

## ТЕХНОЛОГИЯ СТРУКТУРИРОВАНИЯ ПЕРВИЧНОЙ ИНФОРМАЦИИ КАМЕННОГО МАТЕРИАЛА

**Кувшинова Л.А. (nadegdakl@rambler.ru), Кувшинова К.А. (kaelmo@yandex.ru), Куприянова И.И. (kuprijan@aha.ru), Скоробогатова Н.В., Клименцова Н.И. (niknik@yandex.ru)**  
Московское отделение. ФГУП ВИМС

## THE TECHNOLOGY OF STRUCTURING OF THE PRIMARY INFORMATION OF STONE MATERIAL

**Kuvshinova L.A. (nadegdakl@rambler.ru), Kuvshinova K.A. (kaelmo@yandex.ru), Kupriyanova I.I. (kuprijan@aha.ru), Skorobogatova N.V., Klimentzova N.I. (niknik@yandex.ru).**  
Moscow branch. FSUE VIMS

В настоящее время существует большое количество коллекций, собранных на протяжении многих лет при проведении геологоразведочных работ. Однако до сих пор вся эта разрозненная информация не объединена в единый банк данных, доступный широкому кругу пользователей.

Решить эту задачу при стандартном подходе к созданию банка данных, то есть исходя из первичности формата, по нашему мнению, нереально, учитывая особенности информационной инфраструктуры каменного материала: неоднородность, разноуровневость и рассредоточенность исходной геологической информации, низкий уровень ее формализации, а главное - различные принципы формирования и учета каменного материала, выработанные на местах в различных организациях.

Предлагаемая технология дает возможность практически сразу создавать элементы банка данных (например, строить систему запросов). Эта технология в дальнейшем может служить мостом для автономно созданных баз данных при объединении их в единый банк. Таким образом, решается главная проблема - проблема доступа к оцифрованной информации, обработанной по различным компьютерным технологиям.

Технология структурирования первичной информации каменного материала разработана в рамках предложенной нами концепции территориально распределенной информационной системы кластерного типа (ТРИСкт), алгоритмы которой, как и многие вопросы построения, апробированы и уже реализованы в ряде пилотных проектов (Кувшинова, 2006; Кувшинова, 2007; Кувшинова, 2008).

Также решены и некоторые проблемы обработки информации, которая включает «предмашинную» и «предбазовую» подготовку, формализацию информации, создание тематических и внутренних справочников, запросную систему, ввод первоначальной информации и другие функции. В результате этого создается локальный банк, информацию которого можно транспортировать в ТРИСкт и экспортировать в любую другую базу данных.

Формализация включает в себя структурирование, унификацию и коррекцию.

Практическая реализация технологии структурирования текстового описания каменного материала, включающего административные, географические и геологические характеристики, имена авторов коллекций и прочее, основана на использовании понятийных списков, созданных формально из фрагментов исходного текста с помощью различных механизмов автоматизации: учете наличия или отсутствия разделителей в записях (точка, запятая, и т.д.), наличия ключевых слов или понятий в тексте (например к групповым понятиям относятся такие слова как район, область, река, гора и т.д. и т.п.).

Результатом первого этапа структурирования выбранного объема информации является создание справочника групп и иерархического справочника.

На следующем этапе осуществляется процесс увязки иерархического справочника с исходными записями, от которых остается неструктурированная часть.

Неструктурированная смысловая часть информации переводится в простой справочник, который увязывается с исходными записями. В результате получаем увязку иерархического и простого справочников с исходным текстом, то есть фактически получаем локальный банк данных выбранного объема информации.

Структурирование на каждом этапе является итерационным процессом и выполняется до требуемого разбора исходного текста.

Предлагаемая технология является актуальной, она позволяет при минимальных затратах бюджетных средств объединить уже оцифрованную разрозненную информацию в единый распределенный банк каменного материала, не требуя унификации понятийной базы при первоначальной оцифровке информации для разнородных объектов, описываемых несопоставимыми между собой классификаторами. При этом сохраняется в первоначальном виде первоначальная запись, а невостребованная неструктурированная на данном этапе часть вводится в справочники по мере необходимости в процессе итерации.

В данном случае представленная процедура итерационной формализации является составной частью разрабатываемой территориально распределенной информационной системы каменного материала кластерного типа. Вместе с тем разработанное программное приложение итерационной формализации является самостоятельным и может быть успешно и эффективно использовано специалистами, занимающимися сбором и систематизацией любой описательной информации, получая базу локального характера.

*Кувшинова Л.А., Кувшинова К.А., Куприянова И.И., Володина И.С.* Основные проблемы создания единой информационной системы образцов каменного материала. // Минеральное сырье. No 18. М.: ВИМС, 2006. С. 185-196.

*Кувшинова Л.А., Кувшинова К.А., Иоффе Р.А., Куприянова И.И., Володина И.С.* Проблемы организации территориально распределенной автоматизированной системы сбора, учета, хранения и использования информации о геологических коллекциях. // Минералогия Урала-2007. Миасс – Екатеринбург: УрО РАН, 2007. С. 321-324.

*Кувшинова Л.А., Кувшинова К.А., Куприянова И.И., Скоробогатова Н.В.* Макет реализации концепции территориально распределенной информационной системы каменного материала. // Минералогические музеи. Санкт-Петербург: 2008. С. 72-74.