

**Ю.Н.Караулов**

**ГОДОВОЙ ОТЧЕТ ПО ГРАНТУ РГНФ 12-04-12039  
"ИНФОРМАЦИОННАЯ СИСТЕМА  
КОГНИТИВНЫХ ЭКСПЕРИМЕНТОВ (ИСКЭ)" – 2012 Г.**

**Научные проблемы, на решение которых направлен проект и  
его содержание**

Целью проекта является построение экспериментальной программируемой компьютерной системы, моделирующей механизмы работы языкового сознания, которая, по замыслу участников проекта, будет представлять собой информационно-программную среду для проведения эмпирических, а затем и теоретических исследований. Эти исследования будут направлены на решение фундаментальной проблемы, касающейся нескольких смежных научных дисциплин (когнитивной лингвистики, компьютерной лингвистики, когнитивной психологии, психолингвистики, культурологии, социологии) – проблемы типологии знаний, их объема, качества, оперирования ими в масштабе русской языковой личности.

В рамках названной проблемы в проекте предполагается решить несколько конкретных задач:

Первая и наиболее важная — это построение (проектирование и практическая реализация) собственно информационной системы для поддержки эмпирических исследований (условное название — Информационная система когнитивных экспериментов — ИСКЭ).

Последнее задает ограничения содержательного плана, поскольку фактический материал исследования поставляется массовым экспериментом с носителями языка-культуры в условиях использования ими языка обиходно-повседневного общения.

Построение ИСКЭ, прежде всего, позволяет: а) смоделировать механизмы работы языкового сознания и б) установить его размерности; в) на основе теоретической разработки вопроса о типах знания и выделения в их числе минимальной и более крупных единиц (когнем, концептов, суперконцептов, концептосфер и др.) разработать операциональные определения понятий видов знания (живые, остаточные, обязательные = рецептные, факультативные = ретушные, личностные, потенциальные, профессиональные и лингвокультурные.

Объем усредненного личностного знания определяется также экспериментально: показателем уровня насыщения базы знаний является прекращение появления новых когнем в материалах опроса испытуемых. Предположительное количество записей в базе, отражающее усредненный объем личностного знания, должно достигать тридцати тысяч. Качество знаний языковой личности устанавливается на основе: а) разнообразия предметных областей действительности, к которым они относятся; б) соотношения в их числе обязательных и факультативных. В ИСКЭ моделируются процессы оперирования в ходе эксперимента знаниями со стороны языковой личности, включая следующие шаги:

- 1) Установление когнемы на основе заданного смысла.
- 2) Построение ассоциативной цепочки для данной когнемы на основе ассоциативно-вербальной сети.
- 3) Перестройка цепочки в пропозицию.
- 4) Преобразование пропозиции в дискурс.
- 5) Квалификация эквивалентности – полной или асимметричной – между исходной формулой смысла и итоговым дискурсом.

Прикладные аспекты использования ИСКЭ охватывают круг задач, связанных с модернизацией приемов поиска информации, организацией на новых основаниях контроля профессиональных знаний, а также с проблемами обучения русскому языку и культуре иноязычных пользователей.

Устройство самой ИСКЭ конструируется изоморфно структуре минимальной единицы знания, что и позволяет найти возможность согласования принципов оперирования сетевыми структурами, в которых существуют минимальные единицы знания, со структурами координатно-иерархическими, относящимися к концептуальному уровню.

Программно-моделирующая подсистема ИСКЭ (когнайзер) состоит из пяти компонентов, выполняющих следующие когнитивные функции: смысловую, которая оформляется как гипертекст; операционную, содержащую перечень способов задания смысла, которыми пользуется русская языковая личность; знаковую, которая охватывает всю совокупность установленных в эксперименте когнем; предметную, переводящую зафиксированное элементарное знание на концептуальный уровень; и качественно-оценочную, устанавливающую релевантность/нерелевантность (обязательность/факультативность) конкретной когнитивной единицы для языковой личности. При построении когнайзера в качестве инструментария исследования различных размерностей языкового сознания, так же как и его содержательных характеристик предполагается использовать три типа компьютерных моделей: формально-логические, нейросетевые и эволюционные.

## **Содержание фактически проделанной работы за 2012 год**

### **1-й квартал 2012 года**

Все заявленные работы были выполнены.

1. В Интернет развернута рабочая эксплуатационная версия ИСКЭ (доступ по ссылке со страницы проекта по адресу <http://it-claim.ru/Projects/RFH/ISKE.htm>). Краткое описание системы и правил работы с ней приведены в первом выпуске Альманаха в разделе: «I квартал 2012 г. ЯНВАРЬ — МАРТ. 1. МАТЕРИАЛЫ ИССЛЕДОВАНИЙ И РАЗРАБОТОК. Электронные издания CLAIM».

2. Проведены синхронные ассоциативный и когнитивный эксперименты в с лексикой предметной области «Информатика и вычислительная техника». В ассоциативном эксперименте приняли участие ~100 респондентов (студентов старших курсов МГТУ им.Н.Э.Баумана и МГУП им.Ивана Фёдорова, осваивающих образовательные программы по специальностям «Автоматизированные системы обработки информации и управления» — 230102, «Информационные технологии в медиаиндустрии» — 230203 и «Информационные технологии в дизайне» — 230204). В когнитивном эксперименте принимали участие те же студенты. Было сформировано более 1000 когнем, содержащих различные формулы смысла (вопросов тематических кроссвордов). В результате экспериментов сформированы два массива данных. Подробно эксперимент описан в первом выпуске Альманаха в разделе: «I квартал 2012 г. ЯНВАРЬ — МАРТ. 1. МАТЕРИАЛЫ ИССЛЕДОВАНИЙ И РАЗРАБОТОК. Электронные издания CLAIM».

3. Проводилась разработка макетной версии системы тестового контроля знаний на основе когнемной модели СКЗ-КМ. Она развернута в Интернете (доступ по ссылке со страницы проекта по адресу <http://it-claim.ru/Projects/RFH/ISKE.htm>). Описание представлено в первом выпуске Альманаха в разделе: «I квартал 2012 г. ЯНВАРЬ — МАРТ. 1. МАТЕРИАЛЫ ИССЛЕДОВАНИЙ И РАЗРАБОТОК. Электронные издания CLAIM».

### **2-й квартал 2012**

Заявленные работы в основном были выполнены.

1. Разработаны методики интеграции базы данных лингвокультурного тезауруса с лексикографическими базами данных; осуществлена интеграция с базами данных словарей синонимов, антонимов. Представлены в первом выпуске Альманаха в разделе: «II квартал 2012 г. АПРЕЛЬ — ИЮНЬ. 1. МАТЕРИАЛЫ ИССЛЕДОВАНИЙ И РАЗРАБОТОК. Электронные издания CLAIM».

2. В связи с меньшим в сравнении с заявленным участниками проекта объемом финансирования, увеличить лингвокультурный тезаурус на 3000 когнем не удалось. Были проведены когнитивные эксперименты, в результате которых были сформированы более 1000 когнем. Материал экспериментов подготовлен к вводу в базу знаний лингвокультурного тезауруса. Подробно эксперименты описаны в первом выпуске Альманаха в разделе: «II квартал 2012 г. АПРЕЛЬ — ИЮНЬ. 1. МАТЕРИАЛЫ ИССЛЕДОВАНИЙ И РАЗРАБОТОК. Электронные издания CLAIM».

### **3-й квартал 2012 года**

Все заявленные работы были выполнены.

Разработаны алгоритмы и программы реализации *формально-логической* модели функционирования когнайзера; проведены экспериментальные исследования эффективности разработанных программ. На основе иерархии <когнема-концепт-суперконцепт-концептосфера> проведена типологии языковых знаний, операциональные определения типов знаний и установление их размерности. Проведены компонентные количественные исследования изменения содержания когнем, концептов, суперконцептов и концептосфер лингвокультурного когнайзера. Материалы по этим работам представлены в первом выпуске Альманаха в разделе: «III квартал 2012 г. ИЮЛЬ — СЕНТЯБРЬ. 1. МАТЕРИАЛЫ ИССЛЕДОВАНИЙ И РАЗРАБОТОК. Электронные издания CLAIM».

### **4-й квартал 2012 года**

Все заявленные работы были выполнены.

Сделано формально-логическое описание ассоциативных отношений вербальных единиц смыслопорождающей компоненты языкового сознания и описание генетических алгоритмов знакопорождения в ассоциативно-вербальной сети (формально-логическая и эволюционная модели когнайзера). Материалы по этим работам представлены в первом выпуске Альманаха в разделе: «IV квартал 2012 г. ОКТЯБРЬ — НОЯБРЬ. 1. МАТЕРИАЛЫ ИССЛЕДОВАНИЙ И РАЗРАБОТОК. Электронные издания CLAIM».

## **Представление в 2012 году результатов работ по проекту**

Сформировано электронное Интернет-издание альманаха исследований и разработок по проекту на сайте научно-образовательного кластера CLAIM по адресу [http://it-claim.ru/Projects/RFH/ISKE\\_Almanac.htm](http://it-claim.ru/Projects/RFH/ISKE_Almanac.htm).

В течение января — декабря будет (в связи с тем что формальный отчет сдается до 30 ноября 2012 г.) опубликовано 20 печатных и 19 электронных изданий тезисов и презентаций докладов, статей и др. материалов работ по проекту, тексты которых представлены в первом выпуске Альманаха поквартально в разделах: «... 2. ПЕЧАТНЫЕ ПУБЛИКАЦИИ УЧАСТНИКОВ ПРОЕКТА» и «...3. ЗАПИСИ БЛОГОВ, ДИАЛОГИ ФОРУМОВ, ТЕКСТЫ И ПРЕЗЕНТАЦИИ ВЫСТУПЛЕНИЙ».

## **Поименные результаты работы в 2012 г. по проекту**

**Юрий Николаевич Караулов**

1. Осуществлялось руководство и координация работ по проекту.

2. На основе иерархии <когнема-концепт-суперконцепт-концептосфера> проведена типология языковых знаний, даны операциональные определения типов знаний и установлены их размерности.
3. Опубликовано тезисы доклада на международной конференции и одна электронная публикация.

***Екатерина Алексеевна Выломова***

1. Разработаны алгоритмы и программы реализации *нейросетевой* модели функционирования когнайзера; проведены экспериментальные исследования эффективности разработанных программ.
2. Проведены компонентные квантитативные исследования изменения содержания когнем, концептов, суперконцептов и концептосфер лингвокультурного когнайзера.
3. Осуществлено электронное Интернет-издание альманаха исследований и разработок по проекту на сайте научно-образовательного кластера CLAIM по адресу: [http://it-claim.ru/Projects/RFH/ISKE\\_Almanac.htm](http://it-claim.ru/Projects/RFH/ISKE_Almanac.htm) .
4. Опубликовано 3 печатные и 3 электронные публикации.
5. Прошла стажировку в летней школе-конференции NASSLLI (North American Summer School of Logic, Language and Information – <http://nasslli2012.com/>).

***Антон Валерьевич Гуца***

1. Разработана и размещена в интернет макетная версия «Системы тестового контроля знаний на основе когнемной модели — СКЗ-КМ» (доступ к информационно-программному изданию осуществляется со страницы проекта по адресу <http://it-claim.ru/Projects/RFH/ISKE.htm> ).
2. В рамках *эволюционной* модели когнайзера начата разработка «Инструментального программного средства проектирования генетических алгоритмов — DEvol».
3. Опубликовано 4 электронные публикаций.

***Александр Иванович Панченко***

1. Опубликовано 5 печатных и 2 электронных публикации.
2. Подготовлена и передана на рецензирование диссертация на соискание ученой степени PhD «Hybrid Similarity-Based Semantic Relation Extraction» в Catholic University of Louvain. Предполагаемый срок защиты декабрь-январь 2012-13 гг.

***Игорь Васильевич Ружицкий***

1. Осуществлялась подготовка к тестированию и верификации ИСКЭ (разработка тестовых примеров, отработка основных операций — функциональности разработанной системы).
2. Опубликовано одна печатная публикация.

***Александр Викторович Сиренко***

1. Развернута рабочая эксплуатационная версия ИСКЭ (доступ к системе с интернет-страницы проекта по адресу <http://it-claim.ru/Projects/RFH/ISKE.htm> ) .
2. Разработаны и описаны методики интеграции базы данных лингвокультурного тезауруса с лексикографическими базами данных; осуществлена интеграция с базами данных словарей синонимов, антонимов.
3. Разработаны алгоритмы и программы реализации *формально-логической*, моделей функционирования когнайзера; проведены экспериментальные исследования эффективности разработанных программ. Сделаны формально-логическое описание ассоциативных отношений вербальных единиц смыслопорождающей компоненты языкового сознания и описание алгоритмов знакопорождения в ассоциативно-вербальной сети.
4. Опубликовано 4 печатные и 4 электронные публикации, в том числе 2 в рецензируемых журналах.

***Юрий Николаевич Филиппович***

1. Начато формирование экспериментальной базы знаний в предметной области «Информатика и вычислительная техника» — проведены синхронные ассоциативный

(~100 респондентов) и когнитивный (более 1000 вопросов) эксперименты, сформирован и прошел первичную обработку массив данных, разработан эскизный макет базы знаний.

2. На основе иерархии <когнема-концепт-суперконцепт-концептосфера> проведена типология языковых знаний, даны операциональные определения типов знаний и установлены их размерности.

3. Сделано формально-логическое описание ассоциативных отношений вербальных единиц смыслопорождающей компоненты языкового сознания.

4. Опубликовано 7 печатных и 3 электронные публикации, в том числе 3 в рецензируемых журналах.

#### **Галина Александровна Черкасова**

1. Начато формирование экспериментальной базы знаний в предметной области «Информатика и вычислительная техника» — проведены синхронные ассоциативный (~100 респондентов) и когнитивный (более 1000 вопросов) эксперименты, сформирован и прошел первичную обработку массив данных, разработан эскизный макет базы знаний.

2. Проведены компонентные количественные исследования изменения содержания когнем, концептов, суперконцептов и концептосфер лингвокультурного когнайзера.

3. Опубликовано 4 печатных и 2 электронных публикации.

### **Содержание работ спланированных на 2013 г.**

#### **1-й квартал 2013 года:**

1.1. Верификация и тестирование рабочей эксплуатационной версии ИСКЭ.

1.2. Разработка эффективного графического пользовательского интерфейса ИСКЭ.

1.3. Формирование массива данных для создания экспериментальной базы знаний в предметной области «Информатика и вычислительная техника» — проведение синхронных ассоциативных (100 респондентов) и когнитивных (1000 вопросов) экспериментов.

#### **2-й квартал 2013 года:**

2.1. Доведение объема базы знаний макетной версии системы тестового контроля знаний СКЗ-КМ до 4000 когнем.

2.2. Ввод в базу знаний лингвокультурного тезауруса 3000 уникальных когнем.

Доведение её объема до 20000 когнем

#### **3-й квартал 2013 года:**

3.1. Доработка алгоритмов и программ реализации формально-логической, нейросетевой и эволюционной моделей функционирования когнайзера; проведение экспериментальных исследований эффективности разработанных программ.

3.2. На основе иерархии <когнема-концепт-суперконцепт-концептосфера> проведение типологии языковых знаний, и операциональные определения типов знаний и установление их размерности.

3.3. Проведение компонентных количественных исследований изменения содержания когнем, концептов, суперконцептов и концептосфер лингвокультурного когнайзера.

#### **4-й квартал 2013:**

4.1. Формально-логическое и нейросетевое описание ассоциативных отношений вербальных единиц смыслопорождающей компоненты языкового сознания и описание генетических алгоритмов знакопорождения в ассоциативно-вербальной сети (формально-логическая, нейросетевая и эволюционная модели когнайзера).

### **Представление в 2013 г. результатов работ по проекту**

Электронное Интернет-издание альманаха исследований и разработок по проекту на сайте научно-образовательного кластера CLAIM по адресу WWW. It-claim.ru.

Подготовка и защита квалификационных работ, представление к защите кандидатской диссертации.

Проведение 2-х вебинаров (1 и 4 кв. года) по тематике проекта.

Подготовка трех лекций для Интернет вещания по теме проекта.

### **Ожидаемые в конце 2013 г. результаты**

1. Будет закончена верификация и тестирование развернутой в Интернет рабочей эксплуатационной версии ИСКЭ, разработан эффективный графический пользовательский интерфейс.
2. Объем лингвокультурного тезауруса будет увеличен на 3000 уникальных когнем и доведен до 20000.
3. Будет продолжено формирование экспериментальной базы знаний в предметной области «Информатика и вычислительная техника» — проведены синхронные ассоциативный (100 респондентов) и когнитивный (1000 вопросов) эксперименты, расширен сформированный эскизный макет базы знаний, доведен до объема 4000 когнем.
4. Будет разработана рабочая версия системы тестового контроля знаний на основе когнемной модели с базой знаний в объеме 4000 вопросов в предметной области «Информатика и вычислительная техника».
4. Будут доработаны алгоритмы и программы реализации формально-логической, нейросетевой и эволюционной моделей функционирования когнайзера; проведены экспериментальные исследования эффективности доработанных программ.
7. На основе иерархии <когнема-концепт-суперконцепт-концептосфера> будет продолжена типологизация языковых знаний, уточнены операциональные определения типов знаний и их размерности.
8. Будут продолжены работы по формально-логическому описанию ассоциативных отношений вербальных единиц смыслопорождающей компоненты языкового сознания и описание генетических алгоритмов знакопорождения в ассоциативно-вербальной сети (формально-логическая, нейросетевая и эволюционная модели когнайзера).
9. Будут продолжены компонентные количественные исследования изменения содержания когнем, концептов, суперконцептов и концептосфер лингвокультурного когнайзера.
10. Будут подготовлены три лекции для Интернет вещания и проведено два вебинара по тематике проекта. Участники будут вести персональные блоги.
11. Будет продолжено электронное Интернет-издание альманаха исследований и разработок по проекту на сайте научно-образовательного кластера CLAIM по адресу WWW. Philippovich.ru
12. Будут опубликованы не менее пяти статей, в том числе двух в рецензируемых журналах.
13. Представлены к защите две студенческие квалификационные работы и кандидатская диссертация.

### **Планируемые на 2013 г. содержание и объем работы каждого исполнителя проекта**

#### ***Юрий Николаевич Караулов***

Руководство и координация работ проекта. Подготовка отчета. Верификация баз знаний лингвокультурного тезауруса. Типологизация языковых знаний, уточнение операциональных определений типов знаний и их размерности.

#### ***Екатерина Алексеевна Выломова***

Нейросетевое описание ассоциативных отношений вербальных единиц смыслопорождающей компоненты языкового сознания (нейросетевая модель когнайзера).

Доработка алгоритмов и программ реализации нейросетевой модели функционирования когнайзера; проведение экспериментальных исследований эффективности разработанных программ. Электронное Интернет-издание альманаха исследований и разработок по проекту.

***Антон Валерьевич Гуца***

Доведение объема базы знаний макетной версии системы тестового контроля знаний СКЗ-КМ до 4000 когнем. Описание генетических алгоритмов знакопорождения в ассоциативно-вербальной сети (эволюционная модель когнайзера). Доработка алгоритмов и программ реализации эволюционной модели функционирования когнайзера; проведение экспериментальных исследований эффективности разработанных программ.

***Александр Иванович Панченко***

Формально-логическое описание ассоциативных отношений вербальных единиц смыслопорождающей компоненты языкового сознания (формально-логическая модель когнайзера).

***Игорь Васильевич Ружицкий***

Верификация баз знаний лингвокультурного тезауруса. Типологизация языковых знаний, уточнение операциональных определений типов знаний и их размерности.

***Александр Викторович Сиренко***

Разработка эффективного графического пользовательского интерфейса ИСКЭ. Ввод в базу знаний лингвокультурного тезауруса 3000 уникальных когнем. Доведение её объема до 20000 когнем. Доработка алгоритмов и программ реализации формально-логической модели функционирования когнайзера; проведение экспериментальных исследований эффективности разработанных программ. Представление к защите кандидатской диссертации.

***Юрий Николаевич Филиппович***

Верификация и тестирование рабочей эксплуатационной версия ИСКЭ. Формирование массива данных для создания экспериментальной базы знаний в предметной области «Информатика и вычислительная техника» — проведение синхронных ассоциативных (100 респондентов) и когнитивных (1000 вопросов) экспериментов. Типологизация языковых знаний, операциональные определения типов знаний и установление их размерности. Подготовка трех лекций для Интернет вещания. Проведение вебинаров по тематике проекта.

***Галина Александровна Черкасова***

Верификация баз знаний лингвокультурного тезауруса. Формирование массива данных для создания экспериментальной базы знаний в предметной области «Информатика и вычислительная техника» — проведение синхронных ассоциативных (100 респондентов) и когнитивных (1000 вопросов) экспериментов. Проведение компонентных количественных исследований изменения содержания когнем, концептов, суперконцептов и концептосфер лингвокультурного когнайзера.