электронный научно-технический журнал
ИНЖЕНЕРНЫЙ ВЕСТНИК
Издатель ФГБОУ ВПО "МГТУ им. Н.Э. Баумана". Эл No. ФС77-51036. ISSN 2307-0595

Подход к проектированию системы оценки лингвистических знаний

05, май 2015

Пугачёв Е. К.[1,*], Пьо С. Т.[1]

УДК: 004.891

[1]Россия, МГТУ им. Н.Э. Баумана
[*]iu6.pugachev@yandex.ru

Данная работа посвящена созданию интеллектуальной системы, которая позволяла бы оценивать лингвистические знания пользователей, изучающих иностранный язык. В настоящее время системный подход при выполнении проверки лингвистических знаний осуществляется человеком-экспертом.

В целом при изучении иностранного языка можно выделить следующие проблемы:

- большие затраты времени на освоение языка;
- большие затраты времени на тестирование обучающихся;
- отсутствие объективной оценки результатов во время тестирования;
- низкие возможности оценки качества освоение языка и др.

Освоение иностранного языка – это итерационный процесс, предполагающий прохождение трех этапов: обучение, тренировка и тестирование [3]. При этом для каждого этапа можно выделить следующие части:

1.Аудирование (восприятие на слух);

2. Чтение;

3. Грамматика;

4. Письмо (составление связного текста);

5.Говорение (разговорная речь).

В процессе исследования предметной области задачи было определено, что оценка лингвистических знаний зависит от уровня освоения иностранного языка. Можно выделить следующие уровни [1] :

1. Элементарный уровень;

2. Базовый уровень;

3.Первый сертификационный уровень;

4.Второй сертификационный уровень;

5.Третий сертификационный уровень;

6. Четвёртый сертификационный уровень.

Первые три уровня представляют начальный этап изучения языка, второй уровень - это продвинутый этап, а третий и четвёртый уровни для филологов.

В результате обзора существующих программных систем для решения задач, связанных с проверкой лингвистических знаний, было выявлено, что имеющиеся системы дают возможность тренировать, изучать и тестировать только в части «грамматика».

Создание интеллектуальной системы оценки лингвистических знаний (СОЛЗ), которая позволила бы взять на себя функции оценки знаний по различным частям с учетом уровня освоения, является актуальной задачей. Такая система позволит уменьшить время тестирования обучающихся, повысить объективность оценки знаний и качество освоения языка.

Для сознания интеллектуальной системы СОЛЗ необходимо выполнить ряд проектных операций, в которых должны быть учтены результаты различных видов декомпозиции предметной области задачи. В качестве примера на рисунке 1 представлены результаты процесса логической декомпозиции для оценки восприятия речи. Данные результаты необходимо использовать при построении концептуальной модели системы СОЛЗ, в частности подсистемы аудирования.

Рисунок 1 - Результаты логической декомпозиции для оценки восприятия речи.

Чтобы убедиться в правильности концептуальных решений при разработке интеллектуальной системы СОЛЗ воспользуемся положениями теории информационных семантических систем (ISS)[2]. Разрабатываемую систему можно отнести к ISS. Связано это с тем, что в СОЛЗ должна быть заложена разных видов семантическая информация (SI), а именно текстовая (SI_T), аудиальная (SI_S), визуальная (SI_G) и изобразительная (SI_C). Над семантической информацией в рамках ISS должен выполняться ряд семантических операций.

На рисунке 2 представлена структура ISS, которая показывает, какие семантические объекты (SO) и семантическая информация имеются в разрабатываемой системе СОЛЗ.

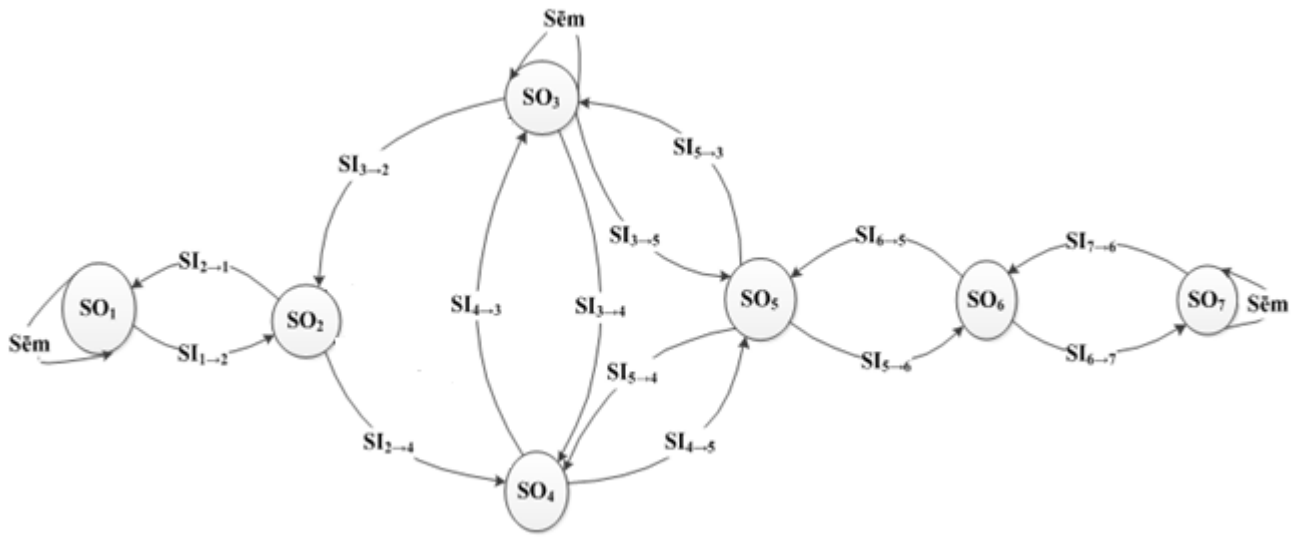

Рисунок 2 - Структура информационной семантической системы.

Каждый семантический объект, представленный на рисунке 2, можно сопоставить с объектом реального или виртуального мира:

SO_1– обучающийся, который сдает тесты;
SO_2– семантический интерфейс;
SO_3– модуль формирования тестов (МФТ);
SO_4 – модуль оценки знаний (МОЗ);
SO_5 – база знаний (БЗ);
SO_6 – модуль накопления знаний (МН);
SO_7 эксперт-лингвист.

Объекты SO_1и SO_7 это люди, а объекты с SO_2 по SO_6 представляют компоненты системы СОЛЗ. Каждый семантический объект в ISS может выполнять одну из следующих ролей: источника SI, приемника SI или посредника при передаче (преобразовании) SI.

При анализе семантической модели необходимо различать SO, которые генерируют SI, от SO, которые являются источниками SI. Также важно понимать, что многие SO могут выступать в рамках одной ISS в нескольких ролях, но основная роль только одна.

Определение ролей семантических объектов в ISS имеет важное значение на этапе принятия концептуальных решений по созданию системы СОЛЗ. Обусловлено это тем, что на основе информации, связанной с определением ролей можно оценить стабильность будущей интеллектуальной системы. Например, повышение количества SO-источников позволяет повысить стабильность, а включение дополнительных SO-посредников уменьшает стабильность. Включение семантического объекта SO_7 повышает стабильность ISS, но при этом целесообразно создать семантический объект SO_6.Если исключить семантический объектSO_6 из структуры ISS, то объект SO_7 должен иметь повышенную квалификацию. Если не предъявлять дополнительные требования к SO_7, то обмен семантической информацией будет невозможным, а следовательно, реализованная система СОЛЗ за короткий интервал времени потеряет свою актуальность. Другими словами, цель данной ISS– оценить лингвистические знания обучающегося с помощью семантической информации, заложенной в базе знаний системы. Если исключить возможность пополнения и изменения БЗ, то это может привести к невозможности функционирования системы СОЛЗ.

Кроме вышесказанного, на основе анализа семантической модели можно определить формы и сложность SI. В представленной на рисунке 2 ISS, при взаимодействии SO используется следующая семантическая информация:

$SI_{1\to2}$– ответы на естественном языке;
$SI_{2\to1}$– тесты на естественном языке;
$SI_{3\to2}$– тесты на формализованном языке;
$SI_{2\to4}$– ответы на формализованном языке;
$SI_{3\to4}$– информация о сформированных тестах;
$SI_{4\to3}$– результаты оценки;
$SI_{4\to5}$– запросы информации для проверки знаний;
$SI_{5\to4}$– информации для проверки знаний;
$SI_{5\to3}$– вопросники тестов;
$SI_{3\to5}$– запросы информации для формирования тестов;
$SI_{5\to6}$– заложенная в БЗ информация на формализованном языке;
$SI_{6\to5}$– новые знания на формализованном языке;
$SI_{6\to7}$– заложенная в БЗ информация на естественном языке;
$SI_{7\to6}$– новые знания на естественном языке.

Основными семантическими операциями для интеллектуальной системы СОЛЗ являются: преобразование SI из естественного представления в формализованное и наоборот, извлечение SI, сжатие SI, преобразование первичной SI во вторичную и др.

Кроме семантической информации и операций, связанных с ней, в системе СОЛЗ необходимо хранить и обрабатывать обычные данные. Основные действия, которые должна выполнять система СОЛЗ, представлены на рисунке 3 в виде обобщенной схемы алгоритма работы. Из данной схемы видно, что для каждого обучаемого формируется план тестирования, определяется уровень тестирования и предметная область и другие действия.

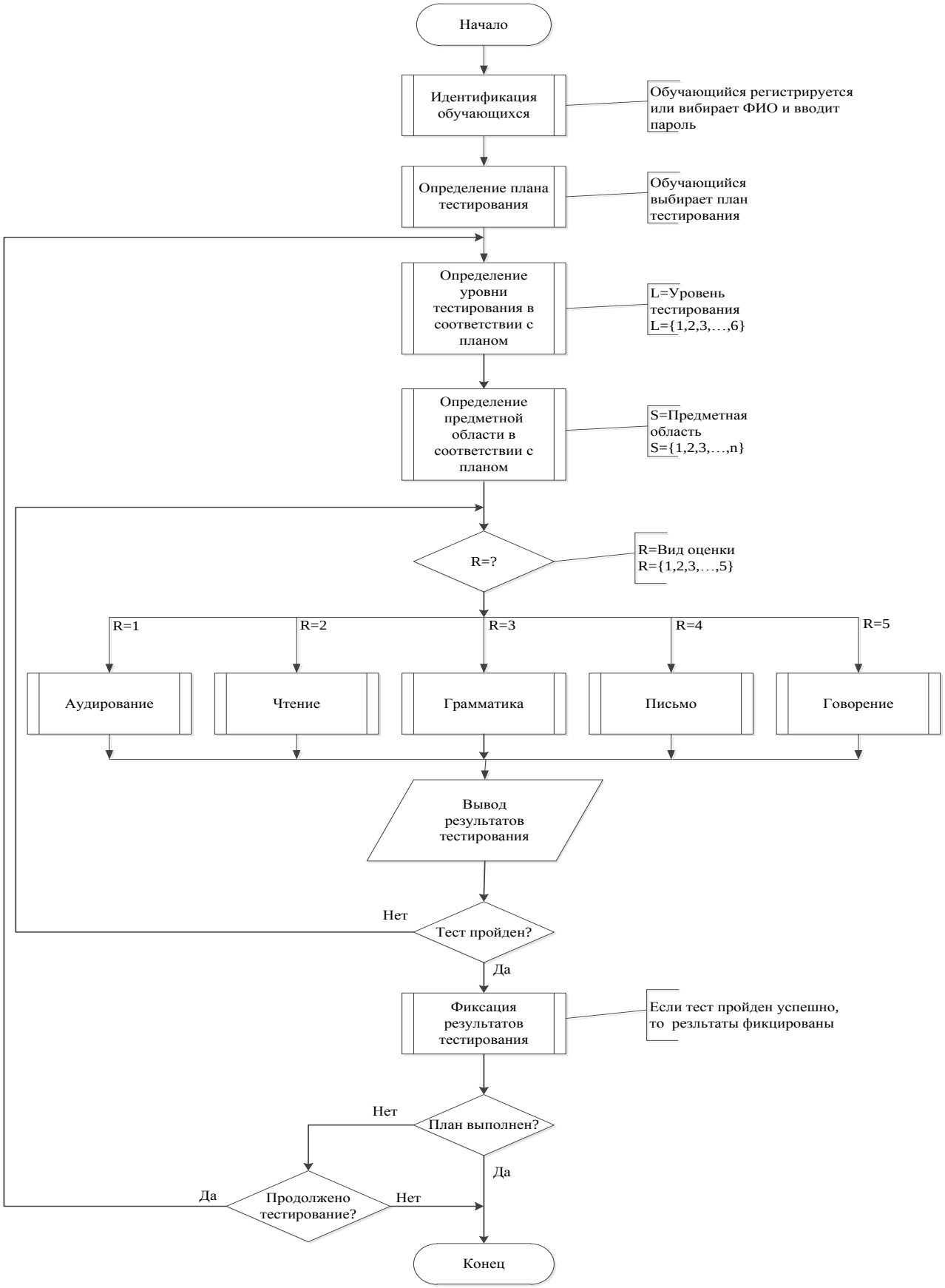

Рисунок 3 - Схема алгоритма работы системы СОЛЗ.

Основными задачами при создании системы СОЛЗ являются: разработка моделей представления знаний для всех видов оценки и их реализация; разработка семантического интерфейса с пользователем, разработка алгоритмов формирования тестов в соответствии с планом тестирования, разработка алгоритмов оценки и др.

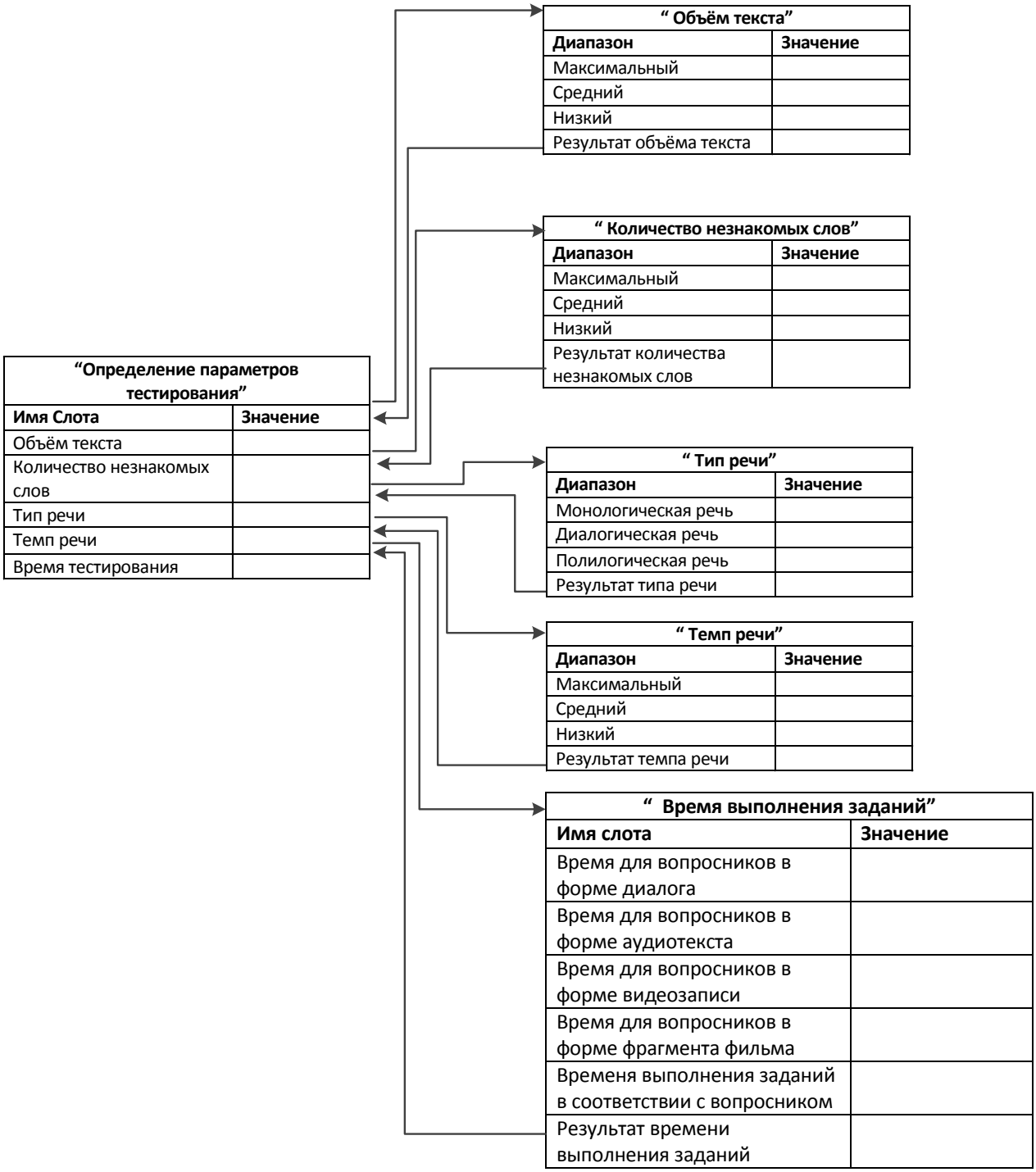

Рисунок 4 – Фрагмент фреймовой модели представления знаний.

На рисунке 4 приведен фрагмент фреймовой модели представления знаний, которую можно использовать для оценки восприятия речи. На данном рисунке представлены толь-

ко процедурные фреймы, которые отвечают за достижение подцели, связанной с определением параметров тестирования. Предлагается использовать способ конкретизации фреймов – «по конъюнкции слотов», а для означивания самых слотов использовать два способа – «с помощью служебных процедур» и «демонов». Процесс конкретизации слотов верхнего уровня является сложным, поэтому возникает необходимость использовать ряд фреймов нижнего уровня.

Реализовать предложенную модель можно на основе объектно-ориентированного программирования [4]. При этом в объектной модели разрабатываемой системы можно использовать классы-сущности для реализации процедурных фреймов, управляющие классы для представления мета-знаний и интерфейсные классы для организации диалога между системой СОЛЗ и внешними действующими лицами.

В результате проведенных исследований были определены виды оценок лингвистических знаний, параметры тестирования для каждого вида, разработана логическая модель. Для проектирования системы предложен подход, основанный на анализе семантической модели с использованием теории информационных семантических систем. Для представления знаний в системе СОЛЗ разработана фреймовая модель. На примере оценки восприятия речи показан подход к проектированию одного из основных компонентов интеллектуальной системы.

Список литературы

1. Андрюшина Н.П., Жорова А.П., Макова М.Н., Норейко Л.Н. Тренировочные тесты по русскому языку как иностранному. III сертификационный уровень. Общее владение. СПБ.: Златоуст. 2010. 152 с.
2. Соломатин Н.М. Информационные семантические системы. М.: Высшая школа. 1989. 127 с.
3. Тестирование иностранных языков. Материалы сайта // LTI Language Testing International ["Международное Тестирование Языков"]. Режим доступа: http://www.languagetesting.com/ (Дата обращения: 30.04.2015)
4. Иванова Г.С., Ничушкина Т.Н., Пугачев Е.К. Объектно-ориентированное программирование: Учебник для вузов / ред. Иванова Г. С. 3-е изд., стер. М.: МГТУ им. Н.Э. Баумана. 2007. 368 с.