

Ключевые слова: дидактика, контент-анализ, обучение, сворачиваемость знаний, сложность, учебник.

THE COMPLEXITY ASSESSMENT PROBLEM OF DIDACTIC OBJECTS AND ITS DECISION

© 2019

Mayer Robert Valerievich, Doctor of Pedagogical Sciences, Docent; Professor
of Chair of Physics and Didactic of Physics
Glazov Korolenko State Pedagogical Institute

(427621, Russia, Glazov, Pervomayskaya st., 25, GGPI, e-mail: robert_maier@mail.ru)

Abstract. A method for evaluating the semantic complexity of didactic objects (educational tests, pictures, formulas, etc.) is proposed. It consists of the following: 1) create a text that gives a complete and simultaneous brief description of the studied didactic object and does not contain semantic voids; 2) determine the amount of semantic information in it; for this an automated calculation of terms (words) is made taking into account their information capacity. To analyze a file with the studied text, a special computer program is used, that interact with another file containing a list of terms and their complexity. The information content of the term is determined by counting the number of significant words in its definition or explanation, understandable for the student with a given thesaurus. As a conventional unit of semantic information, the amount of information contained in ordinary words, that do not require explanation, is taken. The didactic complexity of the educational text is found as the product of its structural complexity and semantic complexity. The structural complexity of the text can be characterized by the product of the average word length (in syllables) and the logarithm of the average number of words in the sentence, which is increased by 1. The results of using this method are presented, consisting in measuring the overall informativeness and average density of information in various educational texts, paragraphs and fragments thereof, pictures and formulas.

Keywords: didactics, complexity, content analysis, knowledge folding, textbook, training.

Введение. Информационно-кибернетический подход к анализу дидактических систем предполагает сведение процесса обучения к: 1) передаче учебной информации от учителя к ученику по каналу связи с помехами; 2) выполнению учеником последовательности специальным образом подобранных заданий. При этом следует учесть количество и скорость передачи информации, длительность обучения, информативность передаваемых сообщений, пропускную способность канала связи, сложность выполнения заданий и т.д. Все это превращает разработку методов количественной оценки сложности и информативности учебных текстов в одну из актуальных проблем современной дидактики. Ее решение позволит оценить информационный объем и сложность различных дидактических объектов.

В процессе обучения учитель и ученик оперируют различными дидактическими объектами (ДО), которые можно разделить на два класса: 1) элементы научного знания: понятия, факты, теории, описания научных приборов и экспериментов; 2) средства обучения: бумажные и электронные учебники, методические пособия, тесты, учебные опыты и компьютерные модели. Все они характеризуются различными параметрами (дидактиче-

скими характеристиками), величины которых влияют на результат использования данного объекта в обучении.

Учебный текст (УТ) – это многомерный объект, характеризующийся большим количеством величин, поэтому обозначенная проблема относится к слабоформализуемым и может быть решена различными способами. Важными характеристиками учебных текстов и других ДО являются объем, общая информативность и дидактическая сложность (ДС). Разработка методов их измерения поможет расположить изучаемые вопросы в порядке возрастания сложности, а также правильно оценить трудоемкость тестовых заданий и точнее определить уровень знаний ученика. Проблема оценки ДС различных текстов является актуальной потому, что: 1) реализация принципа “от простого к сложному” требует изучения вопросов в порядке возрастания их сложности; 2) сложность учебного материала должна соответствовать способностям ученика; 3) сложность (информативность) учебного материала может рассматриваться как характеристика интеллектуального уровня школьника, способного его усвоить.

Цель статьи состоит в обосновании метода оценки дидактической сложности учебных текстов и других

ДО, а также в изучении возможностей его применения на практике.

Решение этой задачи перекликается с проблемами использования математических методов в гуманитарных исследованиях, качественного и математического моделирования обучения, измерения количества и сложности информации в сообщении, передачи учебной информации по каналу учитель–ученик.

Методологической основой настоящего исследования являются работы известных ученых по следующим направлениям: 1) семантическая информация: В.Б. Вяткин [1], О.В. Зеркаль [2], И.П. Кузнецов [3]; 2) тезаурусный подход: Н.В. Лукашевич [4], Вал.А. Луков, Вл.А. Луков [5], А.А. Мирошниченко [6]; 3) сложность систем: Р.Р. Гайсин [7], Д.П. Клейносов [8]; 4) трудность и сложность учебных текстов: Г.А. Балл [9], А.В. Гидлевский [10], А.С. Кисельников [11]; 5) теория учебника, сложность и информативность УТ: В.П. Беспалько [12], Я.А. Микк [13]; 6) контент-анализ текстов: M.D. White, E.E. Marsh [14]; 7) автоматизированная оценка сложности текстов: И.В. Оборнева [15], М.М. Невдах [16]. Для решения поставленной задачи нами использовались метод системного анализа, тезаурусный подход, метод контент-анализа и методология мягких систем. Математическая обработка результатов осуществлялась в электронных таблицах MS Excel. Статья является развитием подходов, рассмотренных автором в работах [17–19].

1. Основные подходы к оценке сложности дидактических объектов.

Дидактические объекты, которые ученик должен усвоить (текст, формулы, рисунки), в конечном счете могут быть заменены текстовым описанием, которое однозначно, кратко и полно передает содержание и структуру оцениваемого ДО. Поэтому задача определения объема и общей информативности ДО может быть сведена к изменению объема и информативности соответствующего ему учебного текста. Вообще, дидактическая сложность УТ может быть найдена как произведение структурной сложности S_{str} , учитывающей средние длины используемых слов, формул и предложений, и семантической сложности S_{sem} , зависящей от смыслового содержания текстовой, формульной и графической составляющих УТ. Структурная сложность одного предложения или текста может быть охарактеризована произведением средней длины слова (в слогах) на логарифм от среднего числа слов в предложении, которое увеличено на 1. Эта формула согласуется с *законом Хика*: время реакции человека (которое пропорционально сложности задачи) на N раздражителей (слов в предложении) пропорционально $\ln(N + 1)$. Чем больше информации поступает в мозг человека, тем больше времени требуется для ее обработки при прочих равных условиях. Оценка семантической сложности фактически состоит в определении общей информативности текста и коэффициента свернутости информации относительно некоторого заданного тезауруса Z_0 (например, уровня знаний пятиклассника).

Мозг – это сложная нейросеть, оперирующая концептами; он воспринимает и порождает информацию в виде последовательности суждений, состоящих из “обычных” слов и научных терминов. Поэтому количество смысловой информации и семантическую сложность УТ следует измерять не в битах, а в словах [18]. За условную единицу смысловой информации (УЕИ) имеет смысл взять количество информации, содержащейся в ежедневно используемых человеком словах, обозначающих конкретные объекты, их признаки или действия и не требующих объяснения: “птица”, “темный”, “светит” и т.д. Это удобно с практической точки зрения потому, что: 1) слово – основная структурно-семантическая единица языка; 2) научные термины и обычные слова отражают объективные особенности восприятия человеком окружающего мира; 3) количество усвоенных учеником понятий можно оценить методом тестирования; 4) объ-

ем текста в словах пропорционален времени его чтения или пересказа на уроке.

В УТ может быть реализован формально-логический способ изложения материала, он может иметь высокую насыщенность терминами и содержать логические рассуждения, являющиеся результатом выдвижения гипотез, объяснения, логического вывода, доказательства, анализа фактов, подведения итогов и т.д. Часто эти операции выделяются автором УТ с помощью соответствующих слов: поэтому, следовательно, предположим, докажем и т.п. Чтобы при измерении информативности УТ учесть его логическую структуру, необходимо выявить логические связи и вместо них в текст добавить слова типа: “потому что” или “следовательно”. Для оценки S_{sem} существенным является количество связей, а не то, какие именно высказывания ими соединены.

Если переходя от одного высказывания к другому, ученик должен догадываться о каких-то сложных фактах, не упомянутых в объяснении, то такой текст называется *напряженным*. Например, читая учебник по обществузнанию, ученик может понять смысл всех слов в предложении, но не осознать заключенную в нем мысль. Правильно составленный УТ не требует большого напряжения от ученика, так как вся важная информация содержится в предложениях явно. Если УТ является напряженным, то соответствующие смысловые пустоты следует дополнить пояснениями так, чтобы ученик с тезаурусом Z_0 мог без напряжения его понять. В этом случае количество информации, представленной неявно, будет близко к нулю, и им можно пренебречь. Трудность понимания одного предложения или фрагмента УТ определяется средней плотностью семантической информации в нем.

Количество информации в УТ зависит от его объема и степени абстрактности используемых понятий, которые характеризуется *коэффициентом свернутости информации* (КСИ).

Вообще, *свертывание информации* – это процесс ее аналитико-синтетической переработки, приводящий к уменьшению физического объема сообщения без потерь информативности. При этом получается “содержательно насыщенное” знание, происходит уплотнение и сжатие информации, создание новых форм научного знания с большей информационной емкостью. Свертывание и уплотнение знаний сопровождается “кумуляцией” информации, уменьшением ее объема путем более короткого, обобщенного изложения и приводит к образованию новых семантических единиц и их проецированию на другие области знаний. При этом информационные емкости однокоренных слов, относящихся к одной или разным частям речи (например, “интерференция”, “интерференционный”, “интерферировать”), примерно одинаковы, так как они являются составляющими одного концепта. Чем выше уровень абстрактности обозначаемого понятия, тем больше информационная емкость слова, а значит и его семантическая сложность.

Итак, КСИ в слове C относительно уровня знаний Z_0 показывает степень концентрации информации в нем и равен числу слов в объяснении сущности понятия, обозначаемого словом C , человеку со знаниями Z_0 . Иначе говоря, необходимо дать определение O термину T , а также определить терминам T_1 , T_2 , ..., которые входят в O , используя слова из тезауруса Z_0 , а затем сосчитать количество всех значимых слов в получившемся тексте.

Если для объяснения термина “ускорение”, выпускнику 5 класса, требуется произнести 8 слов, то информативность (или КСИ) этого термина относительно соответствующего тезауруса равна 8 УЕИ (усл. ед. инф.). Слова, не нуждающиеся в объяснении, имеют информативность 1 УЕИ. Будем считать, что УТ соответствует тезаурусу Z_0 (понятен ученику с тезаурусом Z_0), если кроме обычных слов он содержит только термины из Z_0 и термины, объясненные в тексте.

Информативность предложения относительно тезауруса Z_0 равна объему текста минимальной длины, соответствующему тезаурусу Z_0 , и содержащему ту же полезную информацию. Среднее арифметическое значение КСИ всех слов в УТ фактически является средней информационной свернутостью текста, средней информативностью слова или *средней плотностью информации* в данном УТ.

Желательно, чтобы уровень Z_0 , относительно которого измеряются информативности текстов, не превышал тезауруса, требуемого для понимания самого простого УТ из данного корпуса текстов. Тогда семантические сложности текстов будут больше 1.

2. Предлагаемый метод оценки сложности дидактических объектов.

Итак, для оценки ДС объекта его заменяют максимально полным и в то же время коротким описанием, а затем оценивают ДС получившегося текста относительно выбранного тезауруса Z_0 . В основе измерения ДС текста лежат следующие идеи. ДС текста равна произведению структурной сложности S_{str} УТ на семантическую сложность S_{sem} . Структурная сложность текста равна произведению средней длины слова в слогах D_{word} на $\ln(D_{sent} + 1)$, где D_{sent} – среднее число слов в предложении. Семантическая сложность УТ характеризуется количеством содержащейся в нем семантической информации.

Информативность слова (термина) I_i относительно тезауруса Z_0 равна количеству значимых слов в определении или объяснении данного слова, понятном для ученика с тезаурусом Z_0 . Общая (суммарная, интегральная) информативность Inf УТ относительно тезауруса Z_0 – величина, показывающая количество семантической информации, представленной во всем УТ; она приблизительно равна сумме информативностей $I_i = s_i$ всех терминов, составляющих УТ, или произведению средней информативности слова относительно Z_0 на количество слов N_{word} в УТ:

$$Inf = \sum_{i=1}^{N_{word}} I_i = I_{sr} \cdot N_{word}.$$

Примем без доказательства следующий тезис: учебные тексты, созданные из одной и той же совокупности слов $\{W_1, W_2, \dots\}$, и передающие одинаковое множество мыслей $\{M_1, M_2, \dots\}$, содержат примерно одинаковое количество семантической информации (сложности). Если УТ не содержит повторов и пропущенных логических звеньев, написан лаконичным языком, то: 1) его семантическая сложность примерно равна сумме сложностей предложений; 2) семантическая сложность предложений УТ примерно равна сумме информативностей составляющих его слов.

При наличии в УТ логических переходов, сложных для понимания учеником с тезаурусом Z_0 , необходимо дополнить УТ словами типа “значит”, “поэтому” и поясняющими предложениями, упрощающими для ученика установление логических связей и делающими эти переходы легкими.

В общих чертах методика оценки семантической сложности текстовой составляющей УТ состоит в следующем: 1. Из исходного текста удаляют повторы, стоп-слова, заменяют местоимения соответствующими терминами. 2. Если текст напряженный, вставляют поясняющие предложения, заполняя смысловые пустоты. Выявляют логические связи и вместо них в текст добавляют слова типа: “потому что”, “следовательно” и т.д. 3. С помощью специальной программы подсчитывают число значимых слов в УТ и определяют его объем. 4. Создают текстовый файл slovar.txt, содержащий список из N используемых научных терминов. 5. Подсчитывая количество слов в определениях терминов, находят их сложности s_i ($i = 1, 2, \dots, N$), которые записывают в

файл slovar.txt; 6. Используя компьютерную программу, обращающуюся к файлу slovar.txt, анализируют файл с текстом, подсчитывают количество n_i упоминаний каждого термина и число обычных слов N_{word} . 7. Суммируя сложности всех научных терминов и “обычных” слов, определяют сложность текста: S_{sem} . 8. Вычисляют среднюю информативность слова для данного УТ I_{sr} . 9. Чтобы учесть информативность формул или рисунков, заменяют их максимально кратким описанием и оценивают его сложность тем же методом.

3. Возможности метода оценки сложности дидактических объектов.

Рассмотренный выше метод был апробирован нами для оценки ДС текстов, их фрагментов, формулировок законов, доказательств теорем, формул и их выводов, рисунков и других ДО. Перечислим задачи, которые при этом удалось решить [17-19]:

1. Измерены общее количество информации и средняя плотность информации в отдельных параграфах различных учебников по физике. Это позволило осуществить их классификацию по следующим признакам: 1) сложность текстовой составляющей (терминологическая) ST ; 2) сложностью формульной составляющей УТ (математическая) SM . Эти характеристики слабо коррелируют друг с другом и поэтому могут рассматриваться как независимые атрибуты УТ.

2. Произведена оценка средней плотности информации в параграфах школьных и вузовских учебников физики, посвященных изучению электрического и магнитного полей. Установлено, что среднее количество информации, приходящееся на одно значимое слово или символ в формуле, находится в интервале от 2,4 до 3,9. Учет средней длины предложений и количества символов в формулах позволяет оценить среднюю сложность текста; ее величина изменяется от 3 до 16.

3. Произведена оценка силы меж- и внутрпредметных связей путем подсчета числа терминов, относящихся к одной предметной области (например, к физике микромира – ФММ), в текстах, соответствующих другим областям знаний (разделам физики и химии). В результате контент-анализа учебного пособия [20] установлено: 1) доля знаний ФММ в курсе химии (0,26) существенно выше, чем в курсе физики в целом (0,05) и заметно выше, чем в пятом разделе, посвященном явлениям микромира (0,19); 2) разнообразие ФММ-информации в пятом разделе физики существенно выше, чем в курсе химии; 3) при изучении химии наиболее часто применяются понятия “атом”, “молекула”, “ион” (включая “анион” и “катион”).

4. Измерена сложность объяснения решения десяти типовых задач из различных разделов школьного курса физики. Обнаружено, что общая информативность объяснения стандартной задачи составляет 70 – 750 УЕИ, а средний коэффициент свернутости информации лежит в интервале 1,5 – 6,4.

5. Определена дидактическая сложность важнейших вопросов школьного курса математики. В результате оценки 27 параграфов по математике из различных классов установлено, что их плотность информации (а значит и трудность понимания) за время обучения в школе возрастает в 150–200 раз.

6. Определена плотность информации, содержащейся в теоретических положениях школьного курса математики, соответствующих 1 – 11 классам. Установлено, что средняя плотность информации увеличивается от 2,8 в 1 кл. до 29,9 в 11 кл., то есть примерно в 10 раз. При этом с 1 до 9 класса КСИ возрастает медленно от 2,8 до 10,3 (в среднем на 0,9 в год), а с 9 до 11 класса – очень быстро от 10,3 до 30 (в среднем на 10 в год). Это обусловлено качественными изменениями в преподавании математики после 9 класса.

7. Приблизительно определена сложность математической информации в школьных учебниках физики, оценены суммарная информативность формульной инфор-

мации и коэффициенты свернутости информации. Обнаружено, что КСИ лежит в интервале от 2 до 4, достигая своего максимального значения для 8 класса. Последнее обусловлено тем, что в 8 классе изучаются тепловые, электрические, электромагнитные и световые явления, поэтому информационная емкость физических величин в формулах достаточно велика. В 9 классе КС меньше, так как половину учебного года школьники изучают механические явления. Средняя сложность формул S лежит в интервале от 2,6 до 5,8. Сложность формульной информации, приходящаяся на 1000 знаков, изменяется от 0,9 (7 кл.) до 3,8 (11 кл.).

8. Определено количество информации в электрических схемах [18], произведено их сравнение. Создана компьютерная программа, которая анализирует файл, содержащий список всех приборов, составляющих электрическую цепь, и задает вопросы о количестве узлов и ветвей в цепи. Программа работает со специальным файлом, в котором перечислены названия приборов (резистор, конденсатор, операционный усилитель и т. д.) и для каждого прибора указана сложность. Аналогичным образом решена задача для химических формул.

Выводы. Итак, в статье показана актуальность оценки сложности дидактических объектов, сводимых к текстам: учебных тестов, рисунков, формул и т. д. Предлагаемый метод состоит в создании текста, который дает полное и в то же время краткое описание изучаемого ДО, и автоматизированном подсчете терминов с учетом их информационной емкости. Последняя оценивается путем нахождения количества слов в определении данного термина и фактически равна коэффициенту свернутости информации в нем. Показано, что обсуждаемый метод позволяет решить большое количество разнообразных задач, связанных с измерением общей информативности и средней плотности информации в УТ или других ДО.

СПИСОК ЛИТЕРАТУРЫ:

1. Вяткин В. Б. Синергетическая теория информации: пояснения и терминологические замечания // Научный журнал КубГАУ. 2012. № 80(06).
2. Зеркаль О. В. Семантическая информация и подходы к ее оценке. Часть 1. Семантико-прагматическая информация и логико-семантическая концепция // Философия науки. 2014. № 1. С. 53–69.
3. Кузнецов И. П. Механизмы обработки семантической информации. М.: Наука, 1978. 174 с.
4. Лукашевич Н. В. Тезаурусы в задачах информационного поиска. М., 2010. 396 с.
5. Луков Вал. А., Луков Вл. А. Методология тезаурусного подхода: стратегия понимания // Знание. Понимание. Умение. 2014. № 1. С. 18–35.
6. Мирошниченко А. А. Профессионально ориентированные структуры учебных элементов. Глазов, 1999. 62 с.
7. Гайсин Р. Р. Принцип сложности в естественнонаучном познании: Методологический анализ: Дисс. ... канд. философск. наук. Уфа, 2002. 169 с.
8. Клейнсов Д. П. Изменение структурной и содержательной сложности учебного материала с целью реализации дидактического принципа осознанности знаний. Дисс. ... канд. пед. наук. Москва, 2017. 150 с.
9. Балл Г. А. Теория учебных задач: Психолого-педагогический аспект. М.: Педагогика, 1990. 184 с.
10. Гидлевский А. В. Исчисление трудности дидактической задачи // Вестник Омского университета. 2010. № 4. С. 241–246.
11. Кисельников А. С. К проблеме характеристик текста: читабельность, понятность, сложность, трудность // Филологические науки. Вопросы теории и практики. № 11 (53). 2015. Часть 2. С. 79–84.
12. Беспалько В. П. Теория учебника: Дидактический аспект. М.: Педагогика, 1988. 160 с.
13. Микк Я. А. Оптимизация сложности учебного текста: В помощь авторам и редакторам. М.: Просвещение, 1981. 119 с.
14. White M. D., Marsh E. E. Content analysis: A flexible methodology // Library trends. 2006, Vol. 55, № 1. pp. 22–45.
15. Оборнева И. В. Автоматизированная оценка сложности учебных текстов на основе статистических параметров // Дис. ... канд. пед. наук. Москва, 2006. 165 с.
16. Невдах М. М. Исследование информационных характеристик учебного текста методами многомерного статистического анализа // Прикладная информатика: Изд. «НОУ «МФПУ «Синергия». 2008. № 4. С. 117–130.
17. Майер Р. В. Классификация тем школьного курса физики на основе оценки их физической и математической сложности // Инновации в образовании. 2014. № 9. С. 29–38.

18. Майер Р. В. Контент-анализ школьных учебников по естественно-научным дисциплинам: монография. – Глазов: Глазов. гос. пед. ин-т, 2016. 137 с.

19. Mayer R.V. Computer Model of the Empirical Knowledge of Physics Formation: Coordination with Testing Results // European Journal of Contemporary Education. 2016. Vol. 16. Is. 2. pp. 239–247.

20. Справочник школьника: 5–11 классы // М. Б. Волович, О. Ф. Кабардин, Р. А. Лидин и др. М.: АСТ–ИПЕСС, 2001. 704 с.

Статья поступила в редакцию 29.08.2019

Статья принята к публикации 27.11.2019