

величиной верхнего предела (Рис. 5). На графике изображена характерная стабилизация курса акций ( $V=120$ ).

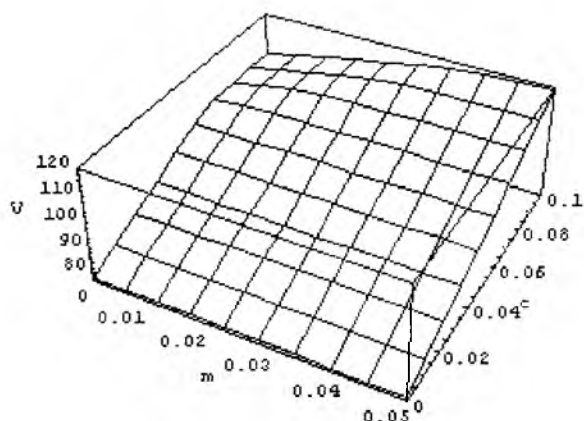


Рисунок 5. График стабилизации курса акций ( $V=120$ )

Таким образом, UnRisk Pricing Engine является универсальным помощником любого экономиста на рынке ценных бумаг. В ней собраны все самые современные и передовые технологии обработки и анализа экономических данных.

#### ЛИТЕРАТУРА

- [1]. Позняк Ю.В., Земсков С.В., Кулешов А.А. Система *Mathematica 4* и возможности ее использования в образовании. Труды Четвертой международной конференции "Новые информационные технологии — НИТе'2000", т.3. Стр. 161-166.
- [2]. UnRisk Pricing Engine for Mathematica. The precision and velocity solution for financial derivatives. 2001. 653p

## МОДЕЛИРОВАНИЕ ПРОЦЕССОВ ИНТЕЛЛЕКТУАЛИЗАЦИИ ИНЖЕНЕРНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

С.Ф. Липницкий, Н.А. Ярмош

Институт технической кибернетики НАН Беларуси, ул. Сурганова, 6, Минск, 220012, тел. (8-017)284-21-47, shi@newman.bas-net.by

### 1. ПРОБЛЕМА ИНТЕЛЛЕКТУАЛИЗАЦИИ ИНЖЕНЕРНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

В условиях интенсивной компьютеризации промышленности постоянно повышаются требования к эффективности информационного обеспечения процессов инженерного проектирования. Это связано главным образом с большими объемами различной по форме представления информации, а также с растущей сложностью и увеличивающимся разнообразием проектируемых объектов. Проследивая историю становления информатики, можно заметить, что до конца 70-х годов развитие автоматизированных информационных систем основывалось преимущественно на технологических усовершенствованиях процедур поиска. В 80-х годах исследователи пришли к убеждению, что дальнейшее повышение эффективности информационных процессов в рамках такого "технологического" подхода уже невозможно. Это подтверждают результаты крупномасштабных экспериментальных исследований эффективности поиска, проводимых в США,

Великобритании и других странах, которые оказывались не выше 50-60 % по суммарному показателю "полнота + точность".

В связи с этим для повышения эффективности информационных процессов необходимо создать новое поколение систем информационного обеспечения, характерной особенностью которых является их интеллектуальность. Разница между обычными (неинтеллектуальными) и интеллектуальными системами заключается в том, что первые работают со сведениями, относящимися к конкретным референтам (например, сведения о проектируемых объектах, документах). Вторые же характеризуются способностью работать с интеллектуальной информацией, т.е. знаниями об объектах или явлениях. В связи с этим и противопоставляются базы данных и базы знаний, интеллектуальные и неинтеллектуальные процессы.

Основные задачи, возникающие при создании упомянутого нового поколения, в совокупности составляют содержание актуальной научной проблемы интеллектуализации инженер-

ных информационных систем. Суть проблемы заключается в выявлении задач интеллектуализации, их формальной постановке и решении.

Несмотря на успехи, достигнутые специалистами стран СНГ, США, Японии и Западной Европы в этой области, проблема интеллектуализации в целом не решена. Это можно объяснить преобладанием эвристических методов исследования, отсутствием единого теоретического подхода к проблеме и сложностью ее формализации. В особенности трудно формализуемы лингвистические понятия, являющиеся основой интеллектуализации информационных процессов.

Отметим две основные особенности проблемы интеллектуализации информационных систем инженерного назначения.

*Первая особенность* - многообразие форм представления информации в инженерных информационных системах. Эта особенность определяет специфические требования к языку общения инженера с информационной системой. С одной стороны, этот язык по своим выразительным возможностям и "дружественности" интерфейса должен приближаться к естественному, а с другой - обеспечивать возможность описания и поиска данных в любой форме представления и с любой степенью детализации.

*Вторая особенность* - высокая интенсивность и большие объемы корректировок в информационной базе. В связи со значительной стоимостью реализации этих процессов создание и актуализация баз данных должны быть в максимальной степени автоматизированы.

Указанные особенности существенным образом влияют на выбор подхода к решению задач интеллектуализации инженерных информационных систем. В первую очередь это касается моделирования языковых процессов, т.е. синтаксического и семантического анализа входной информации и синтеза выходной. Следует отметить, что существующие подходы, принятые при обработке естественно-языковых текстов, здесь не применимы. Это объясняется тем, что морфологическая структура "слов" языка общения инженера с информационной системой не позволяет устанавливать синтаксические зависимости между ними. В связи с этим процессы анализа и синтеза сообщений в инженерной информационной системе должны основываться на формальных процедурах выявления таких зависимостей. Таким образом, проблема интеллектуализации инженерных информационных систем должна быть решена

в такой обобщенной постановке, которая обеспечила бы универсальность разработанных методов, алгоритмов и программ в смысле возможности их использования в различных предметных областях инженерного проектирования. С целью постановки задач интеллектуализации необходимо было провести анализ, типологизацию и декомпозицию задач информационного обеспечения процессов инженерного проектирования. В результате декомпозиции выявлены информационные процессы, имеющие место во всех инженерных информационных системах. Это процессы анализа входных сообщений и синтеза выходных, процессы генерации стратегий поиска и навигационных маршрутов в гипертексте, а также процессы создания и актуализации гипертекстовых баз данных. Всем этим информационным процессам может быть приписан статус интеллектуальных, поскольку при их реализации необходимо использовать базу знаний.

С учетом отмеченных особенностей основные цели моделирования интеллектуальных процессов можно сформулировать следующим образом:

*первая цель* - формирование и формализация совокупности исходных понятий для моделирования интеллектуальных процессов;

*вторая* - исследование в рамках моделирования свойств этих понятий и интеллектуальных процессов;

*третья* - построение и оптимизация алгоритмов реализации интеллектуальных процессов.

С учетом особенностей и целей можно сформулировать задачи интеллектуализации инженерных информационных систем.

## 2. ЗАДАЧА МОДЕЛИРОВАНИЯ ДАННЫХ И ЗНАНИЙ

При решении данной задачи необходимо было разработать систему базовых понятий теории моделирования. Сложность решения заключалась в подборе и формализации этих понятий, поскольку требовалось, чтобы совокупность базовых понятий обеспечивала упомянутый ранее единый теоретический подход к решению всех задач интеллектуализации. Оказалось, что для выявления такой совокупности понятий необходимо было промоделировать структуры данных и знаний, обрабатываемых информационной системой, на основе формализации известных в лингвистике семантических отношений.

Данные в информационных системах инженерного назначения имеют ярко выраженную гипертекстовую структуру в виде описаний объектов и связей между ними. Поэтому для моделирования данных естественно было использовать концепцию гипертекста. 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 предлагается 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. Для контроля эффективности навигационных маршрутов мы ввели три критерия оптимальности продукционного отношения: *представительность, корректность и непротиворечивость*. Представительность означает, что навигационный маршрут, построенный с помощью продукционного отношения, должен содержать все тексты соответствующего тематического раздела гипертекста. Продукционное отношение корректно, если все тексты маршрута упорядочены в соответствии с отношением строгого порядка, определенного на гипертексте. Непротиворечивость соответствует ситуации, когда в навигационном маршруте не присутствуют тексты из других тематических разделов. С учетом этих критериев можно сформулировать формальные требования к совокупности правил продукции.

Основой для моделирования интеллектуальных процессов является база знаний информационной системы. Кроме упомянутых управляющих знаний будем различать *лингвистические и энциклопедические*. В качестве моделей лингвистических знаний нами исследованы синтаксические и семантические деревья сообщений, формализованы понятия входного и внутреннего языков информационной системы. Это позволило ввести в рассмотрение понятия синтаксического и семантического базисов

входного языка как основы для моделирования процесса построения семантических деревьев. Оказалось, что этот процесс сводится к замене поддеревьев синтаксического дерева, являющихся элементами синтаксического базиса, деревьями из семантического базиса входного сообщения. При этом учитывается омонимия, существующая во входном сообщении. Под энциклопедическими понимается часть знаний о предметной области, касающаяся семантики объектов баз данных. В качестве модели энциклопедических знаний нами определена и исследована сеть семантических структур. Это сеть специального вида, узлами которой являются фреймоподобные семантические структуры, а именно семантические примитивы, сцены, эпизоды и сценарии.

### 3. ЗАДАЧА СИНТАКСИЧЕСКОГО И СЕМАНТИЧЕСКОГО АНАЛИЗА ВХОДНЫХ СООБЩЕНИЙ

Синтаксический анализ сообщения можно реализовать путем циклического повторения следующей процедуры: в сообщении ищутся все базовые синтагмы, строятся их синтаксические деревья и исключаются определяющие члены базовых синтагм. Под базовыми понимаются синтагматические структуры, из которых может быть построено любое входное сообщение. С целью алгоритмической реализации этой процедуры необходимо было типологизировать синтагматические структуры входных сообщений. Согласно полученной типологии, при анализе сообщений можно ограничиться рассмотрением только двух-, трех- и четырехсловных синтагматических структур. Доказано, что только типологизированные структуры содержат базовые синтагмы.

Однако для реализации процедур синтаксического анализа входных сообщений недостаточно было установить только условия существования базовых синтагматических структур. Необходимо было выяснить также свойства двух-, трех- и четырехсловных структур, используя которые, эти структуры можно было бы найти в анализируемом сообщении. С этой целью доказаны утверждения, позволяющие выявить в сообщении указанные структуры по совершенно формальным признакам.

Результаты моделирования явились основой для алгоритмизации процессов анализа проективных сообщений. Проективность здесь, как и обычно, означает синтаксическую "правильность" сообщения.

Анализ непроективных сообщений можно реализовать путем использования алгоритмов анализа проективных сообщений. Оказалось, что если непроективное сообщение анализировать с помощью этих алгоритмов, то в результате анализа будет получена совокупность синтаксических деревьев всех проективных подцепочек входного сообщения. 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 сообщения из него необходимо исключить определяющие члены всех его разделенных синтагматических структур и снова применить алгоритмы анализа проективного сообщения. После завершения работы этих алгоритмов и восстановления входного сообщения его анализ можно продолжить путем локализации узловых слов всех определяющих членов разделенных синтагматических структур. Под узловыми словами мы понимаем определяемые члены синтагм сообщения, являющиеся корнями соответствующих синтаксических деревьев.

Семантический анализ сообщений входного языка нами реализован путем формирования и использования прагматического контекста этих сообщений в базе энциклопедических знаний. Анализ осуществляется в два этапа. На первом генерируется псевдосемантическое дерево входного сообщения, а на втором - из вершин этого дерева устраняется лексическая неоднозначность и таким образом строится семантическое дерево.

#### **4. ЗАДАЧА СИНТАКСИЧЕСКОГО СИНТЕЗА ВЫХОДНЫХ СООБЩЕНИЙ**

Для решения этой задачи пришлось ввести понятия семантического и синтаксического покрытий семантического языка. Под семантическим покрытием понимается совокупность орцепей длины 1, позволяющая построить семантическое дерево любого сообщения объединением некоторого подмножества орцепей из этой совокупности. Для определения синтаксического покрытия формализовано понятие синонимии в виде бинарного отношения на множестве синтаксических деревьев: два дерева считаются синонимами в том и только в том случае, если синонимичны соответствующие им сообщения. С учетом отношения синонимии

синтаксическое покрытие определено как совокупность смежных классов фактор-множества множества всех синтаксических деревьев выходного языка по отношению синонимии, являющемуся эквивалентностью. Эта совокупность должна удовлетворять следующему условию: должно существовать такое сюръективное отображение семантического покрытия на синтаксическое, что любое ордерство из семантического покрытия является синонимом своего образа при данном отображении. С использованием понятий семантического и синтаксического покрытий введено понятие синтаксического дерева выходного сообщения как ордерства, полученного в результате объединения всех синтаксических деревьев, взятых по одному из каждого смежного класса синтаксического покрытия соответствующего сообщения семантического языка. Доказано необходимое и достаточное условие существования синтаксических деревьев выходных сообщений (теорема 4.1.1 [1]). Эта теорема имеет конструктивный характер и позволяет алгоритмизировать процесс синтеза синтаксических деревьев.

С целью устранения неоднозначности соответствия между множествами выходных сообщений и их синтаксических деревьев нами определены и исследованы упорядоченные синтаксические деревья. Для определения упорядоченного дерева введены отношение семантической близости и упорядочивающие отображения. Степень семантической близости двух слов в сообщении зависит от расстояния между этими словами. Упорядочивающие отображения ставят в соответствие всем дугам, исходящим из некоторой вершины синтаксического дерева, числовые метки, указывающие степень семантической близости этой вершины и вершин, в которые заходят упомянутые дуги.

На основе упорядоченного дерева строится синтаксически корректное выходное сообщение. Формально под синтаксически корректным сообщением мы понимаем цепочку, все базовые синтагматические структуры которой синтаксически корректны. В свою очередь, базовая структура корректна, если с увеличением расстояния между ее узловым словом и узловыми словами ее подцепочек степень семантической связи между ними не уменьшается.

#### **5. ЗАДАЧА ГЕНЕРАЦИИ СТРАТЕГИЙ ИНФОРМАЦИОННОГО ПОИСКА**

Для решения этой задачи нами обобщены известные в информатике критерии оптималь-

ности 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. В результате моделирования поиска получена поисковая функция (5.2.4) [1], оптимизирующая поиск не только в смысле релевантности, т.е. соответствия найденной информации запросу, но также и пертинентности, т.е. соответствия этой информации информационной потребности. По пертинентности поиск оптимизируется в том случае, если запрос пользователя сформулирован корректно, т.е. отражает его информационную потребность. Корректность достигается за счет итерационности процедуры поиска.

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. Удалось доказать 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.

## 6. ЗАДАЧА НАВИГАЦИИ В ГИПЕРТЕКСТЕ

Процедура создания навигационного маршрута сводится к построению продукционного отношения. Эта процедура может включать один или более сеансов взаимодействия пользователя с гипертекстом. Например, при проведении тематического поиска в базе данных компьютерной энциклопедии линейное упорядочение соответствующей совокупности текстов может быть осуществлено за один сеанс, а

при работе с компьютерным тренажером навигационный маршрут строится пошагово с участием пользователя. В обоих случаях необходимо найти совокупность текстов, каждый из которых может следовать за уже сформированной частью маршрута.

## 7. ЗАДАЧА СОЗДАНИЯ И АКТУАЛИЗАЦИИ ГИПЕРТЕКСТА

Для решения этой задачи необходимо было ввести формальные понятия фрагмента и сегмента текста. Фрагмент - это произвольный кортеж сообщений 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 сообщений этого фрагмента, но не включающий графы-ответы на семантические деревья сообщений, 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. При этом используется база энциклопедических знаний.

## 8. НАПРАВЛЕНИЕ ДАЛЬНЕЙШИХ ИССЛЕДОВАНИЙ

Дальнейшие работы в области моделирования процессов интеллектуализации инженерных информационных систем будут проводиться в направлении формализации и исследования процессов создания и актуализации инженерных баз знаний на основе автоматизированного накопления и обобщения опыта инженерного проектирования.

## ЛИТЕРАТУРА

- [1] Липницкий С.Ф., Ярмош Н.А. Моделирование интеллектуальных процессов в инженерных информационных системах. Минск: Беларуская навука, 1996. 222 с.