

виде графического файла топонимической карты региона, которая также посредством ссылок была бы связана с основной картотекой.

Подводя итог, можно сказать, что сегодня вести речь о некоем прорыве в методике топонимических исследований, разумеется, преждевременно. В этом направлении делаются лишь первые шаги, а конкретные результаты дадут о себе знать, скорее всего, не ранее чем через три-четыре года. Тем не менее, если наши попытки окажутся успешными, данная методика позволит, наконец, объединить усилия всех лингвистов-исследователей Приуральского региона и создать единую электронную топонимическую картотеку бассейна Камы, которая впоследствии сможет найти реализацию и в виде аналога уже упомянутого труда Г. П. Смолицкой, и в виде части комплексной программы «Духовная культура Урала», реализация которой уже на протяжении двух десятков лет ведется в Уральском государственном университете.

Список литературы

- Программа 1976 — Программа сбора материала по региональной лингвистике. (Топонимия) / сост. А. Е. Контарь. — Свердловск, 1987. — 14 с.
- Смолицкая 1976 — *Смолицкая Г. П.* Гидронимия бассейна Оки. (Список рек и озёр) / Г. А. Смолицкая. — М., 1976. — 403 с.
- Шакуров 1973 — *Шакуров Р. З.* Топонимия бассейна реки Дёмы. автореф. дис. ... канд. филол. наук / Р. З. Шакуров. — М., 1973. — 22 с.
- Полякова 1983 — *Полякова Е. Н.* Об особенностях составления топонимической картотеки по пермским деловым памятникам XVI — начала XVIII века / Е. Н. Полякова // Методы топонимических исследований. — Свердловск, 1983. — С. 124-130.
- Матвеев 1974 — *Матвеев А. К.* Тезисы о топонимике / А. К. Матвеев // Вопр. ономастики. — № 7. — Свердловск, 1974, — С. 5–18.
- Матвеев 1987 — *Матвеев А. К.* Топонимические древности / А. К. Матвеев // Формирование и развитие топонимии. — Свердловск, 1987. — С. 4–28.
- Кривощёкова-Гантман 1976 — *Кривощёкова-Гантман А. С.* Структурно-словообразовательные типы коми-пермяцких ойконимов // Вопросы лингвистического краеведения — Вып. 2. — Пермь, 1976. — С. 47–55.
- Трубе 1976 — *Трубе Л. Л.* Отражение хозяйственной деятельности в топонимии Горьковской области // Ономастика Поволжья. — Сб. 4. — Саранск, 1976. — С. 269–271.

Обучающие системы. Контроль семантических противоречий

В. Н. Кучуганов

Ижевский государственный технический университет, Россия

Now it is considered perspective to apply ontology's during development of interactive courseware This work offers an approach to the problem of semantic contradictions and semantic variability reductions during the semantic analysis of textbook definitions, texts and answers tested. It is realized with the help of the logic search functions determined on subject ontology and logic synthesis of a plot circuit.

Введение

В настоящее время считается весьма перспективным [Кафтанников и др. 2003; Исмаилова и др. 2004] применение онтологий при разработке интерактивных обучающих систем (ИОС). Онтология — это множество формализованных описаний терминов предметной области и отношений между ними. Онтология может быть использована в принятии решений при реорганизации и построении информационных систем предприятий, для обмена информацией, для обучения [Ной 2001; Овдей 2004].

Применение предметной онтологии повышает корректность тезауруса и определений, что способствует повышению качества обучения. Однако системы для разработки онтологий в основном ориентированы на автоматизацию проектирования баз данных. Семантический анализ определений, написанных на естественном языке, пока является для них серьезной проблемой, в то время как для ИОС эта задача важна не только на этапе разработки, но и в ходе проверки контрольных заданий.

В работе предлагается подход к задаче разрешения семантических противоречий и сокращения смысловой вариативности в процессе семантического анализа текстов определений и ответов тестируемого с помощью логических поисковых функций, опре-

деленных на предметной онтологии, и логического синтеза схемы сюжета.

В качестве учебного инструмента разработки онтологий предлагается система KG (Knowledge Guide-book), разработанная на кафедре «Автоматизированные системы обработки информации и управления» Ижевского государственного технического университета [Кучуганов и др. 2001]. Система ориентирована на эксперта, не владеющего программированием. Благодаря большей степени конкретизации смысла тех или иных понятий в предметной области (ПО) и отношений между ними, использование KG в учебном процессе позволяет облегчить обучение студентов общим принципам проектирования онтологий и их использованию в практических задачах.

1. Семантические категории предметной онтологии

Под онтологией условимся понимать иерархически классифицированное множество понятий (концептов), представленных тетрадами

Термин — Определение — Модель — Атрибуты,

где: *Термин* — слово или устойчивое словосочетание, имеющее понятие в данной ПО;

Определение — такое словесное описание понятия, что все его ключевые слова определены в этой же онтологии либо в некоторой универсальной;

Модель — вычислимый предикат, формула, система уравнений, схема, образец и т. п., позволяющие получать множество экземпляров, удовлетворяющих данному понятию;

Атрибуты — список атрибутов класса (модели).

Отношения между понятиями определяются через их атрибуты и модели.

В системе KG управления базами знаний имеется *три* базовые физические конструкции для хранения информации [Кучуганов 2002]: *дерево* — концепт — экземпляр и *пять* семантических категорий для наполнения концептов и экземпляров: предмет; процесс; свойство; отношение; эпизод.

Свойства описываются в *разделе знаний о свойствах* материалов, предметов, процессов, который представляет собой классификационное дерево свойств (атрибутов), сгруппированных по назначению, физическим, физиологическим, психофизическим,

социологическим и прочим критериям. Это позволяет унифицировать понятия свойств, использовать их многократно и, самое важное, сопоставлять объекты из разных классификационных деревьев и предметных областей.

Предметами здесь считаются материальные и виртуальные объекты: стол, автомобиль, робот, человек, документ и т. п. Предметы имеют состав (детали) и схему соединения, то есть описываются геометрическими моделями: кинематическая схема, чертеж, карта, изображение, 3D геометрическая модель, экранная форма.

Процесс отображает в базе знаний деятельность коллектива исполнителей в некотором отрезке времени и пространства. В общем случае процессы имеют состав и алгоритм, то есть их описание содержит вычислительные модели. Именно деятельностный процесс порождает множество связей между объектами, поэтому характерной особенностью списка атрибутов процесса является то, что он, помимо общих характеристик, содержит атрибуты, описывающие *роли участников действия*: агент (исполнитель); бенефициант — заказчик, в чьих интересах выполняется действие; реципиент — приемник действия; предмет воздействия: исходный/результатирующий; сцена действия; инструмент; коагент (соисполнитель) и т. д.

Отношения устанавливают факты наличия разнообразных взаимосвязей между объектами. Отношения — это наиболее динамичный раздел базы знаний в том смысле, что они постоянно меняются в ходе осуществления какой-либо деятельности. При описании предметной области и ее бизнес-процессов имеет смысл определить необходимые виды отношений и способы их исчисления, а не хранить постоянно сами эти факты. Тогда методы исчисления отношений можно включать в состав процессов анализа ситуаций и принятия решений.

В системе KG предопределены *четыре* семантически не пересекающиеся категории отношений:

– сравнение/сопоставление — это сравнение значений свойств (предметов, процессов, отношений) и сопоставление графов (предметов/процессов);

– вхождение (в множество, класс, экземпляр);

– деятельностные (ролевые, причинно-следственные, вычислительные) отношения, извлекаемые из экземпляров процессов на основе их концептов;

– коммуникации (толерантности) — личное отношение к другому субъекту или, в общем случае, к объекту, входящему в другую категорию.

Эпизод — это связный полихромный граф, описывающий сюжет (бытовой, производственный и проч.), ограниченный по количеству процессов, участников и времени.

Эпизод/сюжет отличается от составного процесса отсутствием общей целенаправленности подпроцессов и относительной случайностью исполнителей при единстве сцены, в которой эти подпроцессы выполняются. Соответственно, классификация эпизодов возможна только по их главным исполнителям или главным процессам и месту действия. Необходимость введения семантической категории «эпизод» возникла как простой и экономный способ запоминания личного опыта и обучения, хорошо сочетающийся с природой ассоциативной памяти и образного мышления.

Эпизоды-сообщения и отношения на эпизодах. Рассмотрим класс коммуникативных (чувственных) глаголов, выражающих процессы общения и восприятия через органы чувств человека: *говорить, видеть, слышать, думать, читать, писать, просить, требовать, улыбаться* и т. д. Результатом этих действий является сообщение — некоторая информация в той или иной форме, в том числе шум — бесполезная или непонятая информация. Кроме того, *форма сообщения* выражает *отношение актора* к реципиенту тоном, манерой, словом, а также *установку* к сообщаемой информации с помощью атрибута *модальности* (хотеть, мочь, долженствовать) и *наклонения* (повелительное, изъявительное, сослагательное).

Пример. “*Don't just sit around. Come and help me*”.

Здесь мы видим эпизод в эпизоде:

1. Главный герой и второстепенный, один работает, другой сидит.

2. Эпизод, передаваемый прямой речью, в котором говорящий пополняет информацию о первом эпизоде (выражает свое неудовольствие) и выдает указание, согласно которому второе действующее лицо должно стать активным (закончить отдыхать и помочь первому).

Таким образом, семантическое описание каждой сущности в онтологии предметной области — это объединение имени, текстового определения, набора атрибутов и модели:

– предмет представлен геометрической моделью в виде однородного графа (все вершины имеют одинаковый набор атрибутов);

– процесс — вычислительной (алгоритмической) моделью;

– эпизод представлен полихромным графом, в котором вершины имеют различные наборы атрибутов.

Описанная модель знаний позволяет анализировать достаточно широкий спектр отношений между объектами предметной области и бизнес-процессов и, тем самым, уменьшить влияние субъективизма при разработке онтологий.

2. Алгоритмы поиска сущностей

Контроль корректности определений в предметной онтологии предлагается осуществлять путем извлечения семантических значений ключевых слов с помощью *логических поисковых функций* и логического синтеза схемы сюжета. Результатом поиска являются факты, найденные или вычисленные вспомогательными процедурами.

Слова, обозначающие предметы, процессы, состояния, ситуации, отношения, имеют в базе знаний, в общем случае, несколько значений, каждое из которых определяется типовой системой отношений (связей) с другими понятиями. После морфологического и синтаксического анализа текста [Кучуганов 2005] сокращение семантической многозначности решается путем исключения противоречивых фактов.

Пример 1. “*Я выхожу на следующей остановке*”.

Выходить [из]:

1) процесс/ходьба;

2) актер: предмет (одушевленный — 'Я');

3) направление (откуда? куда?):

– вариант 1 — сцена (дом, комната, лес, транспорт);

– вариант 2 — состав участников процесса (игра, бизнес);

– вариант 3 — ситуация;

4) время (настоящее, прошедшее, будущее).

Остановка (следующая):

1. Сцена — специально оборудованное место для временной приостановки движения транспорта с целью входа/выхода пассажиров и(или) груза.

2. Процесс (фаза) движения.

Остановка как фаза движения имеет систему отношений: кто/что; где; когда; для чего/почему; каким образом.

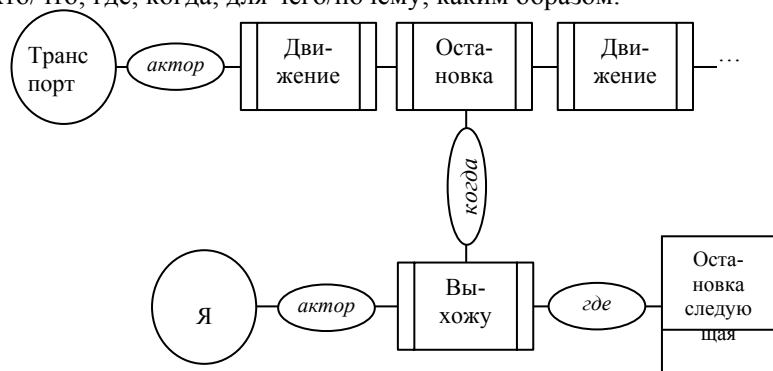


Рис. 1. Граф сюжета

Отсюда видим, что процесс «остановка» как место вызывает ассоциацию с «транспортом» и «выходить» и, как всякий другой процесс, должен быть привязан к шкале времени. В данном случае он связывается с движением транспорта, фазой приостановки. Таким образом, здесь «остановка» — это одновременно и фаза движения и место в пространстве, а атрибут «следующая» говорит о том, что действие произойдет в будущем. На рис. 1 показан граф сюжета, а на рис. 2 — временная диаграмма его процессов.

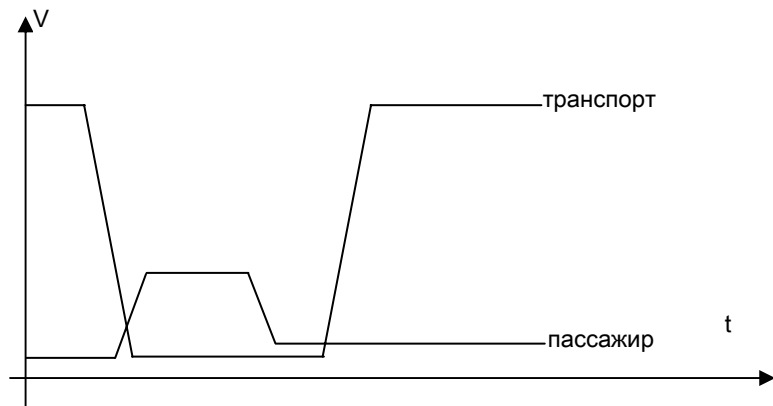


Рис. 2. Временная диаграмма сюжета

В предложении «Остановка автобуса находится за углом» слово «остановка» однозначно указывает на место расположения.

В следующем отрывке известной песни «Пока еще не поздно мне сделать остановку, кондуктор, нажми на тормоза» слово «остановка», очевидно, фаза процесса, хотя и не сказано, какого именно (здесь рефлексия — ассоциирование движения поезда с образом жизни).

Пример 2. “I must fly now. She hates waiting around.”

“Fly”: 1. Имя существительное — летающее насекомое.

2. Глагол:

- 1) летать каким-либо воздушным видом транспорта;
- 2) пилотировать, управлять каким-либо летательным аппаратом;
- 3) летать, разлетаться самостоятельно под воздействием какой-то силы или по инерции;
- 4) переносное — двигаться быстро.

Как имя существительное “fly” противоречит структуре простого предложения, так как следует за “must”. В значении глагола можно отбросить 3-й вариант, потому что имеется атрибут долженствования, но его можно и сохранить для уточнения по расширенному контексту.

“Wait around” (устойчивое словосочетание) — слоняться в ожидании чего-либо.

Тогда с учетом второго предложения «она очень не любит ждать» для глагола “fly” остается единственный вариант — двигаться быстро (торопиться).

Для синтеза схемы сюжета предлагается использовать логику схем программ, позволяющую эффективно синтезировать схемы действий, использующих ограниченные ресурсы [Непейвода 1989; Kuchuganov 1994]). Определение считается корректным, если схема построена и ее граф связан.

Заключение

Предложенный подход апробировался в процессе разработки интерактивных обучающих систем (ИОС) по английскому языку, бухгалтерскому консалтингу, информационным технологиям. Сокращение вариативности смысла ЕЯ-текста может быть достигнуто путем ужесточения требований к качеству предметных онтологий и их тезаурусов, а также благоприятному сочетанию

механизмов логического и образного мышления, то есть логического вывода и функций поиска/распознавания.

Список литературы

- Кафтанников и др. 2003 — *Кафтанников, И. Л.* Перспективы использования онтологий в учебном процессе / И. Л. Кафтанников, С. Е. Корovin // *Educational Technology and Society*. — 2003. — № 6 (3).
- Исмаилова и др. 2004 — *Исмаилова, Л. Ю.* В поисках утраченной онтологии. Категорный подход / Л. Ю. Исмаилова, С. В. Косиков // *Технологии информационного общества — Интернет и современное общество*: тр. VII Всерос. объединенной конф. (Санкт-Петербург, 10–12 нояб. 2004 г.) — СПб.: Изд-во Филол. ф-та СПбГУ, 2004. — С. 56–61.
- Ной и др. — *Ной, Н. Ф.* Разработка онтологий 101: Руководство по созданию Вашей первой онтологии / Н. Ф. Ной, Д. Л. МакГиннесс; Стэнфордский ун-т. — Стэнфорд (Калифорния). — 94305.
- Овдей и др. 2004 — *Овдей, О. М.* Обзор инструментов инженерии онтологий / О. М. Овдей, Г. Ю. Проскудина // *Электронные библиотеки: перспективные методы и технологии, электронные коллекции*: тр. 6-ой Всерос. науч. конф. — RCDL2004. — Пушкино, 2004.
- Кучуганов и др. 2001 — *Кучуганов, В. Н.* Система визуального проектирования баз знаний / В. Н. Кучуганов, И. Н. Габдрахманов // *Информационные технологии в инновационных проектах*: тр. III междунар. науч.-техн. конф. — Ижевск, 2001. — С. 140–143.
- Кучуганов 2002 — *Кучуганов, В. Н.* Семантика графической информации // *Известия ТРТУ*: Тематич. вып. «Интеллектуальные САПР»: материалы междунар. науч.-техн. конф. — Таганрог: Изд-во ТРТУ, 2002. — № 3 (26). — С. 157–166.
- Кучуганов 2005 — *Кучуганов, В. Н.* Визуальное моделирование текстов / В. Н. Кучуганов // «Интеллектуальные системы» (AIS'05) и «Интеллектуальные САПР» (CAD–2005): тр. междунар. науч.-технич. конф. — М.: ФИЗМАТЛИТ, 2005. — Т. 4. — С. 104–114.
- Непейвода 1989 — *Непейвода, Н. Н.* Некоторые семантические конструкции конструктивных логик схем программ / Н. Н. Непейвода // *Вычислительные системы*. — Вып. 129. — Новосибирск, 1989. — С. 49–66.
- Kuchuganov 1994 — *Kuchuganov, M.* A predicate logic of well-founded actions / M. Kuchuganov // *Lecture Notes in Computer Science*. — Vol. 813. — 1994. — P. 218–226.

Распознавание рукописных текстов

А. В. Кучуганов, Г. В. Лапинская
Ижевский государственный технический университет, Ижевск,
Россия

The problem of text recognition in documents processing and information input into the computer is rather actual. Our work offers an approach to the problem solution of hand-written text recognition, consisting of that the scanned sample of the hand-written text vectorized, initial elements of the received image are represented as the count which elements are compared to the reference samples stored in base.

The described model is incorporated in the working version of program HWTR (hand-written text recognition) for recognition of hand-written text.

При обработке документов и вводе информации в компьютер весьма актуальной является задача распознавания текста. Распознавание печатных текстов компьютером — область, сегодня достаточно хорошо исследованная. Что же касается распознавания рукописного текста, то качественный рывок еще впереди. Хотя уже есть несколько отработанных направлений. Во-первых, это системы распознавания форм, заполненных печатными буквами от руки, которые применяются во многих областях. Во-вторых, это распознавание отдельных рукописных букв, написанных особым пером на специальном экране (touch-screen), которое широко применяется в карманных компьютерах и электронных записных книжках. Эти распознающие системы демонстрируют достаточно высокую точность, приближающуюся к точности клавиатуры.

Чтение компьютером слитных букв, то есть обычного письма, сегодня мало разработано. Но исследовательские проблемы, которые надо решить на этом пути, чрезвычайно интересны.

Типичная задача автоматического распознавания образов формулируется примерно так: все множество подлежащих обработке изображений некоторым способом разбивается на конечное число классов, называемых образами [Бутаков 1987]. Автоматическому устройству, снабженному механизмом восприятия образов, предъ-