

РАЗРАБОТКА ПРИЛОЖЕНИЯ ПО ГЕНЕРАЦИИ УЧЕБНЫХ ЗАДАНИЙ К ТЕКСТУ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ НА ОСНОВЕ ПОРОЖДАЕМЫХ ШАБЛОНОВ

Личаргин Д. В., Усова А. А., Сотникова В. В., Липман С.А., Бутовченко В. В.

ФГАОУ ВПО «Сибирский федеральный университет», Красноярск, Россия (660074, Красноярск, ул. Киренского, 26), e-mail: lichdv@hotmail.ru

Цель работы состоит в необходимости сформулировать некоторые принципы генерации учебных заданий, которые могут быть использованы при создании тестов для проверки знаний, в целях повышения эффективности процесса составления учебных материалов. Рассматривается проблема усовершенствования метода генерации учебных заданий на основе многомерного анализа семантических данных за счет привлечения алгоритма автоматической генерации используемых шаблонов. Проблема разработки все больших объемов учебных материалов является актуальной в связи с внедрением творческих, гибких и индивидуальных траекторий обучения. Для решения этой проблемы предлагается применить метод многомерного анализа семантически векторизованных данных. Успешно разрабатывается программа анализа и визуального представления семантики текста, а также автоматической генерации учебных материалов на его основе. Делается вывод о необходимости расширения этого метода с использованием генерируемых семантических шаблонов.

Ключевые слова: компьютерная лингвистика, генерация учебных заданий, генерация осмысленной речи.

APPLICATION DEVELOPMENT FOR THE GENERATION OF EDUCATIONAL TASKS FOR TEXTS IN NATURAL LANGUAGE BASED ON GENERATED PATTERNS

Lichargin D. V., Usova A. A., Sotnikova V. V., Lipman S. A., Butovchenko V. V.

FSAEI HPE "Siberian Federal University", Krasnoyarsk, Russia (660074, Krasnoyarsk, ul. Kirenskogo, 26), e-mail: lichdv@hotmail.ru

The purpose of the work is reduced to the need to formulate some principles of generating the learning tasks that can be used for creating educational tests to check the knowledge in order to improve the process of composing the educational materials. The problem of improving the method of generating the educational tasks based on multi-dimensional analysis of semantic data is considered by evolving an algorithm of automatic generation of templates being used. The problem of the development of ever larger amounts of learning materials is topical regarding the implementation of creative, flexible and individual learning paths. For solving this problem it is proposed to apply the method of multidimensional analysis of semantically vectorized data. Software for text analysis and visual representation of the text semantics, as well as automatic generation of training materials based on it is successfully developed. A conclusion is made about the need to expand this method using generated semantic templates.

Keywords: computational linguistics, generation of educational tasks, generation of meaningful speech.

В работе рассматривается проблема усовершенствования метода генерации учебных заданий на основе многомерного анализа семантических данных за счет привлечения алгоритма автоматической генерации используемых шаблонов.

На сегодняшний день широко распространены и разрабатываются разнообразные системы обучения иностранному языку.

Проблема разработки все больших объемов учебных материалов является актуальной в связи с внедрением творческих, гибких и индивидуальных траекторий обучения.

Проблема решается на стыке таких наук, как дискретная математика, математическая логика, теория формальных языков и грамматик.

Цель работы состоит в необходимости сформулировать некоторые принципы генерации учебных заданий, которые могут быть использованы при создании тестов для проверки знаний, в целях повышения эффективности процесса составления учебных материалов.

Задачи данной работы следующие:

- 1) описание принципа гибридизации многомерного представления семантических данных и порождающих грамматик;
- 2) разработка модуля порождающих грамматик для порождения шаблонов генерации;
- 3) разработка базы правил этой порождающей грамматики;

Основная идея работы состоит в расширении функциональности программы «Генератор учебных заданий», использующей многомерный анализ данных для анализа текста на английском языке и генерация тестовых материалов, за счет привлечения средств автоматической генерации специальных шаблонов, ссылающихся на разделы многомерной базы данных.

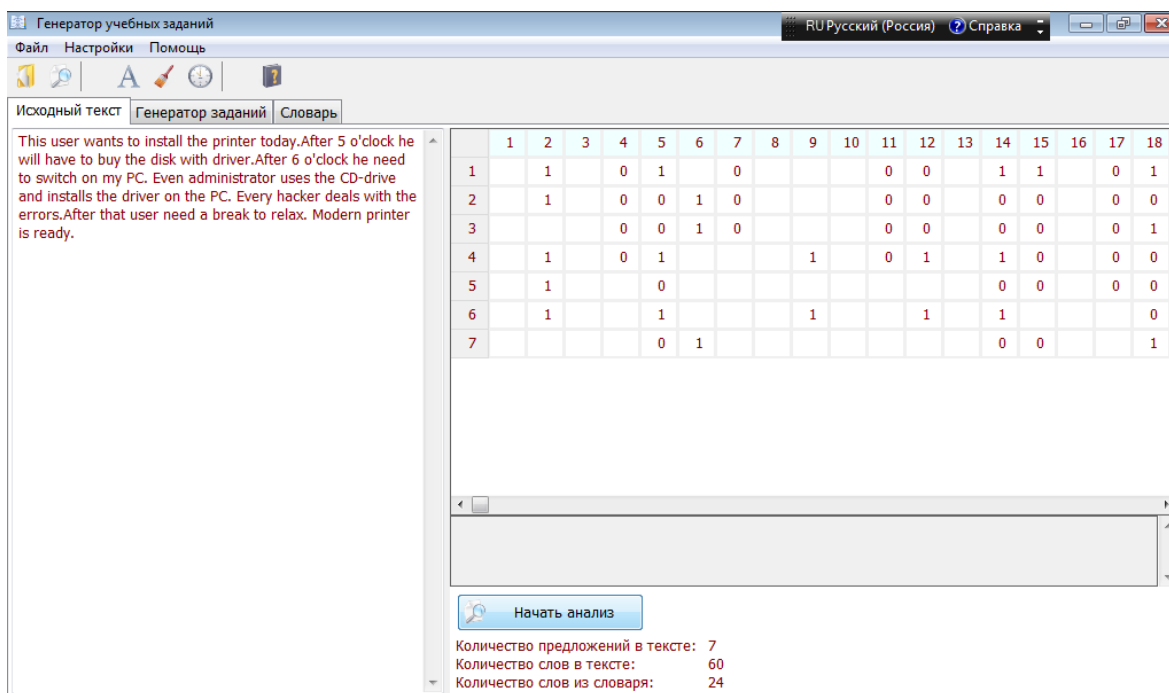


Рис.1. Главное окно программы с загруженным текстом и его визуализацией

Новизна данной работы состоит в гибридизации метода многомерного анализа данных, согласно анализу текста на английском языке с методом порождающих грамматик, применительно к генерации порождаемых шаблонов, ссылающихся на разделы многомерной базы данных, в целях генерации более разнообразных учебных материалов.

На сегодняшний день широко распространены и разрабатываются различные программы для обучения иностранному языку [7-9]. В связи с этим, генерация учебных заданий является актуальной проблемой.

Создание тестов для проверки знаний – это очень трудоемкий процесс, требующий много времени, так как задания составляются преподавателем самостоятельно, а также тесты должны меняться, чтобы избежать списывания со стороны обучающихся.

Появление IT-технологий привело преподавателей к попыткам автоматизировать процесс создания многовариантных заданий для своих курсов.

Целью данной работы является разработка приложения по генерации учебных заданий на основе порождаемых шаблонов на английском языке.

Проблема генерации осмысленных подмножеств естественного [1-6,10] языка является важной для решения проблемы автоматической генерации учебных заданий по иностранному языку.

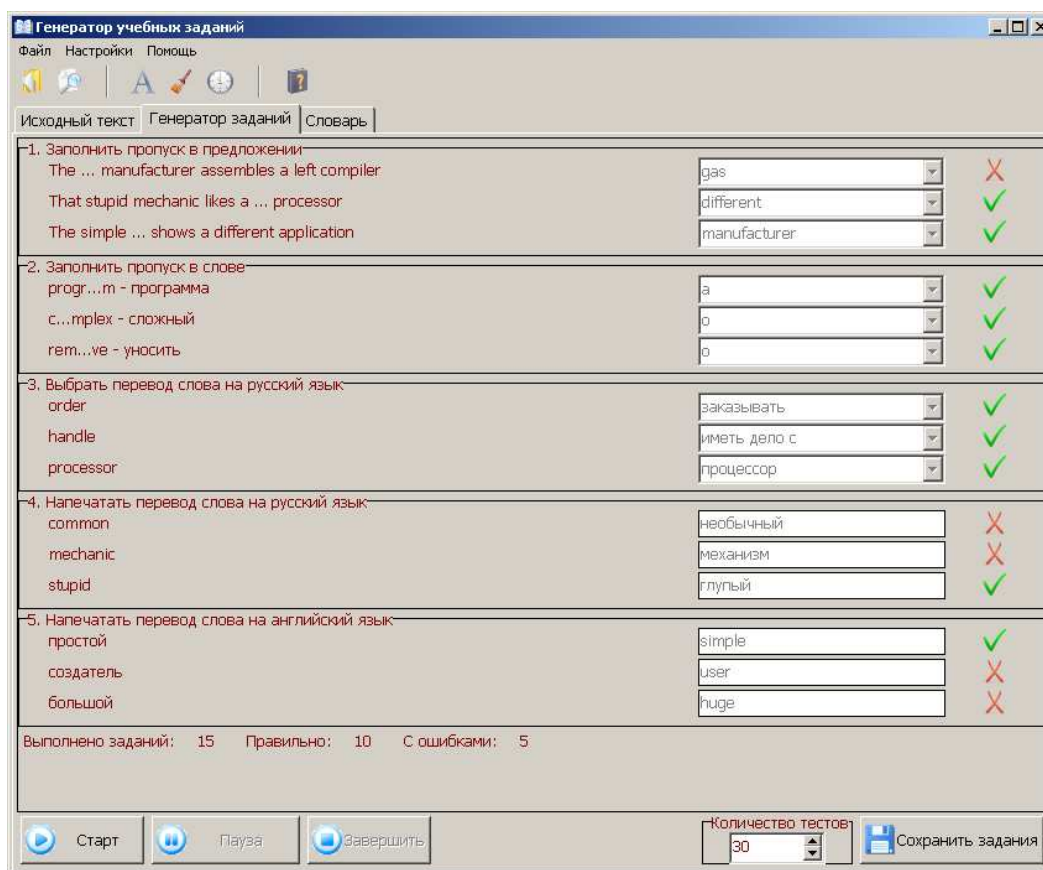


Рис. 2. Программа после выполнения проверки заданий

Коллективом авторов была разработана программа «Генератор учебных заданий». «Генератор учебных заданий» – это программный продукт, который обладает дружелюбным интерфейсом и ориентирован на пользователей, обладающих различным уровнем работы с компьютером. Данная программа получает в качестве входного параметра строку

произвольного текста на английском языке с целью проанализировать его и обеспечить генерацию учебных заданий. Программа позволяет визуализировать структуру текста (см. рис. 1).

В программу «Генератор учебных заданий» требуется внедрить порождаемые шаблоны, которые позволят сделать независимой систему порождения осмысленных фраз и заданий (см. рис. 2), как основы генерируемых учебных заданий.

Тестовое задание 013 (от 2015-09-16)					
1. Заполнить пропуск в предложении					
The stupid programmer verifies a following ...	<input type="checkbox"/> application	<input type="checkbox"/> chicken			
	<input type="checkbox"/> coco	<input type="checkbox"/> turkey			
That ... creator uses a huge application	<input type="checkbox"/> solid	<input type="checkbox"/> talented			
	<input type="checkbox"/> wrong	<input type="checkbox"/> analogue			
The able mechanic connects an ... thing	<input type="checkbox"/> sour	<input type="checkbox"/> sweet			
	<input type="checkbox"/> analogue	<input type="checkbox"/> salty			
2. Заполнить пропуск в слове					
comm__n - обычный	<input type="checkbox"/> a	<input type="checkbox"/> e	<input type="checkbox"/> i	<input type="checkbox"/> o	<input type="checkbox"/> u
r__gister - регистр	<input type="checkbox"/> a	<input type="checkbox"/> e	<input type="checkbox"/> i	<input type="checkbox"/> o	<input type="checkbox"/> u
ord__r - заказывать	<input type="checkbox"/> a	<input type="checkbox"/> e	<input type="checkbox"/> i	<input type="checkbox"/> o	<input type="checkbox"/> u
3. Выбрать перевод слова на русский язык					
compiler	<input type="checkbox"/> компилятор	<input type="checkbox"/> наладонник			
	<input type="checkbox"/> катод	<input type="checkbox"/> диод			
huge	<input type="checkbox"/> огромный	<input type="checkbox"/> цвета морской волны			
	<input type="checkbox"/> ярко-красный	<input type="checkbox"/> большой			
complex	<input type="checkbox"/> прозрачный	<input type="checkbox"/> обычный			
	<input type="checkbox"/> сложный	<input type="checkbox"/> хаки			
4. Напишите перевод слова на русский язык					
mechanic					
element					
stupid					
5. Напишите перевод слова на английский язык					
создатель					
обычный					
программист					

Рис. 3. Сохранение сгенерированных заданий

Порождаемые шаблоны – шаблоны построения фраз естественного языка, которые могут быть заданы, как вручную, так и порождаться на основе, в частности, контекстно-свободных порождающих грамматик.

Они позволяют сделать систему независимой от способа создания, композиции и представления объектов.

Action + -ing + Link + -(e)s + Process + (Manner);

Action + -ing + the + Object + Link + -(e)s + Process + (Manner);

Action + -ing + was + Done + Group of Time;

Surely, + Action + -ing + was + Done + Group of Time.

Маркеры типа Action, Object обращаются к соответствующим ячейкам или группам ячеек условно трехмерного пространства данных с учетом текущей темы генерации учебных заданий.

На основе сгенерированных порождающей грамматикой шаблонов [5] модифицированная версия программы «Генератор учебных заданий» должна выдавать фразы следующего вида.

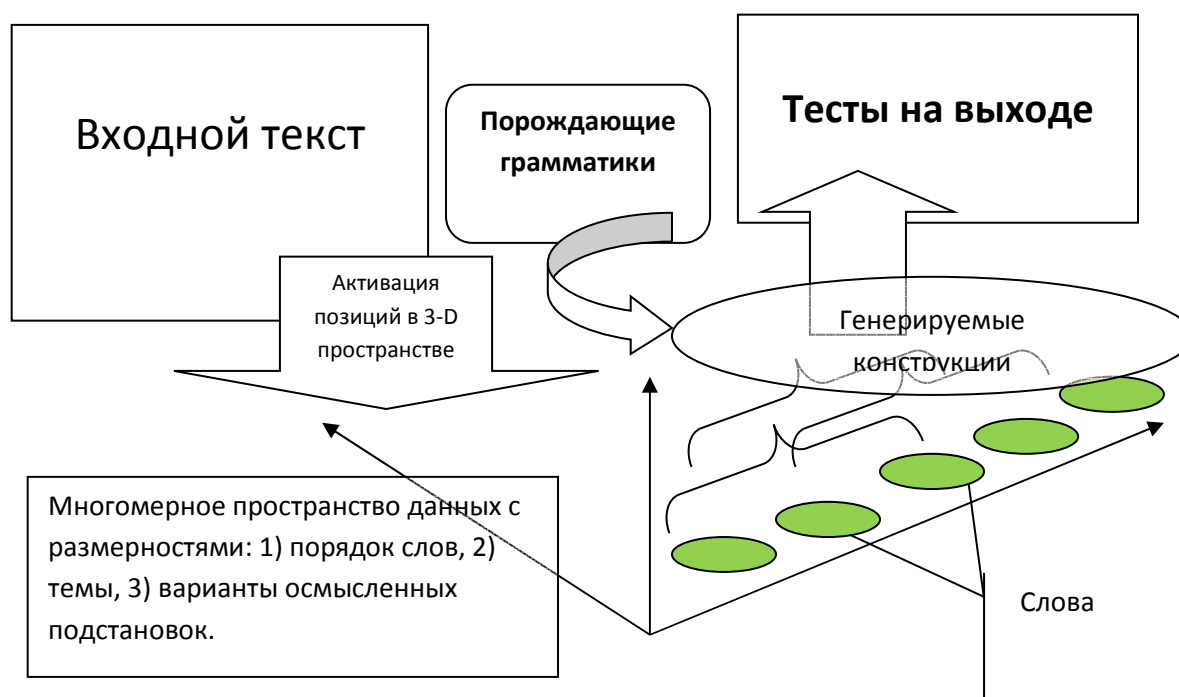


Рис. 4. Схема генерации осмысленных учебных заданий к текстам

«Eating the cake implies good appetite. Building means accuracy. Driving the bus presupposes interruptions. Driving the bus was carried out this morning. Obviously, knitting the coat was performed on Monday».

В качестве одного из этапов в направлении интерпретации текста на естественном языке необходимо осуществить проекцию слов каждого предложения такого текста с учетом реализуемой (в качестве побочного) эффекта экстракции семантического шума. На этом этапе необходимо приведение фраз текста к фразам приведенного вида с определенной

вероятностью искажения смысла этих фраз, что технически не влияет на процесс генерации учебных заданий.

Так, например, входной текст «I am a programmer, preferring C++. C# is considered by some experts to be better...» подается на вход программы, после обработки предложений в многомерной базе данных активируются цепочки вида: «Programmer + (I) + program(s) in + C++. Some + expert(s) + consider + that + (it is) better (for) + programmers + (to) use + C#» на основе порядка слов приведенного вида: «Кто-то1 + думает/сообщает + (кому-то) + что + кто-то2 + склонен + иметь дело с + чем-то + где-то + когда-то + на основании какой-то информации + ...», где для темы «Еда» это будут последовательностью вида: «... повар хочет готовить пиццу/курицу/котлеты ...», а, например, для темы «Здания» – последовательностью вида: «...строитель может строить библиотеку/стадион/театр ...». Ссылки порождающих грамматик относятся к вариантам выбора в многомерной базе данных, словам «комбинонимам», то есть словам, которые при замене друг друга в нужном контексте меняют смысл фраз, но сохраняют осмысленность: «(Я) + хочу/могу/пытаюсь/умею/должен + (приготовить + блюдо)». Далее на основе выборки из этих цепочек строятся учебные задания вида: Выберите правильный вариант подстановки: «is it worthy to program in C# / library / cake?».

Семантический шум представляет собой модель одного из проявлений естественного языка – творческого иносказания и трансформаций фраз человеком на основе определенного диапазона возможностей языка. Формально отвлекаясь от типов связей внутри фраз с семантическим шумом можно породить одну базовую интерпретацию текста на основе порядка терминов в многомерной семантической базе данных (см. рис. 4). Множество фраз неприведенного вида может быть получено из данной формы высказываний (функций над ячейками многомерной базы данных), полученные высказывания представляют собой иносказания стандартных фраз более или менее высокого художественного достоинства. На основе таких фраз строится множество учебных заданий, интегрируемых в различные варианты тестов (см. рис. 3).

Очевидно, что само смысловое значение, включающее все ассоциативные и эмоциональные компоненты, не может быть формализовано на основе простого вектора рационально представленных иррационально-смысловых категорий. Но в рамках определенных аппроксимаций и приближений возможно описывать семантику естественного языка до уровня достаточного для решения конкретных инженерных задач разной степени сложности. В перспективе на основе семантического понятийного аппарата имеется принципиальная, хотя и чрезвычайно сложная, задача прохождения теста Тьюринга лингвистическим программным обеспечением. При этом остается актуальной разработка

программных систем с более узкой функциональной направленностью, в частности, рассматриваемый программный продукт.

В рамках такого рода работы над узким набором задач необходимо развивать принципы генерации осмысленной речи на основе проработки отдельных подмножеств многомерной классификации слов и понятий языка с учетом возможности использования этих фрагментов для генерации подмножеств естественного языка. Необходима реализация этих принципов на основе генерируемых (порождающими грамматиками) семантических шаблонов с извлечением из их структуры семантического шума с частичными семантическими потерями.

Порождение шаблонов генерации осмысленной речи, ссылающихся на ячейки многомерной базы данных, может быть осуществлено при помощи порождающих грамматик и впоследствии уточняться на основе введения операций над векторами сем, когда будут вводиться запреты на подмножества комбинаций фраз, генерируемых на основе принципа семантических подстановок. Данная программа может позволить составлять различные варианты заданий одного типа. Множества вариантов генерируемых шаблонов позволят индивидуализировать задания, так как большое количество тестовых материалов необходимо для исключения списывания со стороны студентов и учеников. Также она позволит обеспечить экономию времени на разработку образовательных ресурсов преподавателям.

В статье рассматривается проблема гибридизации принципа генерации осмысленных заданий и фраз естественного языка на основе многомерных баз данных и принципа генерации проецируемых на неё шаблонов фраз, на основе контекстно-свободных порождающих грамматик, и реализация этого подхода при разработке и усовершенствовании программы «Генератор учебных заданий», которая будет способствовать автоматизации работы преподавателей по разработке электронных образовательных ресурсов, а также позволит ускорить процесс разработки тестов на основе печатных материалов.

Список литературы

1. Личаргин Д. В. Многомерное представление данных по лексике и грамматике английского языка / Сибирский федеральный университет. Красноярск, 2010.
2. Личаргин Д. В., Бачурина Е. П. Обобщенная иерархическая структура учебного электронного курса и рассмотрение на ее основе электронных курсов обучения английскому языку // Информатизация образования и науки. – 2012. – № 15. – С. 20-36.

3. Личаргин Д. В., Суманеева Я. А., Юрьева Е. В. Метод подстановочных таблиц и его применение в сфере обучения русскому языку для иностранцев // Вестник Сургутского государственного педагогического университета. – 2012. – № 6 (21). – С. 179-187.
4. Личаргин Д. В., Таранчук Е. А. Иерархическая структура учебного электронного курса и ее вариабельность для обучения иностранному языку // Дистанционное и виртуальное обучение. – 2011. – № 4. – С. 56-75.
5. Сафонов К. В., Личаргин Д. В. Некоторые принципы автоматической генерации учебных материалов на основе баз знаний и лингвистической классификации // Вестник Сибирского государственного аэрокосмического университета им. академика М. Ф. Решетнева. – 2012. – №2 (42). – С. 72-77.
6. Сафонов К. В., Личаргин Д. В. Разработка векторизованной семантической классификации над словами и понятиями естественного языка // Вестник Сибирского государственного аэрокосмического университета им. академика М. Ф. Решетнева. – 2009. – № 4 (25). – С. 33-37.
7. Goodfellow, R. CALL Programs for Vocabulary Instruction / R. Goodfellow // Computer Assisted Language Learning Journal, Vol. 8. No. 2, 1995. – p. 205-226.
8. Ingraham, B. Language Training for Various Purposes in Several Languages on a Common Hypermedia Framework / B. Ingraham, T. Chanier, C. Emery // Computer & Education, Vol. 23, Iss. 1, 2, 1994. – p. 107-115.
9. Jones, G. Computer Simulations in Language Teaching – the Kingdom Experiment / G. Jones // System, Iss.17, 1986. – p. 35-47.
10. Safonov K.V., Lichargin D.V. Elaboration of a vector based semantic classification over the words and notions of the natural language // Вестник Сибирского государственного аэрокосмического университета им. академика М. Ф. Решетнева. – 2009. – № 5 (26). – С. 52-56.

Рецензенты:

Ченцов С.В., д.т.н., профессор, заведующий кафедрой Системы автоматизации, автоматизированного управления и проектирования Сибирского федерального университета, г. Красноярск;

Бронов С.А., д.т.н., профессор, руководитель научно-учебной лаборатории Систем автоматизированного проектирования кафедры Системы искусственного интеллекта Сибирского федерального университета, г. Красноярск.