

ОПЫТ АНАЛИЗА ТЕРМИНОВ ПРОБЛЕМНОЙ ОБЛАСТИ ПРИ ПОСТРОЕНИИ ЭКСПЕРТНОЙ СИСТЕМЫ

Найденова К. А.
Военно-медицинская академия,
naidenova@mail.spbnit.ru

Создание экспертных систем для распознавания по космо-аэро-снимкам комплексных природных образований, таких как рельеф, ландшафт, геологические структуры земной коры, рудные месторождения и т. п., потребовало создания баз знаний для описания этих объектов на языке специалистов соответствующей проблемной области. Знания, заключенные в языке специалиста, аккумулируют, отражают как структуру обозначаемых объектов, так и способы рассуждений, с помощью которых создавались терминологическая база проблемной области и специфические языковые конструкторы для описания природных образований.

Как таковые, знания, отображенные в структуре терминов и их сочетаниях, дают возможность делать выводы, рассуждать о проблемной области.

В настоящей работе рассматривается опыт анализа терминологической базы для определения и распознавания по снимкам типа равнины по геоморфологическим и ландшафтным признакам. Предварительный анализ терминов и знаний специалистов производился по текстам из монографии «Геологическое строение СССР».

Терминологический анализ состоял из следующих этапов:

1. Образование словаря терминов;
2. Выявление способов образования терминов;
3. Выявление семантической структуры терминов;
4. Классификация терминов по сходному геологическому содержанию;
5. Выявление структуры описания сложных объектов на основе терминов и их сочетаний.

Затем было необходимо:

1. Выбрать форму представления знаний в экспертной системе;
2. Сформулировать правила для дешифрирования типов равнинных ландшафтов.

Далее мы кратко остановимся лишь на некоторых аспектах терминологического анализа.

Термины проблемной области и их содержание

Прежде всего мы расклассифицировали термины по их содержанию, то есть с точки зрения объектов и отношений внеязыковой реальности, для называния которых термины служат. Итак, содержательно термины описывают следующие реалии в данной проблемной области.

Природные объекты: рельеф, ландшафт, элементы рельефа, элементы ландшафта, отложения, осадки, минералы, горные породы, классы горных пород и минералов, геоморфоструктуры и т. д.

Процессы, формирующие рельеф, ландшафт, горные породы, геоморфоструктуры и другие природные образования.

Временные периоды действия процессов или их конечных результатов (существование объектов во времени).

Формы объектов.

Пространственное положение объектов.

Свойства объектов и степени выраженности свойств.

В таблицах 1, 2, 3, 4, 5 даны некоторые фрагменты словаря терминов с указанием их содержания.

Таблица 1. Термины, обозначающие процесс

Термин	Содержание
Аккумуляция	Процесс, формирующий или оказывающий влияние на образование рельефа в целом или некоторого элемента рельефа, способствующий появлению определённых геоморфоструктур
Выветривание	
Выравнивание	
Денудация	
Оледенение	
Опускание	
Осадконакопление	
Поднятие	
Понижение	
Прогибание	
Сглаживание	
Таяние	
Трансгрессия	
Эрозия	

Отметим, что один и тот же термин (слово) может иметь разное содержание. Например, термин, обозначающий процесс, может также обозначать и видимый результат процесса, например, “опускание” есть процесс опускания места, но может быть и обозначением самого места более низкого по сравнению с окружающим ландшафтом.

Таблица 2. Термины, обозначающие элементы рельефа

Термин	Содержание
Болото	Элементы рельефа
Возвышенность	
Гривна	
Гряда	
Дюна	
Западина	
Котловина	
Ледник	
Морена	
Овраг	
Озеро	
Плато	
Равнина	
Склон	
Терраса	
Увал	
Холм	

Термины, обозначающие элементы рельефа, являются словообразующими для конструирования терминов обозначающих формы рельефа. Такую же образующую роль по отношению к формам рельефа играют термины, обозначающие процессы. Например: если ранее происходил процесс таяния ледниковых вод, то рельеф «водно-ледниковый», если в рельефе преобладают гряды и холмы, то рельеф «холмисто-грядовый». Здесь мы имеем уже сложные термины, образованные из двух и более компонентов и, между компонентами этих сложных терминов

реализуются различные по содержанию отношения: отношение «часть-целое» (рельеф, состоящий из холмов и гряд), отношение «быть образованным с помощью процесса», «быть результатом процесса» (рельеф, образованный ледником). Сложные термины подразумевают, что специалист неявно использует правила рассуждения вида «если – то». Например, знание, что некоторый процесс порождает определенный результат, дает возможность сделать непосредственный вывод в форме продукционного правила: «если оледенение относится к четвертичному периоду, то оно является континентально - покровным».

Процессы могут сочетаться или сосуществовать друг с другом, могут идти один за другим, чередуясь и частично перекрываясь. Поэтому для описания формы рельефа часто необходимо оперировать сложными терминами, состоящими из двух и более атрибутивных компонентов.

Таблица 3. Термины, обозначающие формы рельефа

Термин	Содержание
Аккумулятивный	Формы рельефа
Бугристо- грядовый	
Бугристо- западинный	
Водно-ледниковый	
Грядовый	
Гривный	
Ледниковый	
Мерзлотный	
Монотонный	
Морской аккумуляции	
Плоский	
Плоско-волнистый	
Плоско-холмистый	
Увалистый	
Холмисто-моренный	
Холмисто-увалистый	
Эоловый	

Таблица 4. Термины, обозначающие временные периоды

Термин	Содержание
Доледниковый	Временные периоды, ассоциируются с процессами и с результатами процессов, которые произошли в определенный период, – осадками, отложениями, элементами рельефа, геоморфоструктурами
Кайнозойский	
Ледниковый	
Межледниковый	
Мезозойский	
Мезозойско-кайнозойский	
Меловой	
Нижнечетвертичный	
Пермский	
Плейстоценовый	
Плиоценчетвертичный	
Позднечетвертичный	
Послеледниковый	
Третичный	
Четвертичный	

Временная привязка объектов или процессов осуществляется двумя способами:

- непосредственно, с помощью геологических шкал,

- опосредованно, то есть путём соотнесения с каким-либо процессом или местом или объектом, для которого временная привязка известна.

Эти два способа временной привязки часто употребляются одновременно, например, “доледниковый четвертичный”. Здесь привязка по времени осуществляется двумя способами – с помощью геологической шкалы времени и с помощью соотнесения с процессом оледенения. Приставки “до”, “после” (по отношению к процессу), “поздне”, “ранне”, “нижне”, “верхне” уточняют временной интервал процесса или геологического объекта: “позднечетвертичные осадки”, “третичные породы”, “позднечетвертичное оледенение”.

Термины, обозначающие геологические породы, элементы рельефа и процессы накопления осадков, являются словообразующими для конструирования терминов обозначающих характер геологических отложений.

Таблица 5. Термины, обозначающие характер отложений

ТЕРМИН	СОДЕРЖАНИЕ
Аллювиальные	Характер отложений или осадков, отражающий их состав и процессы их формирования
Валунно-галечные	
Глинистые	
Ледниковые	
Моренные	
Песчаные	
Флювиогляциальные	

Существуют термины, выражающие отношение сосуществования двух и более объектов во времени и пространстве. В качестве таких терминов выступают чаще всего глаголы. Эти отношения нередко сопровождаются оценкой степени проявления (хорошо, ярко, четко, наиболее, преимущественно).

Существуют термины выражающие именуемое отношение, отношение «объект – свойство», отношение сравнения, степень выраженности отношения, свойства (качества), степень уверенности (типично, часто, редко, всегда, никогда).

Чаще всего для выражения различных по смыслу отношений употребляется одна и та же форма выражения естественного языка, а именно атрибутивная или определительная форма. Атрибуты могут быть образованы от различных терминов, означающих процессы, периоды времени, элементы рельефа, осадки, местоположение в пространстве, качество. Объект может иметь несколько атрибутов. Например, «длительное денудационное выравнивание», «избирательная денудация», «значительные новейшие поднятия».

Таблица 6. Термины, обозначающие сосуществование во времени и пространстве

Термин	Содержание
Выражены	Отношение сосуществования во времени и (или) пространстве нескольких объектов. Сопровождается оценкой степени проявления
Наблюдаются	
Приурочены	
Развиты	
Распространены	
Сопряжены	
Типичны	
Характеризуются	

Атрибуты количественных и качественных оценок универсальны, они могут сочетаться с любыми по содержанию терминами: длинный, короткий, крутой, главный, избирательный. Часть из них может быть выражена с помощью количественной меры, часть – только с помощью линейно или частично упорядоченных шкал оценок.

Некоторые атрибуты сочетаются только с определёнными терминами, образуя устойчивые словосочетания: «спокойное течение», «спокойная тектоника», «крутой» (пологий) склон”.

Атрибутивные отношения могут быть выражены не только прилагательными, но и существительными с предлогом или без него: «высота склона», «склон с высотой», «колебания высоты».

Анализ содержания терминов и способов образования сложных терминов показывает, что их семантическая (содержательная) структура соответствует отношениям между природными явлениями (процессами, объектами),

так как их понимают специалисты. Семантическая основа терминов, как и знаний специалистов, есть внеязыковая реальность.

Основной структурной составляющей знаний является понятие геологического процесса. Процесс происходит во времени и пространстве, имеет определённый результат, который представляет собой видимую картину природного ландшафта, частями которого являются различные элементы, имеющие разнообразные формы и разнообразное распределение в пространстве. Так как геологические процессы охватывают не только поверхность, а всю земную кору (видимое на поверхности обусловлено и глубинными процессами), то они определяют также глубинное строение земной коры, форму пластов, складок, состав и свойства пород и т. д. Именно эти знания о геологических процессах позволяют осуществлять «перенесение» времени процессов и некоторых их свойств на сформированные этими процессами объекты. Также возможно «перенесение» свойств элементов рельефа на рельеф, содержащий в себе эти элементы.

Модель знаний предметной области объектно-ориентирована. На рисунке 1 дан фрагмент структуры знаний проблемной области, отображающий отношения между процессами, формами рельефа и геоморфоструктурами. Такие отношения как «часть – целое» и «объект – свойство» могут определяться рекурсивно.

Каждый объект предметной области может быть представлен своей специфической объектно-ориентированной схемой. Например, осадки атрибутируются по составу, характеру залегания, мощности отложений и временному диапазону их образования. Рельеф описывается через его элементы, характеризуется степенью его расчленённости, его морфологическими признаками. Но в системе знаний нет ничего обособленного. Система знаний специалиста содержит знания о взаимосвязях между объектами, о взаимосвязях между свойствами одного и того же объекта и между свойствами различных объектов. Эти связи представляются в виде логических правил, позволяющих выводить необходимые следствия из известных фактов или исключать невозможные, несовместимые со знанием значения объектов и свойств.

Равнины определяются как единство рельефа и осадочных отложений. Единство рельефа и отложений осуществляется в природе через процессы, формирующие их. Противоречивость описаний некоторых частей общей геоморфологической картины выявляется при проверке сочетаемости различных следствий известных процессов, их согласованности друг с другом.

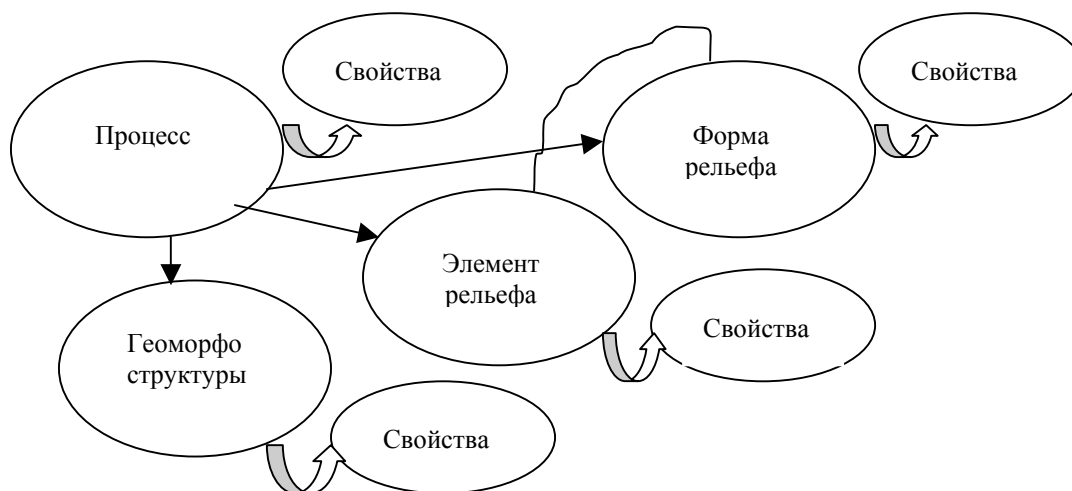


Рис.1. Фрагмент структуры знаний проблемной области

Обозначения: \rightarrow - отношение «имеет результат»; \curvearrowright - отношение «объект – свойства»; \sim - отношение «часть – целое».

Сочетание неязыковых и языковых форм представления знаний

В экспертных системах есть два процесса. Один заключается в распознавании объектов и их свойств по их непосредственным не языковым описаниям (аэро-космо-снимки, набор измерений и результатов тестов). Этот процесс имеет результатом описание распознаваемых объектов на языке более высокого уровня, например, на

языке логики предикатов первого порядка, что широко применяется на практике. Если распознавание успешно завершено, то описание распознанных объектов на естественном языке идеологически осуществить не так трудно. Второй процесс заключается в формировании по тексту на естественном языке тех объектов и отношений предметной области, то есть тех проблемных знаний, которые «несёт в себе» этот текст. Этот процесс неизмеримо труднее первого, и едва ли возможно его полностью автоматизировать.

Понимание проблемно-ориентированного ЕЯ-текста основано на знании семантической структуры терминов, как простых, так и сложных, и семантических отношений между ними. Проблеме понимания ЕЯ-текста и его перевода на язык семантических отношений проблемной области посвящено немало работ. Суть перехода от текста на естественном языке к фрагменту модели предметной области, который лежит в основе сгенерированного текста, сформулирована многими авторами, мы воспользуемся описанием этого процесса, данным в работе [1]. Сначала происходит декомпозиция предложения на фрагменты или сегменты на основе грамматики ЕЯ и критериев связности слов во фрагментах. Затем происходит интерпретация фрагментов на модели предметной области. Традиционно анализ текста поручается на первом этапе лингвистической подсистеме. Лингвистическая подсистема включает модули предварительной обработки текста, морфологический анализатор, построитель сегментов, семантико-синтаксический анализатор (см., например, [2]). Лингвистическая система выделяет интерпретируемые фрагменты предложения. Если учесть, что уже на уровне терминов наблюдается их неоднозначное толкование, и что говорящий использует знания о причинно-следственных зависимостях в проблемной области, то задача выделения интерпретируемых фрагментов оказывается чрезвычайно сложной.

В какой мере и как человек использует модель знаний о предметной области при анализе текста на ЕЯ? Как взаимодействуют его знания о предметной области и о том, как конструируются предложения на ЕЯ? Какие знания доминируют при понимании текста на ЕЯ? Несмотря на большой опыт, накопленный специалистами в области построения систем понимания текстов на ЕЯ, теоретически поставленные выше вопросы остаются не выясненными.

Можно предположить, что те или иные термины в предложении сразу, непосредственно активизируют соответствующие этим терминам фрагменты системы знаний, и активизируемые проблемные знания начинают управлять членением предложения на сегменты. С точки зрения психолингвистики, знания представляют целостную систему, они не изолированы друг от друга. Вопрос главный – как различные знания взаимодействуют в системе.

Рассмотрим простой пример. Пусть в тексте встретилось выражение «рельеф морской аккумуляции». Термин «аккумуляция» активизирует объектно-ориентированную схему знаний о процессе аккумуляции и о связи процесса аккумуляции с формируемым этим процессом рельефом. Теперь, на основе знаний о предметной области, мы можем выделить фрагмент предложения «рельеф аккумуляции». Знание о типах аккумуляции включает морскую аккумуляцию. В системе знаний «морская аккумуляция» есть либо «аккумуляция» плюс её свойство «морская», либо класс аккумуляции. Интерпретация предложения на уровне знаний получена без обращения к лингвистическим знаниям: (объект «это есть» рельеф, сформированный под влиянием (объект «это есть» процесс «имеет имя» аккумуляция, «имеет свойство» морская)).

Это очень простой пример, но в сложных случаях первоначальный лингвистический анализ может приводить к необходимости рассмотрения большого количества возможных интерпретаций. Обращение к модели предметной области позволит сократить область поиска интерпретации. Устойчивые терминологические сочетания есть «фреймы» - готовые фрагменты вывода. Классификационные структуры в модели знаний есть основа рассуждения при поиске интерпретаций. Связи от объектов к их классам, от классов к свойствам объектов, связи между классами объектов по отношению включения, связи между свойствами объектов дают возможность конструировать объектно-ориентированную модель – интерпретацию предложения на ЕЯ. Объектно-ориентированная модель знаний легко укладывается в модель предложения, которая по существу также является объектно-ориентированной. Мы далеки от мысли, что можно вообще обойтись без лингвистического анализа предложений. Однако нам кажется, что определяющим моментом при понимании текстов на естественном языке является модель предметной области, а в более широком аспекте – модель знаний о мире. Чем совершеннее модель знаний, тем успешнее и точнее может происходить интерпретация естественно языковых текстов.

Список литературы

1. Хахалин Г.К. Лингвистический транслятор в семействе систем с обработкой ЕЯ-текстов (ретроспекция). //Сборник научных трудов 6-ой национальной конференции по искусственному интеллекту (КИИ-98) с международным участием / Пушкино: 1998. Том 1. С. 238 – 246.
2. Невзорова О.А., Федун Б.Е. Система подготовки и анализа технических тестов «ЛОТА»: структурно-функциональная схема и модель графического представления текста. //Сборник научных трудов 7-ой

национальной конференции по искусственному интеллекту (КИИ-2000) с международным участием / Москва: ИФМЛ, 2000. Том 1. С. 363 – 371.