

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

УДК 002.513:681.322

А.А. Кравцов, С.Ф. Липницкий, Д.Р. Насуро, Д.В. Прадун

ИНТЕЛЛЕКТУАЛИЗАЦИЯ ПРОЦЕССОВ
ОБРАБОТКИ ТЕКСТОВОЙ ИНФОРМАЦИИ

Представлен обзор работ, выполненных в ОИПИ НАН Беларуси в рамках решения проблемы интеллектуализации информационных процессов поиска и обработки текстовых документов. Рассмотрены модели синтаксического и семантического анализа входных предложений, синтаксического синтеза выходных, генерации стратегий информационного поиска и навигационных маршрутов в гипертексте, создания и актуализации гипертекстовых баз данных.

Введение

Начало современной теории и практики обработки текстовой информации с ориентацией на использование простейших технических средств (карт с краевой перфорацией, счетно-перфорационных машин и т. п.) было положено в 50-х гг. прошлого века. До конца 1970-х гг. развитие информационных систем основывалось преимущественно на технологических усовершенствованиях процедур поиска, а задачам логико-семантического характера уделялось мало внимания [1].

В 1980-х гг. исследователи пришли к убеждению, что дальнейшее повышение эффективности информационных систем в рамках «технологического» этапа становления теории поиска уже невозможно: их техническая эффективность по суммарному показателю «полнота + точность» не превышала, как правило, 100 – 120 % [2 – 4]. Так, например, согласно результатам трехлетнего исследования в Великобритании 40 информационных систем, используемых в девяти странах, выяснилось [5], что традиционными, «ручными», средствами библиотекари отыскивали в среднем 20,7 % релевантных документов; менее эффективным был поиск по карточкам специальных информационных служб; наихудшие же результаты (5 % релевантных документов) были получены с помощью крупнейших автоматизированных систем Великобритании и США. Низкие показатели эффективности поиска и появление новых технических и программных средств (оптические диски, широкополосные сети передачи данных, программные мультисреды) явились достаточным основанием для активизации исследований по интеллектуализации информационных систем на основе новых информационных технологий представления и переработки знаний [6, 7]. Другой важной предпосылкой интеллектуализации стало появление полнотекстовых баз данных, которые представляют собой электронные аналоги печатных документов [8, 9]. Середина 1980-х гг. соответствует третьему, компьютерному, этапу эволюции безбумажного общества (первые два – добумажный и бумажный), характерной особенностью которого является появление публикаций, которые издаются только в машиночитаемой форме и не печатаются параллельно на бумаге, что позволило снизить стоимость доставки информации потребителю на 40 – 60 % [10 – 12].

Стремление к повышению эффективности информационных систем и существование электронных изданий выдвинули в ряд актуальных новую цель интеллектуализации – возможность работы с полнотекстовыми базами данных без посредников-индексаторов, в том числе обработки текстов на разных языках с интерфейсом на языке пользователя. Решение этих задач невозможно без автоматизации процессов анализа и синтеза текстов, интеллектуализации стратегий информационного поиска и процессов создания и актуализации баз данных и знаний, являющихся основой интеллектуализации.

С целью постановки задач интеллектуализации необходимо было провести анализ, систематизацию и декомпозицию задач информационного поиска и обработки текстовых документов [13]. В результате декомпозиции выявлены следующие инвариантные информационные процессы: анализ входных предложений текста и синтез выходных; генерация стратегий информационного поиска и навигационных маршрутов в гипертексте; создание и актуализация гипертекстовых баз данных.

В данной статье приводится обзор работ, выполненных в Объединенном институте проблем информатики Национальной академии наук Беларуси при решении проблемы интеллектуализации информационных систем.

1. Задачи моделирования информационных процессов

Роль моделирования как метода познания в настоящее время неуклонно повышается в связи с увеличивающейся сложностью исследуемых объектов и ростом затрат на проведение экспериментальных исследований. При моделировании информационных процессов обычно различают три уровня: в качестве модели *первого* уровня рассматривается некоторое эвристическое описание информационной базы и процедур поиска ее объектов; на *втором* уровне (математическое моделирование) модель первого уровня уточняется с использованием математического аппарата; под *третьим* уровнем понимают построение компьютерной модели информационной системы, т. е. ее программную реализацию.

1.1. Особенности и цели моделирования

В отличие от традиционных математических моделей естествознания, где для целей моделирования используются непрерывные структуры анализа, при моделировании процессов поиска и смысловой обработки объектов баз данных и знаний применяются главным образом дискретные структуры теории множеств, математической логики, аналитической теории алгоритмов (в силу дискретного характера этих объектов).

Задачи моделирования информационных процессов относятся к классу интеллектуальных, для которых не существует априори известной схемы решения – она должна быть построена в процессе решения.

Учитывая разнообразие моделируемых информационных процессов и целей их интеллектуализации, используемый математический аппарат и взятые за основу моделирования исходные концепции должны обеспечивать единый подход к формализации интеллектуальных процессов.

Основная цель моделирования интеллектуальных процессов в информационных системах аналогична целям моделирования в других областях знаний. Она состоит в повышении эффективности этих процессов и включает несколько подцелей:

- формирование и формализацию совокупности исходных понятий для моделирования интеллектуальных процессов;
- исследование в рамках моделирования свойств этих понятий и интеллектуальных процессов;
- построение и оптимизацию алгоритмов и программ реализации интеллектуальных процессов.

1.2. Типология задач моделирования

Проблема моделирования сводится к решению ряда задач и подзадач интеллектуализации информационных процессов, представленных на рис. 1 [13]. Задачи моделирования данных и знаний решены на основе формализации лингвистических понятий семантических отношений, благодаря чему обеспечивается единый теоретический подход к моделированию и алгоритмизации всех интеллектуальных информационных процессов.

Процессы синтаксического и семантического анализа входных и синтаксического синтеза выходных предложений промоделированы на основе использования моделей данных и знаний. Модели анализа и синтеза обеспечивают эффективность, устойчивость к ошибкам и универсальность соответствующих алгоритмов в смысле возможности их использования в инфор-

мационных системах различного назначения. Универсальность достигается за счет независимости алгоритмов от информационных языков в рамках ограничений, накладываемых на их синтаксис.

Модели генерации стратегий информационного поиска и навигационных маршрутов в гипертексте позволяют оптимизировать разработанные на их основе алгоритмы не только в смысле релевантности найденной информации, но и ее пертинентности.

Модели создания и актуализации гипертекста позволяют автоматизировать не только процесс сегментации текстовых документов большого объема, но и процесс построения на множестве полученных сегментов отношения строгого порядка.

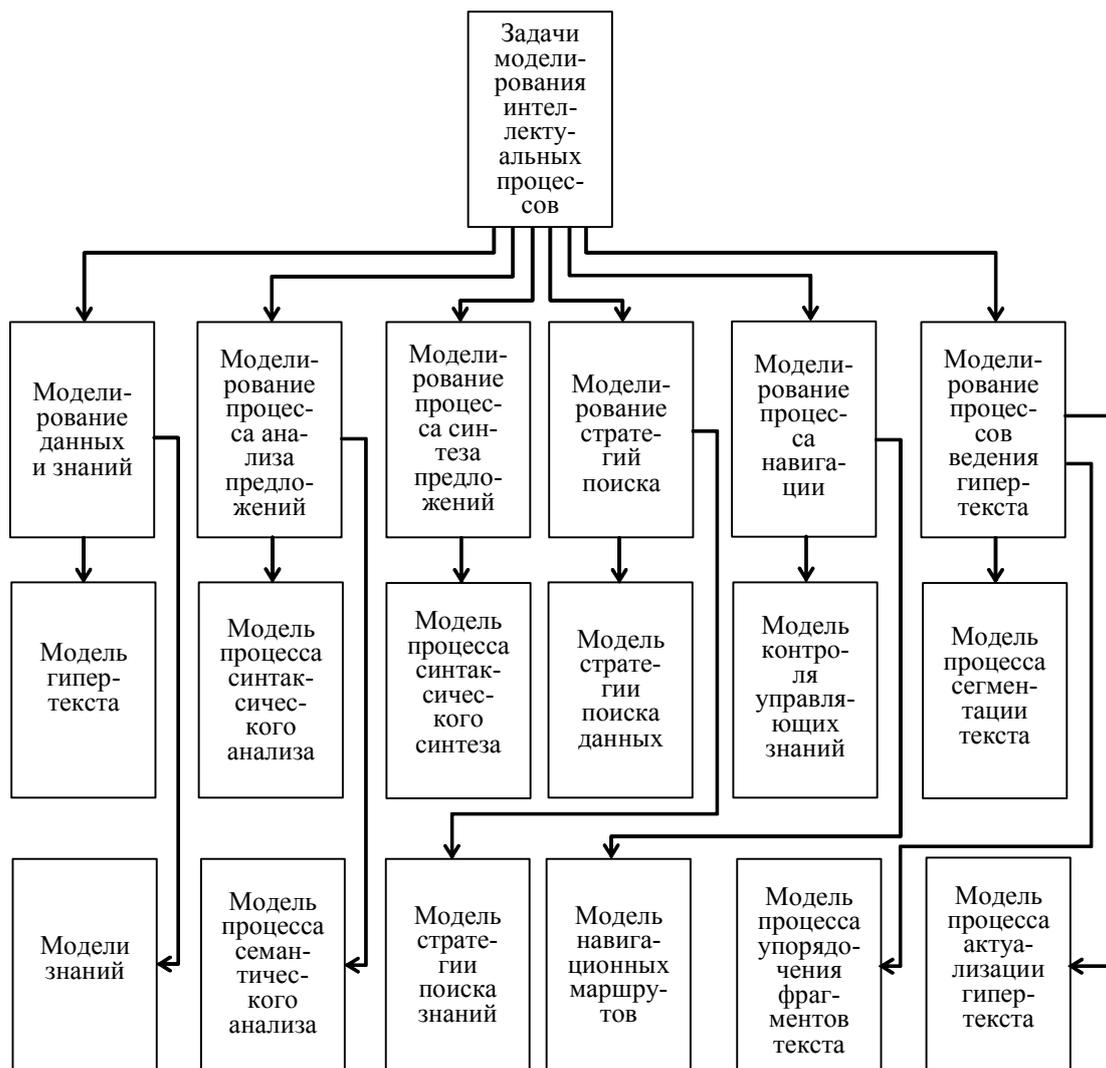


Рис. 1. Типология моделей информационных процессов

2. Моделирование данных и знаний

Для решения задач интеллектуализации необходимо было разработать систему базовых понятий для моделирования информационных процессов. Для выявления такой совокупности понятий были промоделированы структуры данных и знаний. Решение этой задачи стало возможным благодаря использованию формализованных и исследованных в [14 – 16] основных семантических отношений. Это отношения парадигматического подчинения, парадигматической эквивалентности, или синонимии, и синтаксического подчинения.

2.1. Семантические отношения

В лингвистике различают парадигматические и синтагматические семантические отношения. Парадигматические отношения существуют между словами и фразами языка независимо от контекста и объединяют понятия, между которыми имеется устойчивая связь. В противоположность парадигматическим, синтагматические отношения возникают в тексте, т. е. между словами и словосочетаниями каждого конкретного его предложения.

Содержательно отношение парадигматического подчинения соответствует родовидовому отношению между словами и фразами языка (например, фраза «инструмент для обработки отверстий» и слово «сверло» парадигматически подчинены слову «инструмент»). Формально оно определено как рефлексивное и транзитивное бинарное отношение (предпорядок) на множестве всех непустых цепочек в словаре языка. При этом оно должно удовлетворять следующему свойству: если некоторая цепочка δ парадигматически подчинена цепочкам β и γ , то существует цепочка α , которой парадигматически подчинены цепочки β и γ [17, 18]. Используя отношение парадигматического подчинения, формально определяется понятие категории как класса слов и словосочетаний, объединенных некоторыми общими признаками. Элементами одной и той же категории являются цепочки (слова, синтаксические конструкции), которые имеют совпадающие смысловые значения независимо от контекста (например, «языкознание – лингвистика – языковедение», «забастовка – стачка»). Для таких цепочек введено определение отношения парадигматической эквивалентности, или синонимии [16].

При моделировании синтаксической структуры предложений основным является отношение синтаксического подчинения. В естественных языках различают два вида такого подчинения – предикативность и атрибутивность. Примером предикативности служит отношение между сказуемым и подлежащим, атрибутивности – между сказуемым и обстоятельством, подлежащим и определением.

Для определения отношения синтаксического подчинения в работе [15] введена в рассмотрение специальная формальная грамматика, порождающая проективные предложения языка. Ее основное назначение состоит в распознавании их синтаксической структуры при анализе текста и формировании словарей синтагм.

2.2. Данные и знания

Информационная база системы поиска и обработки текстовой информации включает базы данных и знаний. Данные представляют собой информацию, непосредственно используемую абонентами системы. Знания являются вспомогательным компонентом информационной базы и обеспечивают возможность логической переработки данных.

Данные в системе имеют преимущественно гипертекстовую структуру в виде текстов и связей между ними. Поэтому для моделирования данных естественно было использовать концепцию гипертекста. Данные организуются в файлы следующих видов: фактографическая информация, документальная и служебная (рис. 2).

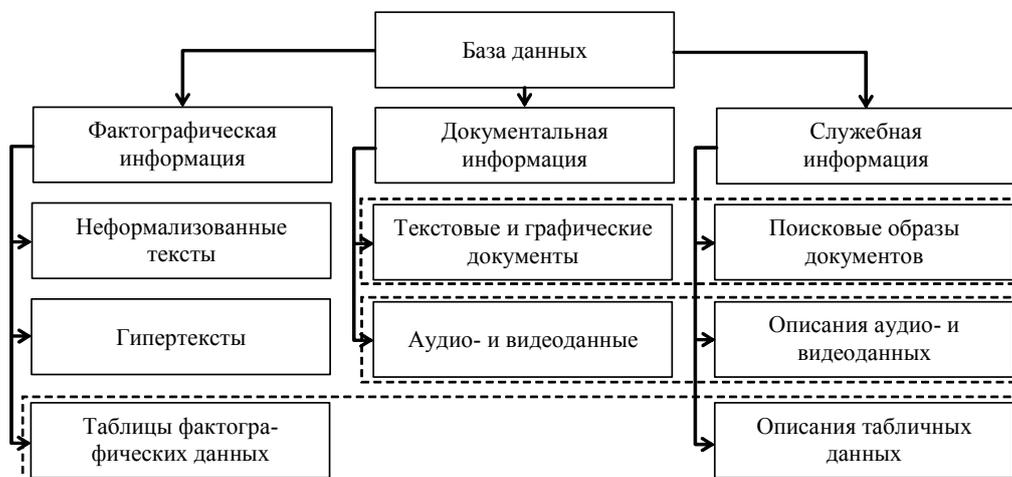


Рис. 2. Состав и структура базы данных

Фактографическая информация представлена в системе неформализованными (неструктурированными или слабоструктурированными) текстами, гипертекстами и таблицами фактографических данных, как заранее введенных в базу данных, так и полученных в результате обработки текстовых документов.

Документальная информация – это текстовые и графические документы, а также различные аудио- и видеоданные.

В качестве *служебной* информации в системе имеются файлы поисковых образов документов, а также описаний аудио-, видео- и табличных данных.

База знаний информационной системы включает лингвистические, энциклопедические и управляющие знания (рис. 3).



Рис. 3. Состав и структура базы знаний

В состав лингвистических знаний входят синтаксические и семантические деревья, словари синтагм, используемые при синтаксическом анализе входных предложений, упорядоченные синтаксические деревья, двухязычные словари 1 и 2 для перевода предложений с входных языков на внутренний и обратно, а также словарь семантических отношений, включающий отношение синонимии и родовидовое отношение.

Синтаксическое дерево предложения определено в работе [16] как ориентированный граф сужения отношения синтаксического подчинения на множество всех слов этого предложения (рис. 4).

Семантическое дерево характеризует внутреннее представление входного предложения в информационной системе [19]. В вершинах семантического дерева представлены слова внутреннего языка, а дугами изображаются семантические связи между ними. Каждая дуга этого дерева помечена парой семантических признаков. На рис. 5 изображено семантическое дерево предложения, полученное из синтаксического дерева этого предложения с использованием двухязычного словаря 1.

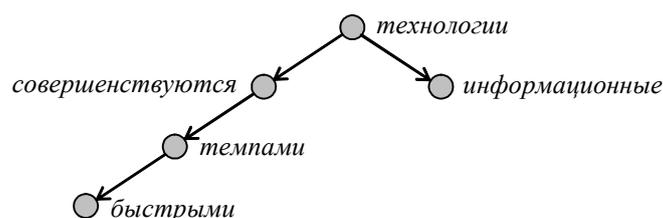


Рис. 4. Синтаксическое дерево предложения «Информационные технологии совершенствуются быстрыми темпами»

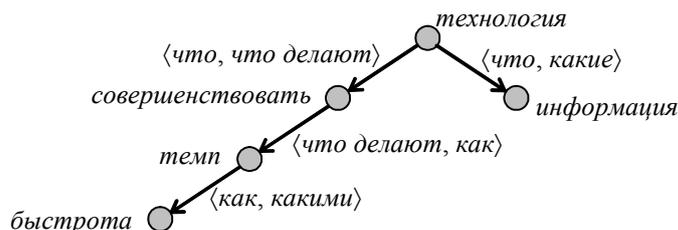


Рис. 5. Пример семантического дерева

Упорядоченное синтаксическое дерево аналогично синтаксическому дереву и отличается от него упорядоченностью дуг, исходящих из каждой вершины (рис. 6). Упорядоченные деревья используются при синтезе выходных предложений в системе [19].

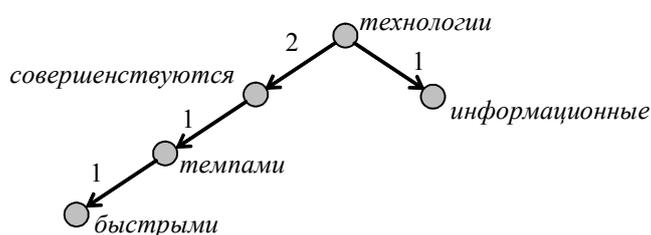


Рис. 6. Пример упорядоченного синтаксического дерева

Словарь синтагм включает пары слов вида «пароход → плывет», «синий ← горизонт», где стрелками указано направление синтаксической связи. В этих цепочках слова «пароход» и «горизонт» являются определяемыми членами соответствующих синтагм, а слова «плывет» и «синий» – определяющими.

Каждая запись двуязычного словаря 1 включает два поля: в первом представлен синтаксический граф некоторого входного предложения (на практике это преимущественно орцепи длины 1), а во втором – семантический граф синонимичного ему предложения внутреннего языка. Записи двуязычного словаря 2 получены из записей словаря 1 их инвертированием.

Словарь семантических отношений состоит из записей, каждая из которых включает три поля: в первом содержится семантическое дерево некоторого предложения; во втором – семантические деревья предложений, состоящих с исходным предложением в некоторой семантической связи; в третьем поле указывается тип этой связи.

Энциклопедические знания представлены в информационной системе сетью семантических структур, включающей семантические элементы, семантические сцены, семантические эпизоды и семантические сценарии [13].

Семантический элемент – это семантический граф (ордеререво), корнем которого служит слово внутреннего языка или семантическая переменная (слот), а все остальные вершины (также слова или слоты) являются смежными корню (рис. 7).

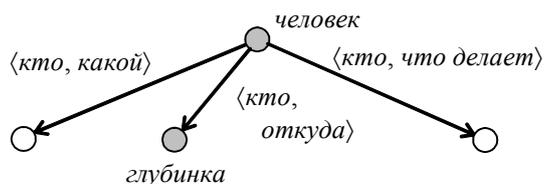


Рис. 7. Пример семантического элемента

Слоты семантического элемента могут быть заполнены словами или другими семантическими элементами. В последнем случае в слот семантического элемента помещается вершина семантического дерева, представляющего заполняющий семантический элемент.

Если один или более слотов семантического элемента заполнить семантическими элементами, то получим семантическую структуру, которую будем называть *семантической сценой*. Слоты семантической сцены, в свою очередь, могут заполняться семантическими элементами или сценами.

Семантический эпизод – это семантическая сцена, слотам которой парадигматически подчинены некоторый семантический элемент или сцена.

Управляющие знания информационной системы – это совокупности правил продукции вида «если..., то...». В позицию «если...» помещаются адреса (идентификаторы) текстов гипертекста, которые тематически предшествуют текстам (адресам, идентификаторам) в позиции «то...». Управляющие знания необходимы для реализации процесса навигации в гипертексте: на их основе моделируются и исследуются процессы построения навигационных маршрутов.

В работах [20, 21] предложена концепция автоматического реферирования текстовых документов с использованием моделей данных и знаний.

3. Моделирование интеллектуальных процессов

Процесс моделирования анализа предложений включает два этапа. На первом исследуется процедура синтаксического анализа, на втором – семантического. Обе модели основаны на формализмах, полученных в результате моделирования лингвистических знаний.

3.1. Синтаксический анализ предложений

Синтаксический анализ предложения можно реализовать путем циклического повторения следующей процедуры [16]: в предложении ищутся синтагмы, строятся их синтаксические деревья и исключаются определяющие члены так называемых маргинальных синтагм. С целью алгоритмической реализации этой процедуры необходимо было систематизировать синтагматические структуры входных предложений. Согласно полученной типологии при анализе предложений можно ограничиться рассмотрением только двух-, трех- и четырехсловных синтагматических структур.

Для реализации процедур синтаксического анализа входных предложений необходимо было выяснить также свойства двух-, трех- и четырехсловных структур, чтобы эти структуры можно было бы найти в анализируемом предложении. С этой целью доказаны утверждения, позволяющие выявить в предложении указанные структуры по совершенно формальным признакам [13].

Результаты моделирования явились основой для алгоритмизации процессов анализа проективных предложений. Проективность здесь, как и обычно, означает синтаксическую «правильность» предложения.

Анализ непроективных предложений можно реализовать путем использования алгоритмов анализа проективных предложений [18]. Оказалось, что если непроективное предложение анализировать с помощью этих алгоритмов, то в результате анализа будет получена совокупность синтаксических деревьев всех проективных подцепочек входного предложения. С целью построения дальнейших шагов анализа пришлось ввести понятие разделенной синтагматической структуры, т. е. структуры, между определяемым и определяющим членами которой располагается подцепочка некоторой другой структуры. Тогда для продолжения синтаксического анализа непроективного предложения из него необходимо исключить определяющие члены всех его разделенных синтагматических структур и снова применить алгоритмы анализа проективного предложения. После завершения работы этих алгоритмов и восстановления входного предложения его анализ можно продолжить путем локализации узловых слов всех определяющих членов разделенных синтагматических структур. Узловые слова – это определяемые члены синтагм предложения, являющиеся корнями соответствующих синтаксических деревьев.

3.2. Семантический анализ предложений

Семантический анализ предложений входного языка реализован путем формирования и использования прагматического контекста этих предложений в базе энциклопедических знаний в базе энциклопедических знаний. Анализ осуществляется в два этапа. На первом генерируется псевдосемантическое дерево входного предложения, а на втором из вершин этого дерева устраняется лексическая неоднозначность и, таким образом, строится семантическое дерево [13, 19]. При семантическом анализе основной проблемой является устранение лексической неоднозначности во входных предложениях. Решение этой проблемы сводится к построению в базе знаний контекста для анализируемого текста и поиску в нем адекватного синонима каждого омонимичного понятия.

3.3. Синтаксический синтез выходных предложений

Модель процесса интерпретации предложений выходного языка информационной системы строится в три этапа. Сначала моделируется процедура перехода от семантического дерева каждого предложения к синтаксическому. Затем на множествах вершин полученного синтаксического дерева определяется система частичных порядков. На последнем этапе моделируется процесс генерации синтаксически корректного выходного предложения.

3.4. Моделирование стратегий поиска

Задача моделирования стратегий поиска включает три подзадачи: моделирование критериев оптимальности стратегий поиска, моделирование стратегии поиска описаний объектов в гипертексте и моделирование процесса поиска знаний в сети семантических структур. На основе моделей генерируются адаптивные стратегии поиска данных и знаний.

Для решения этой задачи обобщены известные в информатике критерии оптимальности поиска «полнота» и «точность» на основе формализации понятий релевантных и пертинентных текстов. Определен также критерий типа «средние потери», с использованием которого исследованы первые два критерия. Стратегия информационного поиска моделируется в виде кортежа поисковых функций [22].

Поиск энциклопедических знаний в сети семантических структур является главной составляющей процедуры поиска и генерации ответов на неформализованные запросы пользователей информационной системы. Будем различать нетерминальную и терминальную сети. Нетерминальная сеть содержит общие сведения об объектах предметной области (например, наименования классов объектов), а в терминальной представлены сведения о конкретных объектах. Вершинами нетерминальной сети являются слоты, а терминальной – понятия. Для реализации поиска в базе энциклопедических знаний определены понятия графа-запроса и графа-ответа. Поиск графа-ответа сводится к нахождению части сети семантических структур, для которой существует гомоморфизм из графа-запроса на граф-ответ. Доказаны необходимые и достаточные условия существования гомоморфизма, моделирующего поиск графа-ответа. В силу этих условий упомянутый гомоморфизм существует, если для каждой простой цепи длины 2 графа-запроса существует гомоморфная ей часть сети семантических структур [23].

3.5. Моделирование процессов навигации в гипертексте

Основная сложность работы с гипертекстом заключается в построении маршрутов навигации в нем, т. е. определении порядка его просмотра. Существующие в настоящее время подходы к построению навигационных маршрутов сводятся главным образом к предоставлению пользователю инструментария для самостоятельного «блуждания» в гипертексте. Неэффективность подобных методов навигации очевидна, поскольку при отслеживании связей между текстами гипертекста пользователь может быть дезориентирован из-за их огромного числа. Это приводит к увеличению времени навигации и интеллектуальным перегрузкам. В нашем случае для навигации предложен подход, основанный на моделировании управляющих знаний в виде продукционного отношения на множестве текстов гипертекста. Элементами этого отношения являются правила продукции с текстами или их идентификаторами в левой и правой частях каждого правила [24, 25].

Моделирование сводится к решению двух задач: построению модели контроля эффективности знаний для управления процессом навигации и моделированию пошаговой генерации навигационного маршрута. Последняя модель обеспечивает адаптивность навигационных маршрутов к информационным потребностям пользователей информационной системы и к их предыдущим действиям в процессе навигации.

На основе модели навигации в гипертексте разработаны алгоритмы и программы гипертекстовой экспертно-обучающей системы [26 – 29] и компьютерного практикума по математике [30].

3.6. Моделирование процессов создания и актуализации гипертекста

Для решения этой задачи необходимо было ввести формальные понятия фрагмента и сегмента текста. Фрагмент – это произвольный кортеж предложений текста, а сегмент – это такой фрагмент, для которого в сети семантических структур имеется подграф, являющийся семантическим графом некоторого семантического эпизода и содержащий графы-ответы на семантические деревья всех предложений этого фрагмента, но не включающий графы-ответы на семантические деревья предложений, непосредственно предшествующих сегменту или непосредственно следующих за ним. «Точность» сегментации зависит от «полноты» сети семантических структур. Оказалось, что по мере заполнения слотов этой сети при повторной сегментации возможно разбиение некоторых текстов гипертекста на два или более сегментов. На основе моделирования и исследования процессов создания гипертекста разработаны алгоритмы его разбиения на сегменты и упорядочения полученных сегментов. Процедура же актуализации гипертекста заключается в коррекции строгого порядка на множестве текстов и повторном применении алгоритмов создания гипертекста [13]. При этом используется база энциклопедических знаний.

Дальнейшее развитие исследований в области интеллектуализации информационных процессов продолжается в направлении создания и анализа корпусов текстов с целью определения информативности текстов и их фрагментов, а также построения базы знаний в виде ситуативно-синтагматической сети [31].

Заключение

Результаты моделирования информационных процессов поиска и обработки текстовых документов обеспечивают:

- универсальность алгоритмов анализа и синтеза предложений в смысле возможности их использования в различных предметных областях;
- адаптивность алгоритмов поиска и навигации к изменяющимся условиям взаимодействия пользователей с базой данных;
- возможность использования алгоритмов создания и актуализации гипертекста для поиска и навигации в базах данных, не являющихся гипертекстовыми.

Список литературы

1. Мельчук И.А. Опыт теории лингвистических моделей «смысл – текст». – М.: Наука, 1974. – 314 с.
2. Абрамов А.В. Системы навигации в автоматизированных обучающих системах // Известия вузов. Электромеханика. – 1995. – № 4. – С. 75-77.
3. Кибрик А.Е. Лингвистические предпосылки моделирования языковой деятельности в интеллектуальных системах. – М.: Наука, 1987. – С. 33-52.
4. Скороходько Э.Ф. Семантические сети и автоматическая обработка текста. – Киев: Наукова думка, 1983. – 219 с.
5. Стогний А.А., Глазунов Н.М. Современные проблемы создания интегрированных баз данных // Кибернетика. Становление информатики. – М.: Наука, 1986. – С. 128-139.
6. Поспелов Д.А. Логико-лингвистические модели в системах управления. – М.: Энергоиздат, 1981. – 232 с.

7. Осипов Г.С. Приобретение знаний интеллектуальными системами: Основы теории и технологии. – М.: Наука. Физматлит, 1997. – 112 с.
8. Глушков В.М. Основы безбумажной информатики. – М.: Наука, 1987. – 552 с.
9. Попов Э.В. Общение с ЭВМ на естественном языке. – М.: Наука, 1982. – 360 с.
10. Любарский Ю.Я. Интеллектуальные информационные системы. – М.: Наука, 1980. – 232 с.
11. Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах. – М.: Наука, 1989. – 191 с.
12. Поспелов Д.А. Вычислительные машины становятся интеллектуальными // Кибернетика. Становление информатики. – М.: Наука, 1986. – С. 139-155.
13. Липницкий С.Ф., Ярмош Н.А. Моделирование интеллектуальных процессов в инженерных информационных системах. – Мн.: Беларуская навука, 1996. – 222 с.
14. Липницкий С.Ф. Математическая модель синтаксического анализа входных сообщений в интеллектуальной информационной системе // Вестник Белорусского университета. Сер. 1: Физ. Мат. Мех. – 1989. – № 2. – С. 54-58.
15. Липницкий С.Ф. Математическая модель распознавания синтаксической структуры предложений при обработке текстовой информации // Доклады НАН Беларуси. – 2002. – Т. 46. – № 1. – С. 60-63.
16. Липницкий С.Ф. Математическая модель синтаксического анализа текста в информационно-аналитической системе // Информатика. – 2004. – № 1. – С. 28-36.
17. Липницкий С.Ф., Яковишин В.С. Свойства синтаксических структур и анализ входных сообщений в информационной системе // Весці АН БССР. Сер. фіз.-тэхн. навук. – 1987. – № 4. – С. 89-93.
18. Липницкий С.Ф. Математическая модель синтаксического анализа непроективных и слабо проективных сообщений в интеллектуальной информационной системе // Вестник Белорусского университета. Сер. 1: Физ. Мат. Мех. – 1990. – № 2. – С. 48-50.
19. Липницкий С.Ф. Математическая модель семантического анализа и синтеза сообщений в интеллектуальной информационной системе // Весці АН БССР. Сер. фіз.-мат. навук. – 1990. – № 5. – С. 99-105.
20. Кравцов А.А., Липницкий С.Ф., Насуро Д.Р. Концепция системы автоматического реферирования текстовых документов на основе моделирования семантических отношений // Тез. докл. Междунар. науч. конф. «Интеллектуализация обработки информации» (ИОИ-2004). – Симферополь: Крымский научный центр НАН Украины, 2004. – С. 91.
21. Кравцов А.А., Липницкий С.Ф., Насуро Д.Р. Концепция системы автоматического реферирования текстовых документов на основе моделирования семантических отношений // Искусственный интеллект. – 2004. – № 2. – С. 314-319.
22. Липницкий С.Ф. Математическая модель информационного поиска, основанного на синтаксическом анализе входных сообщений // Весці АН БССР. Сер. фіз.-тэхн. навук. – 1988. – № 1. – С. 83-89.
23. Липницкий С.Ф. Элементы математической модели поиска в информационно-логической системе // Кибернетика. – 1987. – № 4. – С. 119-121.
24. Липницкий С.Ф. Метод навигации в гипертекстовой базе данных. I // Весці АН Беларусі. Сер. фіз.-мат. навук. – 1994. – № 1. – С. 100-105.
25. Липницкий С.Ф. Метод навигации в гипертекстовой базе данных. II // Весці АН Беларусі. Сер. фіз.-мат. навук. – 1994. – № 2. – С. 122-127.
26. Липницкий С.Ф., Шибут М.С., Ярмош Н.А. Гипертекстовая экспертно-обучающая система. I. Модель данных и знаний // Весці НАН Беларусі. Сер. фіз.-мат. навук. – 1998. – № 3. – С. 111-114.
27. Липницкий С.Ф., Шибут М.С., Ярмош Н.А. Гипертекстовая экспертно-обучающая система. II. Алгоритмы автоматизированного обучения // Весці НАН Беларусі. Сер. фіз.-мат. навук. – 1999. – № 1. – С. 107-111.

28. Липницкий С.Ф., Шибут М.С., Ярмош Н.А. Метод автоматизированного контроля знаний в гипертекстовой среде // Весці НАН Беларусі. Сер. фіз.-тэхн. навук. – 2000. – № 1. – С. 86-89.

29. Котова Л.В., Липницкий С.Ф., Ярмош Н.А. Объяснительная функция в системе автоматизации обучения // Интеллектуальные системы. Вып. 2. – Мн.: Ин-т техн. кибернетики НАН Беларуси, 1999. – С. 149-156.

30. Липницкий С.Ф. Компьютерный практикум: математическая модель и алгоритмы автоматизированного обучения решению задач // Вестник Белорусского университета. Сер. 1: Физ. Мат. Мех. – 1992. – № 3. – С. 54-56.

31. Кравцов А.А., Липницкий С.Ф., Насуро Д.Р. Создание базы знаний информационной системы на основе тематических корпусов текстов // Докл. Междунар. науч. конф. «Суперкомпьютерные системы и их применение. SSA'2004». – Мн.: ОИПИ НАН Беларуси, 2004. – С. 136-140.

Поступила 25.11.04

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: kravtsov@newman.bas-net.by,
lipn@newman.bas-net.by,
nasuradr@newman.bas-net.by*

A.A. Kravtsov, S.F. Lipnitski, D.R. Nasuro, D.V. Pradun

INTELLECTUALIZATION OF TEXTUAL INFORMATION PROCESSING

Review of UIIP NASB works in the context of search processes intellectualization and textual documents processing is presented. Models of syntactic analysis and semantic interpretation of input sentences, models of syntactic synthesis of output sentences, models of search strategies creation and navigation routes in hypertext, models of hypertext databases creation and actualization are considered.