

## **АВТОМАТИЗИРОВАННОЕ ИЗВЛЕЧЕНИЕ ТЕОРЕТИЧЕСКИХ ЗНАНИЙ СУБЪЕКТА В РАМКАХ ВОПРОСНО-ОТВЕТНОГО ВЗАИМОДЕЙСТВИЯ НА ОСНОВЕ АНАЛИЗА ТЕКСТА ОТВЕТА НА ЕСТЕСТВЕННОМ ЯЗЫКЕ**

### **Введение**

Статья посвящена проблеме извлечения теоретических знаний субъекта, относящихся к определенной предметной области в процессе взаимодействия субъект-компьютер: рассмотрена общая проблематика извлечения знаний субъекта, выделен определенный класс задач автоматизированного извлечения знаний субъекта, а также описаны основные идеи и концептуальные положения предлагаемого подхода к решению выделенного класса задач на основе применения «фигур знания».

Сразу необходимо уточнить, что задача извлечения знаний субъекта рассматривается здесь вне контекста известной в искусственном интеллекте и теории экспертных систем задачи приобретения знаний эксперта в определенной предметной области для их последующей формализации и хранения в базе знаний интеллектуальной системы. Приобретение (извлечение) и формализация знаний эксперта – это отдельная, специфическая задача, в первую очередь направленная на выявление «глубинных» знаний, которые применяются при решении той или иной конкретной задачи. Описанный в этой статье подход к извлечению знаний субъекта, напротив, в значительной степени ориентирован на работу с теоретическими, базовыми, декларативными знаниями определенной предметной области – это, в первую очередь, основные понятия предметной области и основные связи между этими понятиями. Такие знания составляют основу той или иной области знания и не связаны непосредственно с решением каких-либо конкретных задач.

### **Общая проблематика извлечения знаний субъекта**

Перед тем как рассматривать подходы к решению задачи извлечения знаний субъекта, прежде всего нужно определить, в чем заключается природа сложности этой задачи, какие основные трудности препятствуют ее успешному решению.

Фундаментальная сложность проблемы извлечения знаний у человека состоит в том, что все ментальные, когнитивные процессы, мышление, память, работа сознания субъекта доступны для непосредственного восприятия только самому субъекту. Сознание, знания, память, мысли каждого человека всегда остаются для других людей «черным ящиком». Поэтому об уровне интеллектуального развития, содержании памяти, знаниях, способах мышления какого-либо человека можно судить только по каким-либо внешним проявлениям этих характеристик, которые для остальных людей выступают своеобразными «индикаторами» для формирования суждений. Естественно, это, прежде всего, деятельность человека, его речь, какие-либо предметы, созданные человеком и имеющие определенное смысловое наполнение. Здесь необходимо также учитывать, что механизмы работы сознания и памяти человека, процессы формирования его знаний являются в настоящее время во многом непознанными, парадоксальными и неподдающимися формализации. Например, в [1] отмечается «отсутствие хоть какого-нибудь приемлемого решения» проблемы сознания в психологии и рассматриваются очень многие парадоксы работы сознания и памяти человека, выявленные при проведении психологических экспериментов. В [2] приводятся самые различные аргументы для обоснования принципиально неалгоритмической природы работы человеческого сознания. В [3] утверждается, что «исчерпывающего описания структуры индивидуально-го знания не может быть дано принципиально».

Формирование суждений о знаниях другого человека имеет свою специфику по сравнению с формированием суждений о каких-то других его личностных качествах. В

общем случае, при решении задачи извлечения знаний субъекта на основе взаимодействия субъект-компьютер можно выделить две группы проблем: 1) проблемы, возникающие независимо от того, решается задача извлечения знаний человеком или компьютером; 2) специфические проблемы, возникающие при извлечении знаний с помощью компьютера. Рассмотрим отдельно эти группы проблем.

*Общие проблемы извлечения знаний субъекта*

1. Проблема отсутствия информации, или проблема организации взаимодействия.

Чтобы сформировать суждения о знаниях субъекта, необходимо получить информацию об этих знаниях, иначе говоря, необходимо организовать такое взаимодействие с субъектом, в процессе которого субъект так или иначе проявит свои знания. Информация, полученная от субъекта, в которой выражаются его знания, может быть представлена в различной форме: числовой, текстовой, графической, звуковой. При формировании суждений о знаниях субъекта обязательно должна учитываться специфика используемой формы взаимодействия. В качестве возможных вариантов формы взаимодействия можно привести свободный диалог, анкетирование, интервьюирование, тестирование, игровую форму.

2. Проблема интерпретации полученной информации о знаниях субъекта.

Чтобы распознать, какие именно знания в той или иной предметной области присутствуют в информации, полученной от субъекта, необходимо, как минимум, самому обладать знаниями в этой предметной области. Помимо этого, разные субъекты (или компьютерные системы), обладающие знаниями о предметной области, могут сформировать различные суждения о знаниях субъекта в результате интерпретации полученной информации, в силу субъективности восприятия и разной «картины мира» в данной предметной области. Также, для того чтобы интерпретировать полученную информацию, нужно обладать знанием той знаковой системы, в которой представлена информация. Например, двое людей, обладающих знаниями в одной и той же предметной области, но говорящих на разных языках, не смогут обмениваться знаниями друг с другом.

3. Проблема достоверности формируемой модели знаний субъекта.

Чтобы оценить адекватность сформированных суждений о знаниях субъекта, нужно сравнить полученную модель его знаний с «реальными» знаниями. Так как «реальные» знания субъекта недоступны для непосредственного восприятия, такое сравнение провести невозможно. Каким образом в таком случае можно гарантировать достоверность, адекватность построенной одним субъектом (или компьютером) модели знаний другого субъекта? Одним из возможных методов обеспечения (повышения) достоверности является получение нескольких моделей знаний субъекта различными способами, например, на основе анализа различных видов информации, полученной от субъекта, или на основе различных форм взаимодействия. Тогда можно считать, что те знания субъекта, которые присутствуют сразу в нескольких моделях, имеют достаточно высокий уровень достоверности. Очевидно, что чем больше количество полученной от субъекта информации о том или ином элементе знания, тем более достоверным будет описание этого элемента в построенной модели знания. Также, достоверность зависит от того, насколько хорошо знает предметную область и знаковую систему, в которой представлена информация, тот субъект (компьютерная система), который осуществляет извлечение знаний.

4. Проблема неполноты, неточности, нечеткости формируемой модели знаний.

Понятно, что построенная модель знаний субъекта всегда будет неполной в силу ограниченного времени взаимодействия, ограничений, накладываемых выбранной формой взаимодействия, массы случайных факторов, влияющих на процесс взаимодействия. На способность выражения знаний субъекта оказывают влияние его физиологическое и эмоциональное состояние, те или иные индивидуальные особенности субъекта: например, уровень знаний субъекта может быть высоким, но ясно и кратко выразить

эти знания в вербальной форме ему (ей) трудно. При извлечении знаний человеком большую роль играет человеческий фактор, возможная предвзятость, субъективизм. Компьютер не может построить адекватную модель в силу примитивности интерфейса взаимодействия (по сравнению с взаимодействием людей), отсутствия у компьютера полноты знаний о предметной области, в силу эргономических ограничений. Поэтому при формировании модели знаний субъекта в той или иной степени нужно учитывать так называемые НЕ-факторы – неполноту, неточность, нечеткость, неопределенность и т.д.

*Специфические проблемы извлечения знаний субъекта на основе взаимодействия субъект-компьютер.*

1. Ограниченность интерфейса взаимодействия.

Человек способен воспринимать информацию с помощью различных органов чувств; для компьютера, на данный момент, основными устройствами ввода информации в процессе взаимодействия с субъектом, являются клавиатура и мышь.

2. Ограниченность знаний компьютера о той знаковой системе, в которой он получает информацию от субъекта.

Главную роль здесь играет ограниченность знаний компьютера о естественном языке. Обмен знаниями (опытом) между людьми и извлечение знаний одного человека другим происходят, прежде всего, при общении людей на естественном языке. Самый простой путь узнать о знаниях другого человека – попросить его выразить эти знания вербально, в письменной или устной форме. Поэтому применение методов понимания компьютером естественного языка – один из главных ключей к решению задачи извлечения знаний субъекта.

3. Ограниченность знаний компьютера о предметной области.

Понятно, что чем лучше компьютер знает предметную область, чем больше объем и качество его знаний, тем более эффективно можно организовать извлечение знаний субъекта в диалоге с компьютером. Эффективное представление знаний и их использование при решении задач, традиционно решаемых человеком, являются основным предметом исследования искусственного интеллекта. Безусловно, ограниченность знаний о предметной области присутствует и у человека, но в гораздо меньшей степени (здесь имеется в виду потенциальная возможность познания человека и компьютера).

### **Рассматриваемый класс задач автоматизированного извлечения знаний субъекта и возможные подходы к решению таких задач**

Задача автоматизированного извлечения знаний субъекта может рассматриваться с самых разных точек зрения и иметь разные области практического применения. В данной статье рассматривается особый класс задач автоматизированного извлечения знаний субъекта, обладающий следующими свойствами:

1. Целью взаимодействия субъект-компьютер является извлечение базовых, теоретических, декларативных знаний субъекта в определенной предметной области. Такие базовые знания любой предметной области включают знание основных понятий этой предметной области и отношений между ними.
2. Субъект при взаимодействии с компьютером отвечает на относительно небольшое количество вопросов, связанных с определенным фрагментом (подобластью, темой) предметной области, причем ответ на эти вопросы субъект формирует на естественном языке.
3. Ответ субъекта на каждый вопрос представляет собой примерно одно предложение, то есть является достаточно кратким.
4. В результате взаимодействия с субъектом компьютер формирует модель знаний субъекта об определенном фрагменте предметной области, которая отражает знание основных понятий этого фрагмента предметной области и отношений между понятиями.

Основным отличительным свойством данного класса является формулирование ответов субъектом на естественном языке, таким образом, задача извлечения знаний субъекта в данном случае сводится к задаче извлечения знаний из текста на естественном языке, введенного субъектом, или, иначе говоря, к задаче преобразования текстовой информации в знания.

Обзор существующих подходов к пониманию естественного языка выходит за рамки настоящей статьи. Тем не менее, можно сразу отметить, что многие из существующих подходов к анализу естественно-языковых текстов в различных ЕЯ-системах не подходят для решения данной задачи в силу различных причин. Например, частотные методы анализа семантики текста неприменимы в данном случае из-за слишком небольшого объема текста, который не позволяет анализировать различные частотные характеристики, такие как частота совместной встречаемости слов.

Также сразу можно утверждать, что решение данной задачи должно быть ориентировано на так называемые семантико-ориентированный и прагматико-ориентированный подходы к анализу естественного языка [4], иначе говоря, при анализе нужно максимально учитывать особенности выбранной формы взаимодействия, контекст задаваемого вопроса, особенности класса предметных областей, для извлечения знаний по которым будет применяться разработанное решение. Такое утверждение обосновывается тем, что создание универсальных средств анализа естественного языка и описания его семантики – крайне трудоемкая и, как показывает практика, вряд ли разрешимая задача. Реальное языковое разнообразие всегда оказывается шире искусственно созданных формализмов. Поэтому, для решения конкретной задачи целесообразно применять такие методы анализа ЕЯ, которые бы максимально учитывали специфику этой задачи.

Исходя из всего сказанного выше, на рис. 1. изображена схема взаимодействия основных компонентов (моделей и алгоритмов), необходимых для решения рассматриваемого класса задач. Главным компонентом является алгоритм интерпретации (преобразования), формирующий модель знаний субъекта исходя из полученной от него в процессе ответов на вопросы текстовой информации. Для решения этой задачи необходимы как знания о языке, так и знания о мире (предметной области).

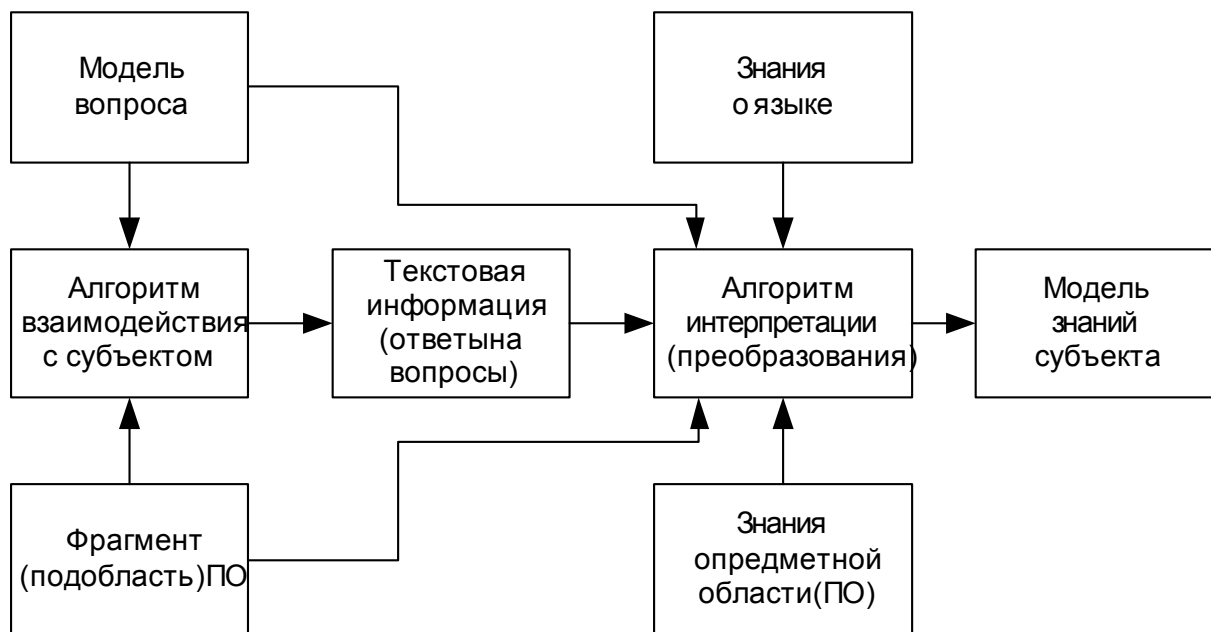


Рис. 1. Схема взаимодействия основных компонентов, необходимых для извлечения знаний субъекта на основе анализа текста ответа

## Предлагаемый подход к решению рассматриваемого класса задач

Ниже кратко описаны основные идеи и концептуальные положения, которые составляют предлагаемый подход к решению рассматриваемого класса задач. Описание разбито на части, соответствующие различным компонентам схемы на рис. 1.

### *Знания о языке и предметной области*

Основу предлагаемого подхода составляет применение когнитивного тезауруса, который представляет собой описание «интегрального», «усредненного» языкового сознания в данной предметной области в виде совокупности элементарных вербальных единиц знания – так называемых фигур знания. Такой подход к описанию языкового сознания, а также рабочие термины «когнайзер» и «фигура знания» предложены Карауловым Ю.Н. и описаны в его работах, например, в [5]. Сразу нужно указать, что в работах Караулова Ю.Н. речь идет об описании так называемого обыденного языкового сознания, наивной картины мира; идея описания когнитивного тезауруса в данной предметной области состоит в применении аналогичного подхода для описания языкового сознания в этой предметной области, являющейся частью научной картины мира. Фигура знания является пятикомпонентной структурой. Ее интенционал, т.е. сущностное содержание, образуют три следующих компонента: знак – это слово или словосочетание, формула смысла – некоторое предложение, задающее смысл знака, и способ задания смысла, однозначно определяемый формулой смысла и знаком. Необходимо подчеркнуть, что способы задания смысла не зависят от знака, т.е. смысл одного и того же знака может быть задан различными способами. Например, смысл знак «информатика» может быть задан дескрипцией «дисциплина об информации», метафорой «дитя кибернетики», синонимом «computer science» или же, например, дефиницией «фундаментальная естественная наука, изучающая процессы передачи и обработки информации». Помимо указанных трех интенциональных компонентов, в фигуру знания входят еще два компонента, определяющих ее экстенционал – это когнитивная (референтная) область, к которой относится фигура знания, а также функция, определяющая ценность (вес) данной элементарной единицы знания. Когнитивная область представляет собой отражение в сознании субъекта той реальной предметной области, к которой относится фигура знания; в случае обыденного языкового сознания совокупность таких когнитивных областей и формирует наивную языковую картину мира. При описании фигур знания когнитивного тезауруса в данной предметной области (ПО) научной картины мира, каждую фигуру знания можно отнести к той или иной подобласти (теме, разделу) внутри данной ПО. Например, внутри ПО «Информатика» можно выделить когнитивные области «Основные понятия», «Устройство компьютера», «Программирование», «Интернет» и т.д. Для характеристики ценности вербальной единицы знания Караулов Ю.Н. предлагает использовать два уровня: знание-рецепт, т.е. существенное, важное, основное знание, и знание-ретушь – второстепенное, фоновое, обладающее низкой ценностью знание. При этом фигура знания рассматривается не как некая статичная, застывшая структура, а именно в динамике, когда в разные моменты времени могут актуализироваться различные связи между рассматриваемыми пятью компонентами. Выделяется два основных режима работы когнайзера: активный и пассивный. В активном или смыслопорождающем режиме исходным является словесный знак; исходя из знака, когнайзер определяет некоторый смысл, и, тем самым, способ задания смысла. Цепочка параметров фигуры знания в активном режиме выглядит так: Знак → Способ → Смысл → Область → Функция. В пассивном режиме исходной информацией является формула смысла, в которой уже задан определенный способ задания смысла, а целью поиска когнайзера становится знак. Полную цепочку параметров фигуры знания для пассивного режима можно представить в следующем виде: Смысл → Способ → Знак → Область → Функция. В фигуре знания все пять компонентов связаны друг с другом, т.е. каждый компонент определенным образом зависит от четырех других. Схематически фигура знания изображена на рис. 2.

## 3. Слово (знак)

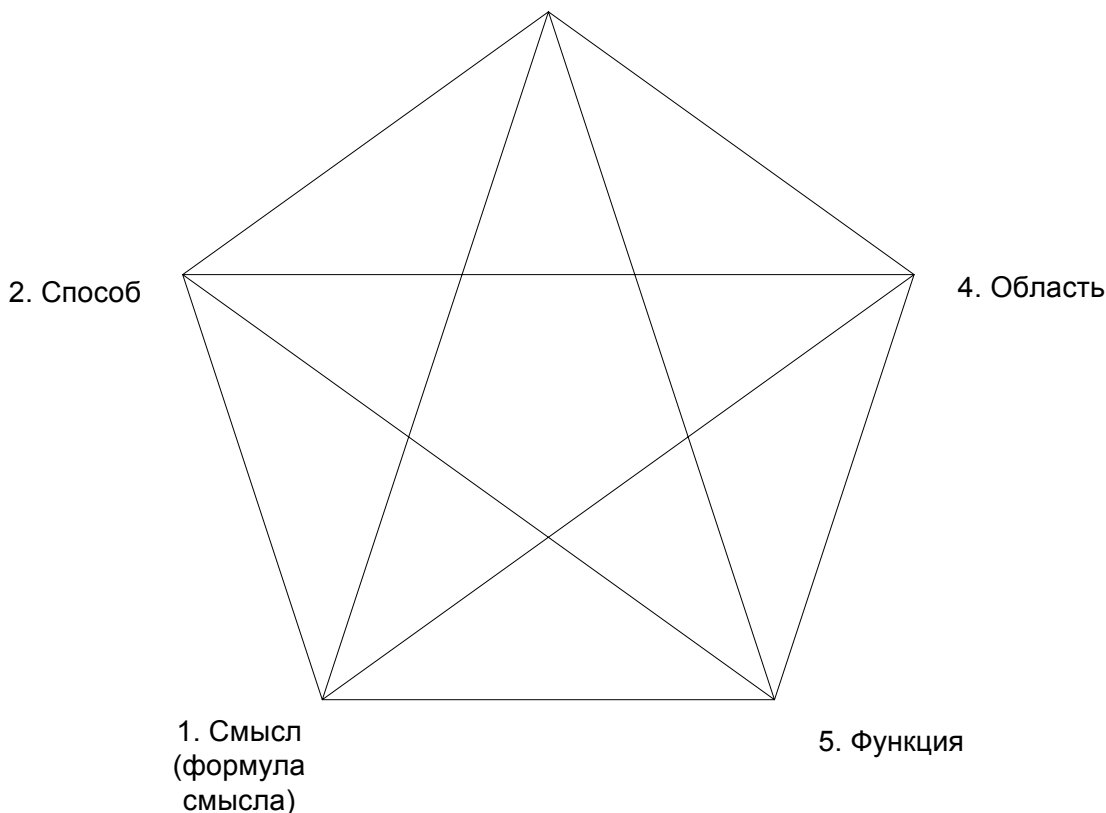


Рис. 2. Схематическое изображение фигуры знания [5].

Функционирование активного режима когнитивера описывается определенными траекториями движения в ассоциативно-вербальной сети. Взаимодействие параметров фигуры знания в активном режиме в работах Караулова Ю.Н иллюстрируется с помощью связей от стимула к реакции и от реакции к стимулу в ассоциативном тезаурусе русского языка [6]. В случае описания фигур знания в определенной предметной области научной картины мира, не основываясь на ассоциативных словарях (которых для большинства предметных областей не существует), для каждой фигуры знания можно, по крайней мере, выделить основные бинарные связи между знаком данной фигуры знания, и основными знаками-словами, выражающими понятия предметной области, и входящими в формулу смысла данной фигуры знания. Ниже в таблице даны некоторые примеры фигур знания в предметной области «Информатика» с выделенными для каждой фигуры знания связями.

Таблица 1. Некоторые примеры фигур знания в предметной области «Информатика»

Знак	Способ задания смысла	Формула смысла	Когнитивная область	Функция знания	Связи
Бит	Дескрипция	Единица измерения информации	Единицы измерения	Рецепт	Бит--- Информация
Процессор	Метафора	Мозг компьютера	Устройство компьютера	Рецепт	Процессор--- Компьютер
Информация	Дескрипция	Она передается от источника к приемнику в виде сигналов.	Основные понятия	Рецепт	Информация--- Источник Информация--- Приемник Информация--- Сигнал
Бит	Дескрипция	Один разряд двоичного кода.	Единицы измерения	Рецепт	Бит--- Двоичный код

Таким образом, когнитивный тезаурус, отражающий знания о языке и предметной области, можно представить в следующем виде:  $T = \{E, R, F\}$ . Здесь  $E$  – множество знаков (слов и словосочетаний), выражающих понятия данной ПО,  $R$  – множество отношений (связей) между понятиями ПО,  $F$  – множество фигур знания, актуализирующих те или иные отношения из  $R$  с помощью экспликации смыслов понятий из  $E$  через их связь с другими понятиями из  $E$  или понятиями смежных предметных областей и обыденной картины мира.

#### *Модель вопроса*

В качестве основных вариантов используемой формы вопроса можно предложить следующие формы:

1. Дайте толкование понятия  $E_i$ .
2. Кратко опишите, в чем состоит суть связи между понятиями  $E_1$  и  $E_2$ .
3. Заданы два понятия: 1.  $E_1$ . 2.  $E_2$ . Дайте толкование первого понятия с использованием второго понятия; это толкование должно отражать суть связи между понятиями.

Наиболее предпочтительным является использование третьей формы, так как она максимально сужает контекст, т.е. максимально четко определяет ожидаемый ответ в виде толкования смысла одного понятия с использованием другого понятия. Естественно, перед началом диалога (опроса) пользователю будут даны дополнительные пояснения в виде конкретного примера вопроса и примеров возможных ответов на этот вопрос. Множество фигур знания, соответствующих (релевантных) такому вопросу, может быть разбито на 3 подмножества: 1) множество фигур знания, где знаком является  $E_1$  или синонимы  $E_1$ ; 2) множество фигур знания, где знаком является  $E_2$  или синонимы  $E_2$ ; 3) множество фигур знания, где знак отличается от знаков фигур знания предыдущих подмножеств, но при этом в формуле смысла используются как знак  $E_1$  (или его синонимы), так и знак  $E_2$  (или его синонимы). Здесь под синонимом понимается использование разных знаков для выражения одного и того же смысла, а не широкое понимание синонима как семантически близких слов.

#### *Фрагмент (область) ПО*

В соответствии с выбранной формой вопроса (третья форма) исходный фрагмент ПО представляется как семантическая сеть, а именно, в виде ориентированного графа, вершины которого соответствуют всем понятиям, которые будут использоваться в наборе вопросов, а дуги – самим вопросам. При этом дуга каждого вопроса направлена от соответствующего понятия  $E_1$  вопроса к понятию  $E_2$ .

#### *Алгоритм взаимодействия с субъектом*

Алгоритм взаимодействия заключается в последовательном предъявлении всех вопросов, соответствующих дугам ориентированного графа, представляющего исходный фрагмент ПО, для которого проводится извлечение знаний. Порядок вопросов устанавливается экспертом до проведения контроля, таким образом, чтобы формулировка предыдущих вопросов исключала наличие подсказок для следующих вопросов.

#### *Алгоритм интерпретации (преобразования)*

Основная идея работы алгоритма интерпретации очень проста: для каждого вопроса выполняется процедура семантического сравнения текста ответа на вопрос с каждой из фигур знания, входящих в множество фигур знания, соответствующих (релевантных) вопросу. Для данного ответа и данной фигуры знания (ФЗ) результат выполнения процедуры семантического сравнения состоит в подтверждении (опровержении) следующих гипотез: 1) Какое-либо совпадение ответа и ФЗ отсутствует; 2) Ответ включает ФЗ; 3) ФЗ включает ответ; 4) Ответ и ФЗ эквивалентны; 5) Ответ и ФЗ частично пересекаются. При выполнении процедуры могут учитываться параметры фигуры знания, например, способ задания смысла, а также определенные знания о вопросе. Например, в зависимости от типа связи между понятиями в вопросе для каждого из вопросов может быть описан соответствующий фрейм. Так, связи «информатика - ин-

формация» соответствует фрейм «Описание науки», а связи «бит - информация» соответствует фрейм «Описание единицы измерения». В целом при работе процедуры семантического сравнения могут быть в той или иной степени реализованы морфологический и синтаксический анализы. Разработка конкретного алгоритма, реализующего эту процедуру, является предметом дальнейших исследований.

#### *Модель знаний субъекта*

В соответствии с описанной выше идеей алгоритма интерпретации модель знаний субъекта может быть представлена в двух различных «проекциях». Во-первых, это множество фигур знания, для каждой из которых в той или иной степени подтверждено присутствие данной фигуры в языковом сознании субъекта. Например, для каждой из фигур знания, входящих в модель знаний субъекта, может быть вычислен некий фактор уверенности в том, что данная когнитивная единица входит в знание субъекта. Во-вторых, учитывая параметры фигуры знания (прежде всего знак) и связи между понятиями ПО, соответствующие каждой из фигур знания, на основании первой «проекции» может быть синтезирована другая, представляющая собой семантическую сеть, отражающую те понятия и отношения, знание которых в той или иной степени было продемонстрировано субъектом в процессе взаимодействия.

### **Заключение**

В статье представлены только основные идеи и концептуальные положения предлагаемого подхода к решению выделенного класса задач. Основное преимущество данного подхода состоит в том, что в подавляющем большинстве аналогичных систем, организующих анализ ЕЯ для извлечения «базовых» знаний, модель знаний субъекта получается за счет сравнения либо ЕЯ-формулировок ответа, либо их концептуального представления с некоторой «эталонной», «правильной» моделью знаний. При сравнении данной формулировки с эталонными учитывается только достаточно ограниченное число формулировок, данных экспертами. При преобразовании ЕЯ-текста к концептуальному представлению крайне трудно учесть такие способы задания смысла, как метафора, цитата, суждение, ассоциация и др. Предлагаемый подход основан на построении не «эталонной», а реальной, «живой» модели языкового сознания в данной предметной области, отражающей самые различные возможные способы задания смыслов понятий разными субъектами, начиная от ассоциации и метафоры, и заканчивая строгой дефиницией. В такую модель языкового сознания одновременно входят и «книжное», академическое знание, и знание, выражаемое на «обычном», собственном языке субъекта. Для реализации предлагаемого подхода необходимо вести работу по следующим основным направлениям: 1) накопление фигур знания в определенной предметной области; 2) накопление реальных ответов субъекта на выбранные формы вопросов; 3) исследование применимости к данному подходу различных математических моделей и лингвистических процедур анализа текста. Одной из наиболее очевидных областей применения данного подхода является сфера образования.

### **Список литературы**

1. Аллахвердов В.М. Методологическое путешествие по океану бессознательного к таинственному острову сознания. – СПб.: Издательство «Речь», 2003. – 368 с.
2. Пенроуз Р. Новый ум короля: О компьютерах, мышлении и законах физики: Пер. с англ. / Общ. ред. В.О. Малышенко. Предисл. Г.Г. Малинецкого. Изд. 2-е, испр. – М.: Едиториал УРСС, 2005. – 400 с. (Синергетика: от прошлого к будущему).
3. Александров И.О., Максимова Н.Е. Закономерности формирования нового компонента структуры индивидуального знания. // Психологический журнал, 2003, том 24, № 6, с. 55-76.



4. Сулейманов Д.Ш. "Аналитический обзор отечественных и зарубежных работ в области обработки естественного языка в аспекте прагматически-ориентированного подхода". // Электронный журнал Казанского госуниверситета "Информационные технологии". - Казань, 1999.
5. Караулов Ю.Н. О единицах знания. // Полифония образования и англистика в мультикультурном мире. Тезисы первой международной конференции Ассоциации англоведов и преподавателей английского языка 25-26 ноября 2003 г. Москва, МГЛУ, 2003.
6. Караулов Ю.Н., Сорокин Ю.С., Тарасов Е.Ф., Уфимцева Н.В., Черкасова Г.А. Русский ассоциативный словарь. Т. 1-6, М., 1994-1998.
7. Филиппович Ю.Н., Черкасова Г.А., Дельфт Д. Ассоциации информационных технологий: эксперимент на русском и французском языках. С предисловием Уфимцевой Н.В. М.: МГУП, 2001. 304 с. – Книга в комплекте с CD-ROM.
8. Филиппович Ю.Н., Прохоров А.В. Семантика информационных технологий: Опыт словарно-тезаурусного описания. С предисловием А.И. Новикова М.: МГУП, 2002. 368 с. – Книга в комплекте с CD ROM.
9. Шаров Д.А. Система анализа формулировок. // Статьи, принятые к публикации на сайте международной конференции Диалог' 2004.
10. Гаврилова Т.А., Червинская К.Р. Извлечение и структурирование знаний для экспертных систем. – М.: Радио и связь, 1992.
11. Попов Э.В. Общение с ЭВМ на естественном языке. Изд. 2-е, стереотипное. – М.: Едиториал УРСС, 2004. – 360 с. (Науки об искусственном).
12. Черепанова Ю.Ю. Тестирование теоретических знаний на основе применения тезауруса семантических полей для построения концептуальных моделей текста ответа обучаемого на естественном языке. // Материалы международной научно-методической конференции «Образование и виртуальность». - Ялта: ЯИМ, 2001.
13. Всеволодский С.Н., Гаврилов А.В. Архитектура интеллектуальной системы тестирования знаний с анализом ответов на естественном языке. // Международная конференция ИСТ-2003 "Информационные системы и технологии", Новосибирск, 2003. - Т.3, С. 114-115.