

ТРУДЫ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ
Итоги Электронного Геофизического Года
3–6 июня 2009 • Переславль-Залесский, Россия

Основы понимания геологических знаний в геологических обучающих комплексах

А. А. Писарев¹ и В. Н. Дудецкий¹

Получено 29 октября 2009; принято 8 ноября 2009; опубликовано 11 ноября 2009.

КЛЮЧЕВЫЕ СЛОВА: геология, обучающая система, классификация, геологическое знание.

Ссылка: Писарев, А. А. и В. Н. Дудецкий (2009), Основы понимания геологических знаний в геологических обучающих комплексах, *Росс. ж. наук о Земле*, 11, RE1004, doi:10.2205/2009ES000356.

Компьютерная обучающая система по геологии полезных ископаемых представляет собой программу, предназначенную для изучения процессов образования горных пород, свойств минералов и типов месторождений. Данная система включает в себя три комплекса по нескольким геологическим дисциплинам. Первый – минералогический комплекс содержит информацию о минералах и их свойствах, второй описывает типы горных пород и их генезис, третий – типы месторождений полезных ископаемых, условия образования и их основные черты. Каждый комплекс может выступать в качестве отдельно взятой обучающей программы. Система объединяет эти комплексы по средствам классификационной схемы. Схема представляет собой иерархическую структуру, в которую входят генетическая классификация месторождений полезных ископаемых, схема процессов минералообразования и схема систематики минералов. Каждая составляющая построена по принципу “от большего к меньшему”. Классификационная схема представлена единым текстовым файлом. Общее количество знаков, определяющих роль словосочетания в иерархической структуре равно 15. Первые три знака до запятой неизменны и резервируются в зависимости от наличия свободных номеров в базе. Количество знаков после запятой варьируется и может быть равно от 1 до 12, в зависимости от положения в иерархическом ряду.

В систему загружается шаблон классификационной схемы, в котором объявлена каждая из позиций иерархической структуры. Настройка системы производится за счет чтения классификационной схемы. Гибкость в настройке позволяет резервировать несколько порядковых номеров для каждого ранга, а также задавать со-

ответствие порядковых номеров иерархическому значению в пределах одной классификации, когда объединяются несколько схем с одинаковыми рангами в начале и различными рангами в конце. В результате чего формируется база интерфейсов, в которой система в соответствии с шаблоном выстраивает все элементы иерархической структуры. Далее в систему загружаются схемы физических и химических характеристик элементов иерархической структуры и их видеообразы. Происходит формирование баз физических и химических характеристик, а также базы видеообразов. Сформированные базы объединены друг с другом посредством центральной базы – Главного индекса. Главный индекс состоит из записей фиксированной длины, каждая запись принадлежит к одному из следующих типов: слово; слово-продолжение; понятие; понятие-продолжение; шаблон; шаблон-продолжение. Слово является определяющей структурной единицей Главного индекса. Адрес слова в Главном индексе совпадает с указателем на Главный индекс в базе слов. В зависимости от частоты употребления слова в определениях других слов, понятий и шаблонов, формируются структуры слов-продолжений, замыкаемые в кольца. Для определения сложных объектов в главный индекс введены особые структуры – понятия. Понятия, аналогично словам, образуют замкнутые кольца. В отличие от слова, адрес понятия в Главном индексе не представлен в базе слов. Для определения устойчивых выражений, не определяемых в текстах явно, в Главный индекс введены шаблоны, которые также замыкаются в кольца в зависимости от частоты употребления. Между словами, понятиями и шаблонами введены отношения наследования, омонимии, синонимии, антонимии. Адреса слов, понятий и шаблонов в Главном индексе совпадают с указателями на Главный индекс в базах интерфейсов, физических и химических характеристик, и видеообразов (Рис. 1). Таким образом, формируется база знаний системы, в которую помимо вышеперечисленных баз также входят базы слов, источников и шаблонов. Для пополнения и расширения

¹ Российский государственный геологоразведочный университет, Москва, Россия

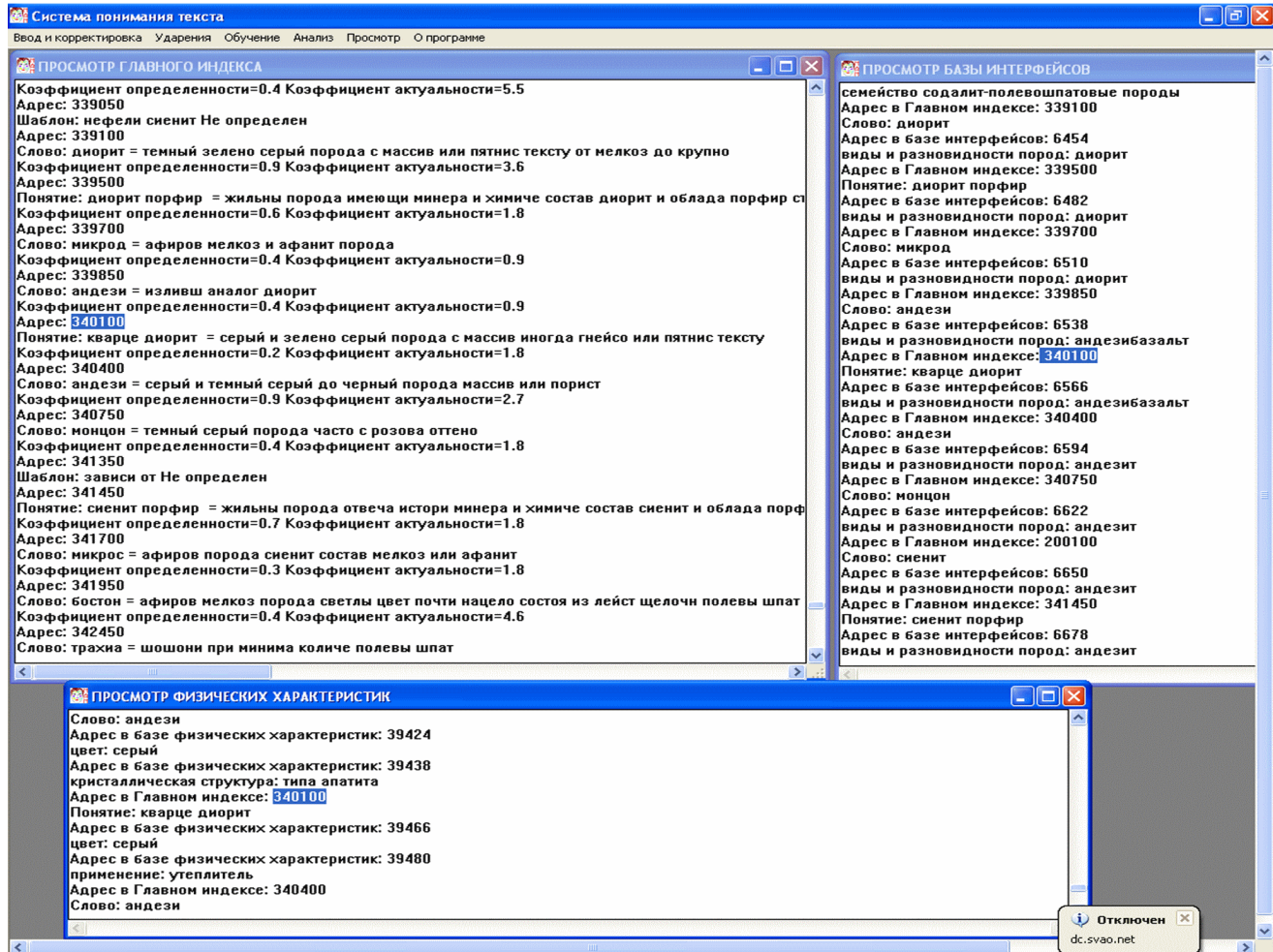


Рис. 1. Главный индекс.

базы знаний, на последующем этапе в систему загружаются геологические тексты. Все предложения вводимого текста проходят четырехуровневый анализ для определения семантического типа предложения: морфологический, синтаксический, логический и семантический анализ. Чтение компьютером геологической текстовой базы производится на этапе прагматического анализа абзаца. Прагматический анализ проводится после того, как текст прошел без ошибок предыдущие этапы. В результате чего формируется база геологических знаний.

Обучающая система по геологии полезных ископаемых позволяет получить знания о генетических типах месторождений полезных ископаемых, об условиях их образования, минеральном составе и локализации полезного компонента, выявить к какому классу, группе и серии относится тип месторождения. В зависимости от этого сделать вывод, к каким типам пород приурочены место-

рождения, в результате какого процесса произошло их образование и что для них характерно.

Известно, что область геологических знаний сложна и противоречива с лексической точки зрения. Зачастую одному термину могут соответствовать несколько различных понятий и наоборот одинаковые понятия встречаются в различных терминах. Главная особенность данной системы состоит в том, что она позволяет идентифицировать понятия такого рода и сопоставлять присущую для них информацию, найденную в тексте. Тем самым была решена проблема противоречивости геологического знания.

А. А. Писарев, В. Н. Дудецкий, Российский государственный геологоразведочный университет, Москва, Россия (mushroompi@mail.ru)