1. Использован метод математического планирования эксперимента при проектировании состава шлакового бетона.

2. Получены зависимости прочности бетона на шлаковом вяжущем от расхода крупного заполнителя (доменного щебня) и мелкого заполнителя (кварцевого песка), которые позволят прогнозировать марку бетона по прочности в зависимости от соотношения мелкого и крупного заполнителя.

Выполненная работа позволяет сделать вывод, что необходимы дальнейшие исследования с большим количеством входных факторов – варьировать водоцементное отношение, объем вяжущего, возможность использования различных добавок, повышающих технологические параметры бетона. На уровне Череповецкого промышленного региона дальнейшие исследования позволят разработать технические рекомендации по эффективным составам бетонных смесей из существующих вторичных продуктов промышленных производств, которые не только расширят номенклатуру шлаковых бетонов, но и позволят одновременно решать вопросы по снижению экологической нагрузки региона.

Литература

1. *Вознесенский, В.А.* Современные методы оптимизации композиционных материалов / В.А. Вознесенский, В.М. Выровой, В.Я. Керш и др. – Киев, 1983.

2. *Волкова, С.Б.* Математическая статистика и планирование эксперимента / С.Б. Волкова. – Череповец, 2009.

3. *Грызлов, В.С.* Шлакобетоны в крупнопанельном домостроении / В.С. Грызлов // Строительные материалы. – 2011. – № 3. – С. 40 – 41.

4. *Грызлов, В.С.* Рекомендации по использованию вторичных ресурсов металлургической, химической промышленности в строительстве / В.С. Грызлов, А.Г. Каптюшина, А.И. Фоменко. – Череповец, 1989.

5. *Дворкин, Л.И.* Проектирование составов бетона с заданными свойствами / Л.И. Дворкин, О.Л. Дворкин. – Ровно, 1999.

УДК 681.3.016

А.Н. Швецов, А.М. Куртасов

СЕМАНТИЧЕСКИЙ ПОДХОД К ЗАДАЧЕ ГЕНЕРАЦИИ УЧЕБНЫХ ТЕСТОВ НА ОСНОВЕ ИЗВЛЕЧЕНИЯ ЗНАНИЙ ИЗ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕКСТОВ

В статье предлагается подход к задаче генерации учебных тестов, основанный на технологии извлечения знаний из естественно-языковых текстов. Рассматривается проблема семантического анализа обрабатываемых текстов.

Учебные тесты, генерация, извлечение знаний, естественно-языковые тексты, семантика.

The paper proposes the approach to the problem of educational tests generation, based on the technology of knowledge acquisition from natural-language texts. The problem of semantic analysis of processed texts is considered.

Educational tests, generation, knowledge acquisition, natural-language texts, semantics.

В ходе внедрения информационных технологий во многих образовательных учреждениях России сложилась практика создания электронных учебно-методических комплексов дисциплин (ЭУМКД), т. е. наборов учебно-методических материалов по определенной учебной дисциплине или учебному курсу, размещаемых на электронных носителях. С целью повышения эффективности применения ЭУМКД в них часто интегрируются средства оценки знаний учащихся, что позволяет осуществлять контроль успеваемости параллельно с выдачей учебных материалов. Наиболее распространенной формой контроля в ЭУМКД является тестирование.

В настоящее время на рынке существует большое количество программных продуктов, предназначенных для компьютерной реализации учебного тестирования: *UniTest System*, *AUGUST 4*, «СИнТеЗ», «Прометей», *Moodle* и др. [1]. Все эти продукты обеспечивают широкие возможности для проведения тестирования и оценки результатов. Однако наиболее важная и сложная для выполнения задача — композиция тестовых заданий — до сих пор остается слабо автоматизированной. Тестовые вопросы и варианты ответов создаются вручную, и этот процесс отличается большой трудоемкостью.

В данной статье развивается авторский подход к автоматизации рассматриваемой задачи, основанный на извлечении знаний из естественно-языковых текстов. В качестве таких текстов могут использоваться пособия и учебники по различным учебным дисциплинам. Исходные принципы подхода были заложены в нескольких ранее опубликованных работах [2], [5]. Основная идея подхода заключается в том, что из составляющих текст утвердительных предложений можно автоматически генерировать вопросы, которые затем будут отбираться, корректироваться и редактироваться преподавателем.

С целью практической апробации подхода в настоящее время разрабатывается программный продукт, позволяющий формировать тестовые задания для последующей их передачи в *LMS*-систему *Moodle* (http://moodle.org/), которая используется в качестве основного средства дистанционного обучения в Вологодском государственном техническом университете.

Основными средствами, реализующими данный

подход, являются лингвистические процессоры, которые друг за другом обрабатывают входной текст. Вход одного процессора является выходом другого. Процессоры выполняют анализ текста на разных уровнях языка:

– графематический анализ (выделение предложений, слов, чисел, формул и т. д.);

– морфологический анализ (построение морфологической интерпретации слов входного текста);

– синтаксический анализ (построение дерева зависимостей предложения);

– семантический анализ (построение семантического графа текста).

Необходимой предварительной процедурой для дальнейшего анализа текста в любой системе анализа естественных языков является выделение предложений из сплошного текста. В первом приближении предложение есть последовательность символов, заканчивающаяся на символы «.», «!» или «?», однако на практике следует учитывать возможность использования точки в качестве символа сокращения и другие нюансы [7]. В разработанном графематическом алгоритме используются предопределенные наборы общепринятых («г.», «гг.», «и т.д.») и распространенных («т.к.», «т.е.», «т.н.») сокращений, а также учитывается возможность сокращения инициалов в именах собственных («А.С. Пушкин» и т.п.). В результате работы алгоритма получается массив предложений (фраз), которые в дальнейшем могут обрабатываться алгоритмами морфологического и синтаксического анализа.

Для получения тестовых заданий различных видов в настоящем подходе применяются различные виды алгоритмов синтеза вопросов. Эти алгоритмы отличаются глубиной анализа естественного языка, и их можно разделить на две группы.

Алгоритмы первой группы осуществляют замену искомого слова в предложении на комбинацию символов «?» (по каждому предложению исходного текста может быть построено не более одного вопроса). К данной группе относятся следующие алгоритмы: поиск сокращений (аббревиатур), поиск численных значений, генерация на основе определений, генерация из конструкций «если …, то …».

Алгоритмы этой группы основываются на тривиальном просмотре фразы и поиске необходимых символов (или же конкретных слов). Эти алгоритмы наиболее просто реализуются на практике и отличаются относительно высоким быстродействием, однако часто возвращают неприемлемые результаты.

Алгоритмы второй группы выполняют построение вопроса по результатам синтаксического анализа текста. Сюда относятся следующие: вопросы к подлежащему (*что?*, *кто?*, …), вопросы к прилагательным (*какой?*, *какая?*, …), вопросы к обстоятельству места (*где?*), вопросы к обстоятельству времени (*когда?*). Эти алгоритмы требуют наличия развитого морфологического словаря. В используемом для исследований программном продукте они реализованы с помощью библиотек *RML* (http://www.aot.ru/).

Для примера опишем алгоритм формирования вопроса к прилагательному. Блок-схема этого алгоритма представлена на рис. 1. В начале работы инициализируются переменные *A*, *B*, *C*: *A* – обрабатываемое предложение из массива текста; *B* – ответ на вопрос (присваивается пустая строка); *C* – готовое предложение для тестирования (по умолчанию равно *A*). Затем с помощью метода *FindSituation()* библиотек *RML* производится синтаксический анализ предложения *A*. Далее инициализируются вспомогательные переменные для работы алгоритма: *flag = истина* (сигнализатор найденной фразы для генерации вопроса), $j = 0$ (переменная цикла, номер текущего узла предложения). В переменную *K* записывается число узлов в анализируемом предложении. Затем начинается цикл с предусловием: пока *j* меньше либо равно ($K-1$) и *flag = истина*. В цикле последовательно разбирается каждый синтаксический узел предложения. Если в узле находится тип отношения «свойство», то с помощью метода *GetGramInfo()* находятся грамматические характеристики зависимого слова. Если данное слово представляет собой прилагательное, то оно записывается в переменную *B*. Наконец, формируется вопрос с учетом формы прилагательного (его характеристик, получаемых от указанного выше метода). Для закрытия цикла переменной *flag* присваивается значение *ложь*.

Следует отметить, что, если текст обрабатывается только до уровня синтаксиса без учета семантики, не все генерируемые вопросы могут являться релевантными в данной предметной области. Вследствие этого на пользователя может лечь довольно трудоемкая задача отбора вопросов. Тем не менее, испытания подхода на конкретных учебных дисциплинах показали, что даже при такой реализации подход дает относительно неплохие результаты. Например, в случае учебного пособия по дисциплине «Интеллектуальные информационные системы» (авт. А.Н. Швецов) с помощью алгоритмов второй группы удалось получить 40 % заданий, пригодных для использования в тесте без изменения, и 22,8 % заданий, из которых можно получить пригодные задания путем редактирования (т. е. больше половины полученных заданий оказались подходящими для составления тестов при той или иной степени участия преподавателя). Алгоритмы первой группы демонстрировали высокую полезность при анализе пособий, содержащих большое количество чисел, но в иных случаях часто оказывались малоэффективными (около 15 % пригодных заданий).

Для оценки временной эффективности алгоритмов они были испытаны на компьютерах различной аппаратной конфигурации (от нетбука до мощной рабочей станции на базе процессора *Intel Core i3*). Между алгоритмами первой и второй группы наблюдалась существенная разница: вследствие более высокой сложности алгоритмов второй группы для их работы часто требуется больше времени даже в случае меньшего размера пособия. Но в целом на всех конфигурациях генерация заданий заняла относительно мало времени: не более 2 мин. на загрузку пособия и не более 40 с на генерацию заданий тем или иным алгоритмом.

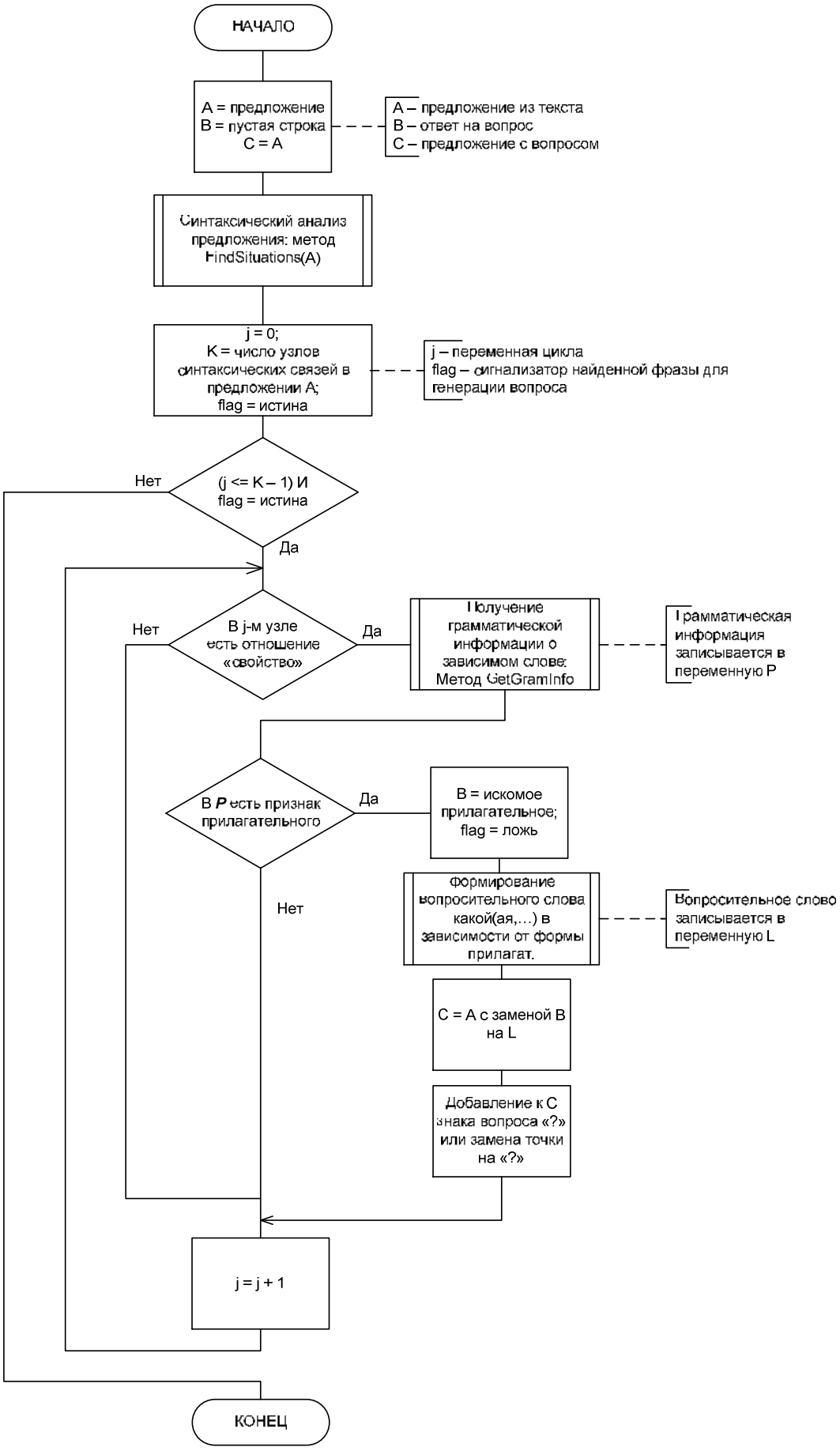


Рис. 1. Блок-схема алгоритма генерации вопросов к прилагательным

По результатам испытаний был сделан вывод, что эффективность генерации заданий будет изменяться от пособия к пособию и для достижения лучших результатов необходимо усовершенствовать подход таким образом, чтобы учитывалась семантика выделяемых из текста предложений [4].

Для перехода на уровень семантики в настоящем подходе предлагается формировать базу знаний на основе семантических сетей, которая позволяла бы определять принадлежность текстов к той или иной предметной области и, соответственно, выбирать тексты для генерации тестов по той или иной учебной дисциплине. По мнению авторов, большой потенциал при реализации в современных интеллектуальных системах имеет логический подход к семантическому анализу естественных языков, который интенсивно исследуется в работах лингвистов и логиков.

При создании моделей и методов семантического анализа могут использоваться различные логические формализмы: семантика смысла и денотата Г. Фреге, теория объектов и пропозиций Б. Рассела, теория истины А. Тарского, семантика возможных миров С. Крипке, теория типов Б. Рассела и К. Айдукевича. Новое направление в этой области, получившее название формальной семантики, сформировали работы Р. Монтегю. Основная идея данного направления выражена в названии одного из его основополагающих трудов – *English as a formal language* [6]. Любой естественный язык предлагается понимать как формальный логический язык, который является более сложным по отношению к существующим формальным языкам. При описании естественного языка предлагается использовать такие же понятия и конструкции, как для других логических языков.

Логика Р. Монтегю является основой метода семантической обработки информации, который предлагается использовать в настоящем подходе [3]. Процесс обработки текстовой информации по данному методу представляется следующим образом. Входные данные для системы – это текст на естественном языке. На основе лингвистической обработки текста строится набор категорий интенсиональной логики для дальнейшего применения правил трансформации (ПТ) синтаксических конструкций в элементы единой формулы, отражающей смысл высказывания. Формальное представление не зависит от конкретного естественного языка и представляет собой набор типов и операций над формулами. Результатом обработки является формализованное представление смысла текста в виде набора формул, отражающих смысл предложений и множества постулатов значений, представляющих фоновые знания о мире. Метод заключается в применении алгоритмов формализации смысла естественно-языковых текстов, заполнении базы знаний и интерпретации на ней запросов, формируемых при генерации тестовых заданий.

В методе выделяются следующие этапы:

1) формализация естественно-языковой фразы,

2) интерпретация формулы формальной семантики,

3) заполнение базы знаний.

При формализации естественно-языковой фразы сначала проводится лингвистическая обработка фразы языковым процессором с целью построения синтаксического дерева. Далее осуществляется рекурентный обход узлов дерева с целью применения функции генерации формулы на основе ПТ. Каждому узлу дерева сопоставляется некоторая категория интенсиональной логики. Для определения категории узла используется табличная функция отображения категорий синтаксического анализатора в категории интенсиональной логики.

Под интерпретацией формулы в формальной семантике понимается установление ее истинностного значения на основе представленной теоретико-множественной картины мира. Для корректного применения функции интерпретации в технической системе представляется необходимым разделение данных о состоянии мира и процедур интерпретации фрагментов языка логики.

Для реализации механизма заполнения базы знаний предлагается расширить толкование термина «интерпретация», используемого в формальной семантике. Под интерпретацией здесь понимается не просто вычисление истинностного значения формулы, но и отображение знакового представления выражения на определенную картину мира. При этом может проводиться как интерпретация с целью вычисления выражения, так и интерпретация для изменения модели предметной области, на которой производится отображение. Для заполнения базы знаний на основе данных текста строится формальное представление естественно-языкового текста в виде множества набора формул интенсиональной логики. Затем определяется целевая семантическая сеть и осуществляется ее клонирование, результатом которого является семантическая сеть, имеющая пометы на всех узлах, показывающие ее принадлежность к определенному тексту. Далее происходит интерпретация каждой формулы с целью нанесения признаков объектов, представленных в формуле, на семантическую сеть.

Литература

1. *Башмаков, А.И.* Разработка компьютерных учебников и обучающих систем / А. И. Башмаков, И. А. Башмаков. – М., 2003.

2. *Воронец, И.В.* Универсальная автоматизированная система тестирования знаний и самообучения, основанная на анализе естественно-языковых текстов учебных пособий / И.В. Воронец, А.Н. Швецов, В.С. Алешин // Пилотируемые полеты в космос: сб. докл. V международ. научн.-практ. конф. (9 – 10 апреля 2003 г.). – Звездный городок Моск. обл., 2003. – С. 65 – 67.

3. *Летовальцев, В. И.* Программная формализация естественного языка средствами формальной семантики / В.И. Летовальцев, А.Н. Швецов // Программные продукты и системы. – 2010. – № 3. – С. 85 – 90.

4. Методология создания агентно-ориентированных учебных комплексов для подготовки специалистов технического профиля: отчет о НИР (промежуточ.) / Вологодский государственный технический университет; рук. А.Н.

Швецов; исполн.: В.А. Горбунов [и др.]. – М., 2011. – № ГР 01201056386.

5. *Швецов, А.Н.* Построение приближенной концептуальной модели предметной области на основе анализа смысла естественно-языковых текстов / А.Н. Швецов, В.С. Алешин // Международная конференция по мягким вычислениям и измерениям SCM'2003: сб. докл. Т. 2. – СПб., 2003. – С. 120 – 123.

6. *Montague, R.* English as a formal language / R. Montague, edited by R. H. Thomason // Formal Philosophy. – Yale University Press, 1974.

7. *Riley, Michael D.* Some applications of tree-based modeling to speech and language indexing / Michael D. Riley // Proceedings of the DARPA Speech and Natural Language Workshop. – Stroudsburg, PA, USA, 1989. – P. 339 – 352.

УДК 621. 778.04.

Р.А. Юдин, И.Р. Юдин

ОСОБЕННОСТИ ГОРЕНИЯ ТОПЛИВА ПРОИЗВОЛЬНОГО ХИМИЧЕСКОГО СОСТАВА

В статье приведены физико-химические реакции горения однокомпонентных газов и горючих элементов в воздухе при стехиометрических условиях. На их основе получено аналитическое уравнение горения в воздухе топлива произвольного химического состава, позволяющее рассчитывать важнейшие теплотехнические параметры горения, например, объем воздуха, который необходим для полного сгорания единицы горючего вещества.

Углерод, водород, сера, алканы, алкены, алкины, физико-химические реакции горения топлива.

Physico-chemical reactions of combustion of one-component gases and flammable elements in the air at stoichiometric conditions are considered in the article. An analytical equation of fuel combustion of random chemical composotion in the air is got on their basis allowing to count the significant heat engineering burning parameters, for example: volume of the air, which is necessary for the full combustion of the unit of combustible substance.

Carbon, hydrogen, sulfur, alkanes, alkenes, alkynes, physico-chemical reactions of fuel combustion.

Горючими составляющими топлива являются: углерод, сера, водород, входящие в твердое топливо, оксид углерода, молекулярный водород, сероводород, а также жидкие и газообразные углеводороды. Особый практический интерес представляет углеводородное газообразное топливо, так как им наиболее часто отапливают теплотехнические установки различного технологического назначения. Это топливо содержит газообразные составляющие трех гомологических рядов – алканы, алкены и алкины. Алканы являются насыщенными углеводородами и относятся к гомологическому ряду метана (CH_4). Общая химическая формула ряда алканов – C_nH_{2n+2}. Алкены – ненасыщенные углеводороды, гомологи ряда этилена (C_2H_4). В структуре молекул алкенов имеется двойная физико-химическая связь между атомами углерода. Общая химическая формула ряда алкенов – C_nH_{2n}. К алкинам относятся ацетилен (C_2H_2) и бензол (C_6H_6) с общей химической формулой – C_nH_n. Алкины также являются ненасыщенными углеводородами. В структуре молекулы ацетилена имеется одна тройная физико-химическая связь между атомами углерода, а в бензольном кольце между атомами углерода имеются три двойные физико-химические связи.

Химические реакции горения топлива в воздухе для стехиометрических условий и при избытках воздуха ($\alpha \geq 1$) хорошо изучены и широко известны. Однако для углеводородных газов имеется только одно математическое обобщение химических реакций при сжигании этих газов в кислороде, представленное в монографии А.В. Арсеева [1, с. 51], следующей формулой:

$$C_mH_n + (m + n/4)\, O_2 = m\, CO_2 + n/2\, H_2O. \quad (1)$$

Известно, что в объемных процентах азота в воздухе в 3,76 раза больше, чем кислорода. Как правило, азот не участвует в процессе горения, а присутствует в начальных и конечных продуктах реакции как балласт. Отсюда очевидно, что при сжигании любого углеводородного газа в воздухе формула (1) примет следующий вид:

$$C_mH_n + (m + n/4)O_2 + 3{,}76(m + n/4)N_2 = nCO_2 + n/2\, H_2O + 3{,}76(m + n/4)N_2. \quad (2)$$

С одной стороны, химические уравнения (1 – 2) компактны, абсолютно справедливы для всех гомологических рядов и приемлемы для использования в теплотехнических расчетах. Однако они имеют формальный характер, так как в них не отражен дифференцированный вклад алканов, алкенов и алкинов в совокупный процесс горения, поэтому делать однозначный вывод о том, что компактность формул обуславливает уменьшение объема вычислений, по меньшей мере, преждевременно.

Кроме того, во-первых, в такой записи формулы (1), (2) неоправданно усложнены тем, что содержат два переменных коэффициента: n, m. Однако все три гомологических ряда углеводородных газов содержат только один переменный коэффициент – n. Во-вторых, в конкретных теплотехнических расчетах необходимо вычислять только отдельные составляющие реакций горения, например, стехиометрические объемы кислорода и воздуха, коэффициент расхода воздуха (α_Σ). Для этого необходимо исполь-