

**К ВОПРОСУ СОЗДАНИЯ
ИНФОРМАЦИОННО-
АНАЛИТИЧЕСКОЙ СИСТЕМЫ
МЕСТОРОЖДЕНИЙ РЕДКИХ
МЕТАЛЛОВ**

Бездушный А.Н., Костин В.В.

Вычислительный Центр им. А.А.

Дородницына РАН



Актуальность

1. Правительство РФ перед министерством природных ресурсов и экологии РФ поставило задачу сбора информации по всем природным ресурсам страны в единое хранилище.
2. Специалистам геологической области необходимо средство анализа имеющейся информации.



Постановка задачи:

1. Создание хранилища данных
2. Создание системы визуального анализа больших объемов данных
3. Поиск скрытых закономерностей у месторождений редких металлов



Ключевая проблема:

Сведения о месторождениях в России:

- Являются частной собственностью
- Засекречены государством
- Не собраны

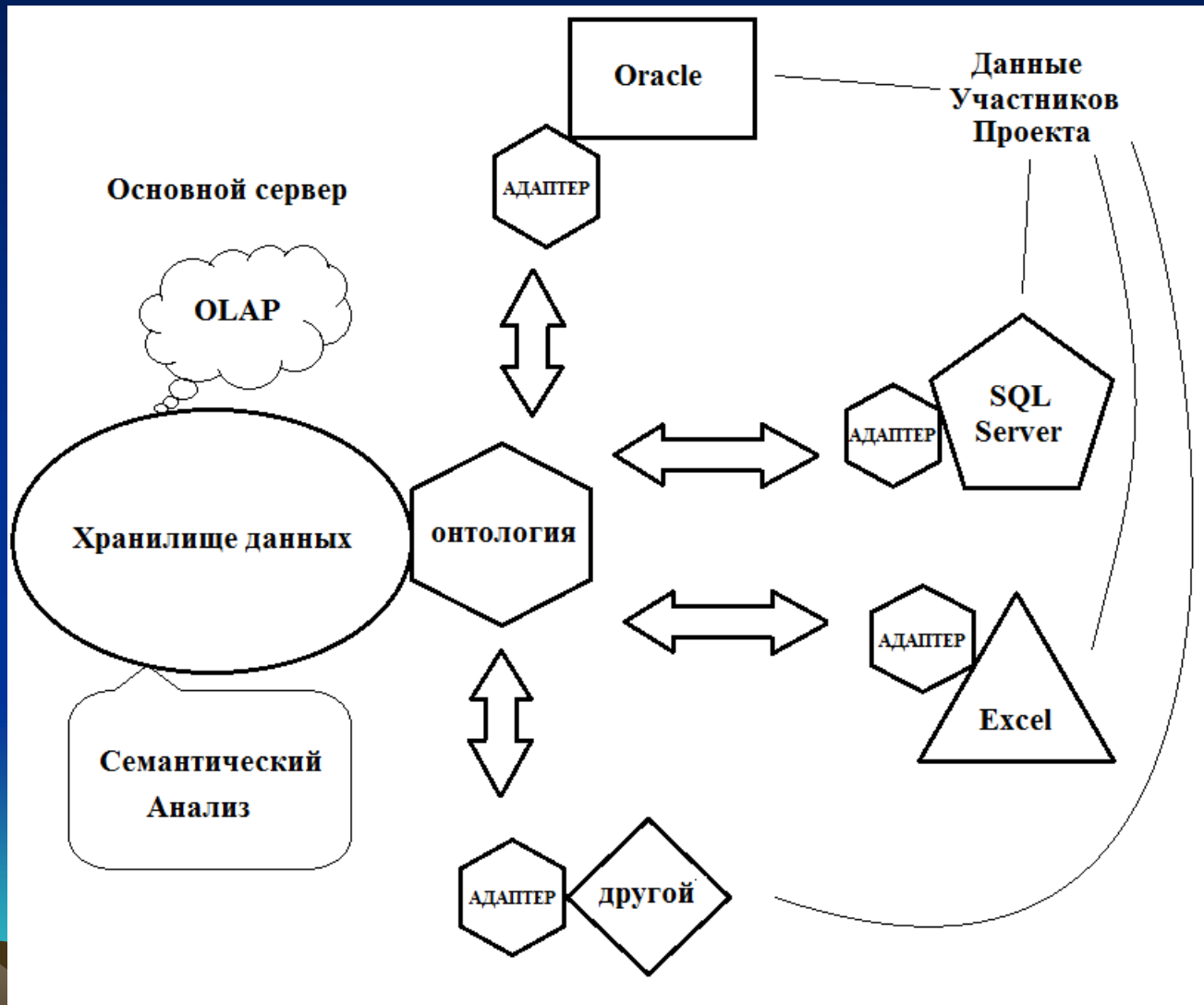


Описание проекта

- Большое количество равноправных участников
- Участники проекта обмениваются информацией разных областей
- Вся информация копируется на центральный сервер
- Результаты анализа доступны всем участникам проекта



Архитектура системы



Основные проблемы при обмене информацией:

- Формирование единой системы понятий о месторождении
- Создание универсальной онтологии
- Создание адаптеров для передачи информации
- Возможность корректировки уже установившейся системы терминов

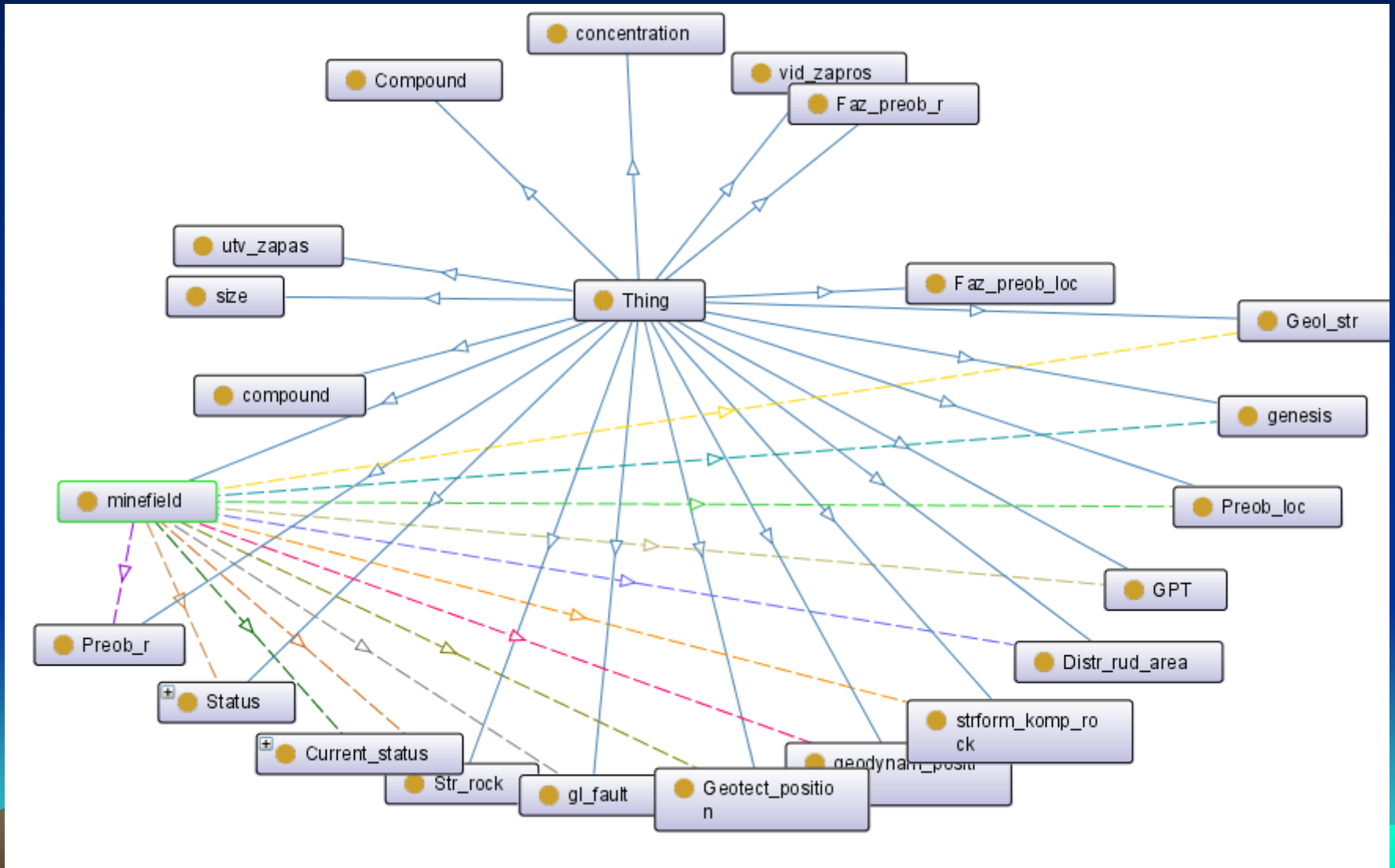


Характеристики, описывающие месторождение:

- Номенклатурные
 - Наименование
 - Статус
 - группы сложности по разным категориям
 - долгота и широта центра месторождения
 - уровень утверждения запасов или ресурсов и др.
- Геологические
 - минералы и руды, добываемые на месторождении
 - методы разработки и исследования
 - геотектоническое и геодинамическое положение месторождения
 - морфологический фактор
 - размах содержания компонентов и др.
- Геоэкономические
 - концентрация главного полезного компонента
 - запасы элементов по разным категориям (А, В, С1, С2)
 - забалансовые запасы
 - извлекаемые запасы и др.



Схема соотношения сущностей сетевой онтологии



Рассмотренные средства визуализации сетевых онтологий

- Ontograf
- OWLViz
- NavigOWL
- RDF Gravity
- OntoVisT



Требования к идеальной системе визуализации:

- Наличие автоматической системы структурирования изображения (layout).
 - Возможность автоматически упорядочивать схему в наглядную картинку.
 - Возможность при упорядочивании выделять основные сущности.
 - Возможность механически редактировать изображение.
 - Оптимизация с точки зрения пересечения проекций ребер графа.
- Возможность одновременно отображать классы, их свойства и экземпляры.
- Возможность хранить созданные ранее срезы и быстро к ним обращаться.
 - Возможность оперативно хранить несколько срезов.
- Наглядное представление подклассов и подсвойств.
 - Указание на связь свойства и его подсвойств.
- Возможность присоединять другие онтологии.
- Возможность отображать одновременно несколько онтологий.
- Возможность управлять онтологией – вносить в нее изменения, сохранять их.
- Наличие удобного фильтра для быстрого поиска.
- Возможность выделить уровни иерархии.
- Возможность экспортировать в другие форматы.



Система визуализации Ontograf:

- Наличие автоматической системы структурирования изображения (layout).
 - Возможность автоматически упорядочивать схему в наглядную картинку.
 - Возможность при упорядочивании выделять основные сущности.
 - Возможность механически редактировать изображение.
 - Оптимизация с точки зрения пересечения проекций ребер графа.
- Возможность одновременно отображать классы, их свойства и экземпляры.
- Возможность хранить созданные ранее срезы и быстро к ним обращаться.
 - Возможность оперативно хранить несколько срезов.
- Наглядное представление подклассов и подсвойств.
 - Указание на связь свойства и его подсвойств.
- Возможность присоединять другие онтологии.
- Возможность отображать одновременно несколько онтологий.
- Возможность управлять онтологией – вносить в нее изменения, сохранять их.
- Наличие удобного фильтра для быстрого поиска.
- Возможность выделить уровни иерархии.
- Возможность экспортировать в другие форматы.



Особенности хранилища данных:

- Невозможность изменения хранящихся данных
- Более оперативный доступ к информации
- Увеличенная скорость обработки информации



OLAP

(аналитическая обработка данных в реальном времени)

- Многомерная структура данных
- Измерения – параметры, по которым проводится анализ
- Значения ячеек – значения, которые подлежат анализу



Способы визуального представления данных

- Сводные таблицы
- Одномерные и двумерные диаграммы
- Проекции трехмерных диаграмм
- Лепестковые диаграммы



Двумерный срез 12-мерного куба

Иерархическая структура географии

Elements	All Elements	Eras	All Eras
genesis	All genesis	GSD	All GSD
GSSUBD	All GSSUBD	Minerals	All Minerals
ODSC	All ODSC	OSC	All OSC
Rocks	All Rocks	Status	All Status

MeasuresLevel	- Level 02	- Level 03	Level 04	Name	Amount	basin	deposit
	All Geography	All Geography Total		All Division	1 390.00	1.00	980.00
	+ AFRICA	AFRICA Total			186.00		104.00
		ASIA Total			292.00		214.00
		+ AFGHANISTAN	AFGHANISTAN Total		5.00		2.00
		ARMENIA					
		+ CHINA	CHINA Total		93.00		69.00
		+ GEORGIA	GEORGIA Total		1.00		1.00
		+ INDIA	INDIA Total		22.00		13.00
			IRAN Total		8.00		7.00
			AZARBAYJAN E SHARQI		1.00		1.00
		- IRAN	KERMAN		2.00		1.00
			KHORASAN		1.00		1.00
			YAZD		3.00		3.00
			ZANJAN		1.00		1.00
		+ IRAQ	IRAQ Total		2.00		1.00
		+ ISRAEL	ISRAEL Total		3.00		3.00
		+ JAPAN	JAPAN Total		4.00		3.00
		+ JORDAN	JORDAN Total		4.00		3.00
		+ KAZAKHSTAN	KAZAKHSTAN Total		53.00		43.00
		+ KOREA	KOREA Total		1.00		1.00
		+ KYRGYZSTAN	KYRGYZSTAN Total		11.00		9.00
		+ LAOS	LAOS Total		2.00		
		+ MALAYSIA	MALAYSIA Total		4.00		1.00
		+ MONGOLIA	MONGOLIA Total		16.00		14.00
		+ MYANMAR (BURMA)	MYANMAR (BURMA) Total		6.00		5.00
		+ NORTH KOREA	NORTH KOREA Total		1.00		1.00
		+ PAKISTAN	PAKISTAN Total		1.00		
		+ PHILIPPINES	PHILIPPINES Total		1.00		1.00
		+ SINGAPORE	SINGAPORE Total		1.00		1.00
		+ SOUTH KOREA	SOUTH KOREA Total		1.00		1.00
		+ THAILAND	THAILAND Total		1.00		1.00
		+ TAIWAN	TAIWAN Total		1.00		1.00
		+ UZBEKISTAN	UZBEKISTAN Total		1.00		1.00
	- ASIA						

3-мерный срез 5-мерного куба

Распределение количества месторождений по соединению, глобальной разломной тектонике и концентрации ПК

Measures: Amount

Name: All Name

Name	Name	All Concentration	Бедное	Богатое	Рядовое
BeO	All GL_fault	14,00	10,00		4,00
	Глубинные разломы, разломы	2,00	2,00		
	Глубинный разлом	4,00	3,00		1,00
	Зона субпараллельных г.	2,00	1,00		1,00
	Зоны пересечения разломы	3,00	2,00		1,00
	Линеамент	2,00	1,00		1,00
	Трансформный разлом	1,00	1,00		
Cs2O	All GL_fault	2,00		2,00	
	Зоны пересечения разломы	1,00		1,00	
	Линеамент	1,00		1,00	
Li2O	All GL_fault	18,00	5,00	6,00	7,00
	Глубинные разломы, разломы	4,00	1,00	1,00	2,00
	Глубинный разлом	5,00	1,00	1,00	3,00
	Зона субпараллельных г.	3,00	1,00	1,00	1,00
	Зоны пересечения разломы	3,00		2,00	1,00

3-мерный срез 20-мерного куба.

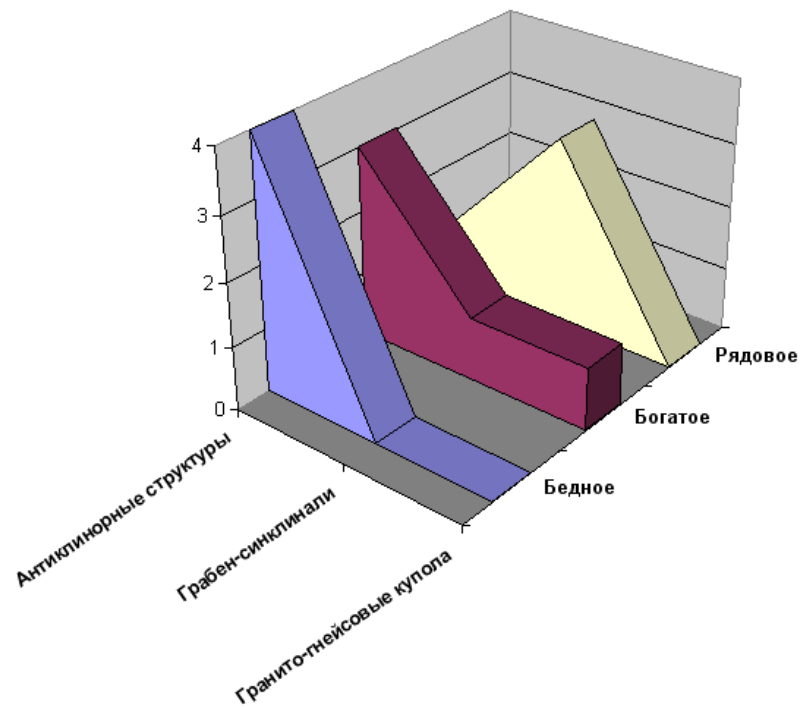
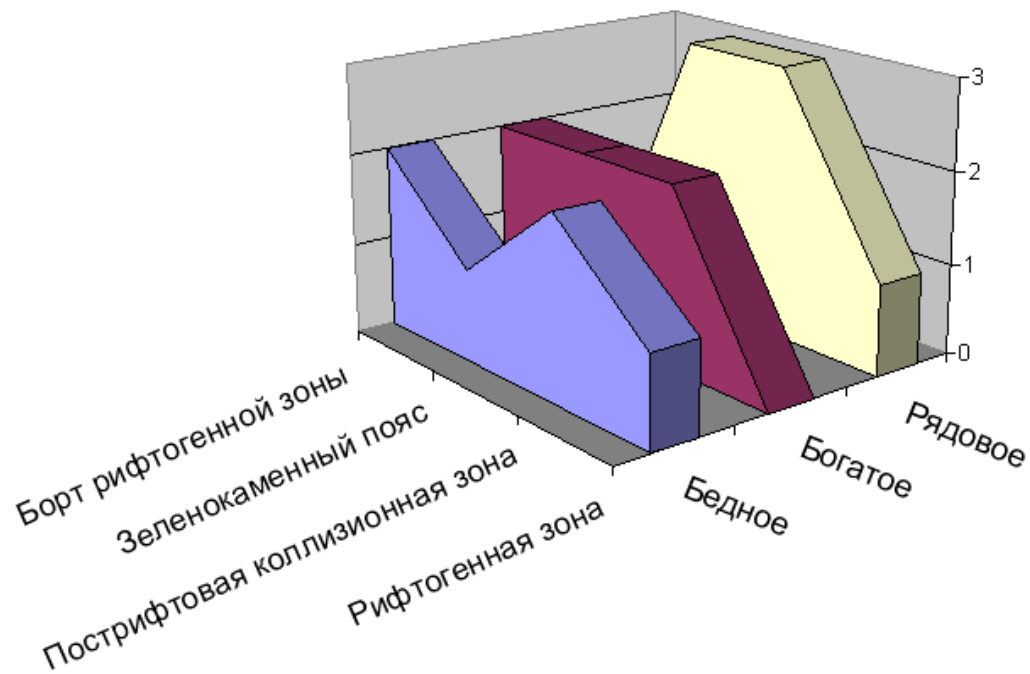
Распределение запасов по элементам, геотектонической позиции и геологической структуре

Concent...	All Concen	Distrib_r...	All Distrib_	Fed_Okr	All Fed_Ok	Geodyn...	All Geodyr	VZR_Str...	All VZR_St
GL_fault	All GL_faul	Preob_r	All Preob_I	Size	All Size	Status	All Status	StrForm...	All StrForm
Subject...	All Subject	Utv_zapas	All Utv_zaj	Measures	Zapas	Str_Rock	All Str_Roc	GPT	All GPT
Preob_loc	All Preob_I	Genesis	All Genesis						

		Name		
Name	Name	All Compound	BeO	Li2O
All Geotect_position	All Geol_str	51 552,00	337,00	51 215,00
	Грабен-синклиналь	6 052,00	337,00	5 715,00
	Ядро антиклинали	45 500,00		45 500,00
Кратон	All Geol_str	6 052,00	337,00	5 715,00
	Грабен-синклиналь	6 052,00	337,00	5 715,00
Обрамление окраины кр:	All Geol_str	45 500,00		45 500,00
	Ядро антиклинали	45 500,00		45 500,00

Schema Data

Трёхмерные графики распределения месторождений по размеру и геологическим параметрам месторождений



Полученные результаты OLAP анализа:

- Богатые месторождения лития:
 1. Не располагаются в рифтогенной зоне
 2. Имеют гранито-гнейсовую геологическую структуру
 3. Расположены на складчатой области
- Бедные месторождения лития:
 1. Расположены в обрамлении срединного массива
 2. Расположены на линзовидно-полосчатой структуре или на изоклиальной складчатости



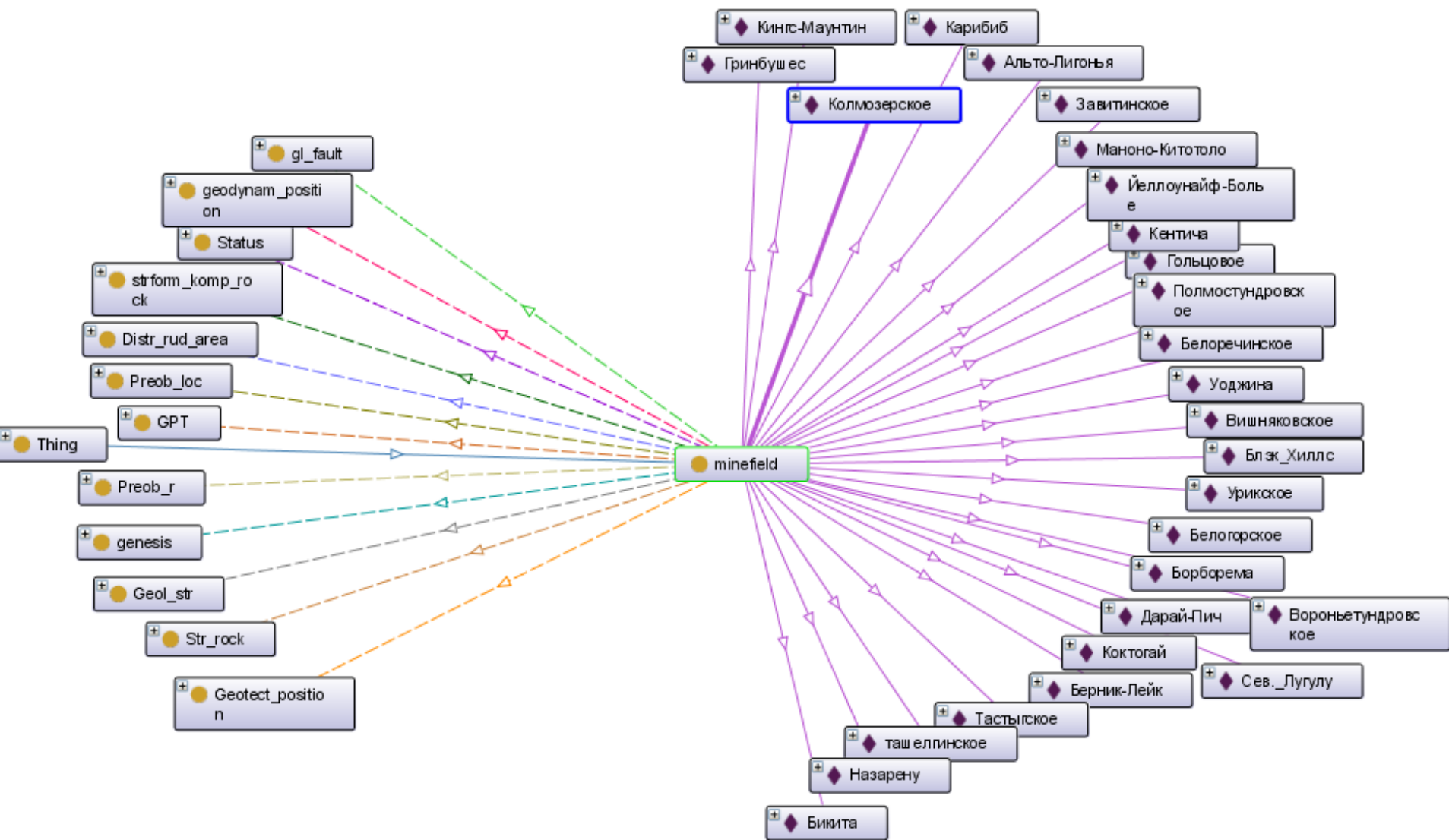
Семантический анализ

Рассмотренные машины вывода:

- Sesame
- OWLIM
- KAON2
- HermiT
- RacerPro
- Pellet



Схема OWL онтологии – связи класса «Месторождение»



Правила вывода

1. Если месторождение содержит минерал сподумен, то месторождение содержит элемент литий.
2. Если глобальный тектонический контроль у месторождения – «сподуменные пегматиты», то месторождение содержит литий.
3. Если на месторождении добывается руда гранитных пегматитов, то на месторождении содержится элемент литий.



Результат семантического анализа

1. Завитинское месторождение
содержит литий.
2. Урикское месторождение содержит
литий.



Методы интеллектуального анализа:

- Нейронные сети
- Рассуждения на основе аналогичных случаев
- Построение дерева решений
- Алгоритм ограниченного перебора
- Визуальный анализ

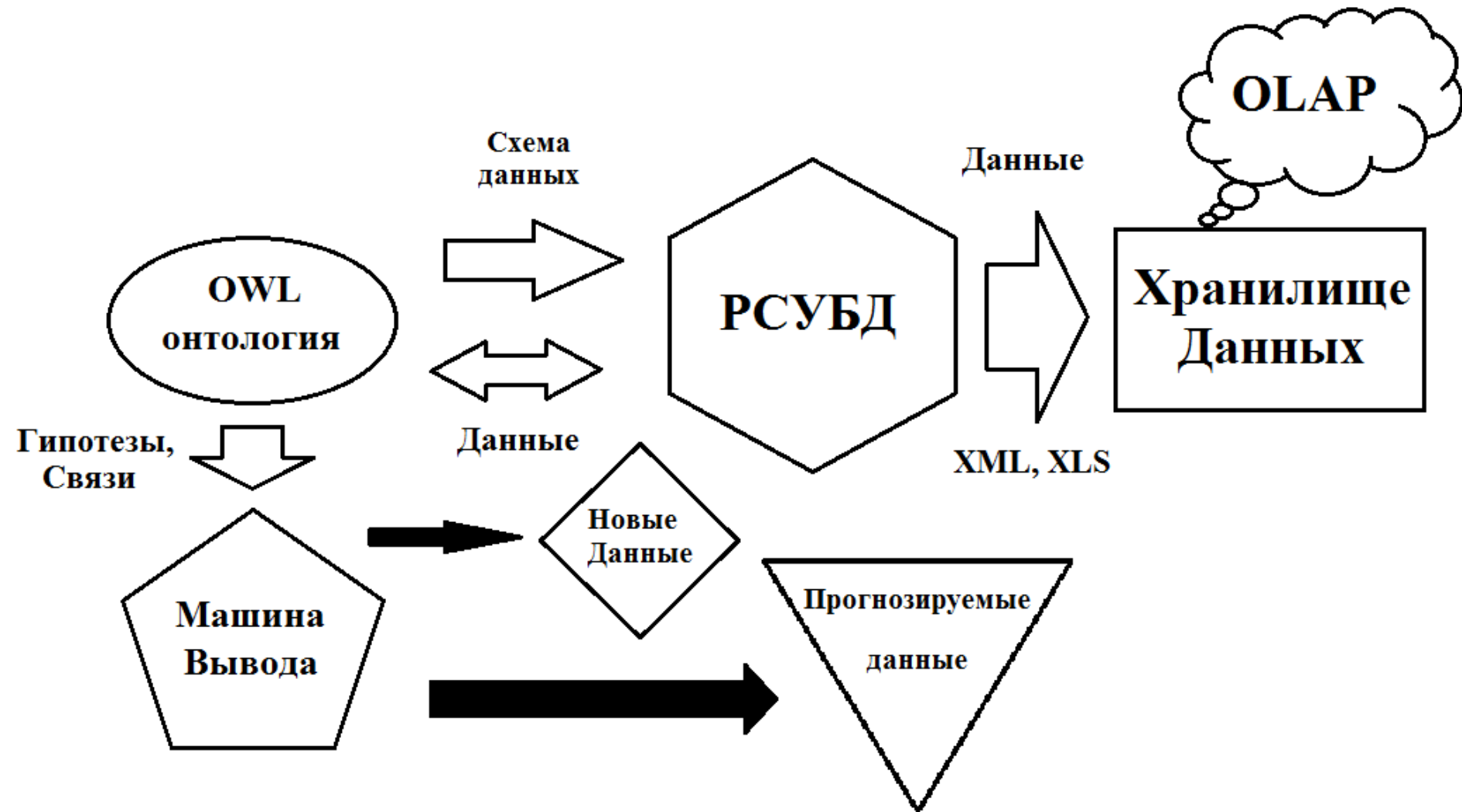


Источники полученной на данный момент информации:

1. Ряховский Владимир Михайлович,
зав. отделом информатизации
Государственного геологического музея им.
В.И. Вернадского РАН.
2. Усова Татьяна Юрьевна,
зав. Сектором экономических исследований
Института минералогии, геохимии и
кристаллохимии редких элементов.



Архитектура реализованной на данный момент системы



Выводы

1. Используя методы OLAP-анализа мы смогли выявить некоторые закономерности в представленных нам данных о месторождениях.
2. Применяя методы data mining на более крупных массивах данных можно выявить большее количество закономерностей и скрытых связей.
3. Используя машину вывода, были получены новые знания.



Дальнейшее развитие

- Методика, отработанная в данной работе, может быть эффективно использована:
 - для оценки экономической эффективности добычи соединений на месторождениях или в отвалах
 - для проведения интеллектуального анализа данных в смежных областях
 - при оценке влияния токсичных отходов на загрязнение окружающей среды
 - при прогнозировании залежей
 - при оценке геоэкономических характеристик других полезных ископаемых.



Спасибо за внимание!



Схема соотношения классов базы данных Ряховского Владимира Михайловича.

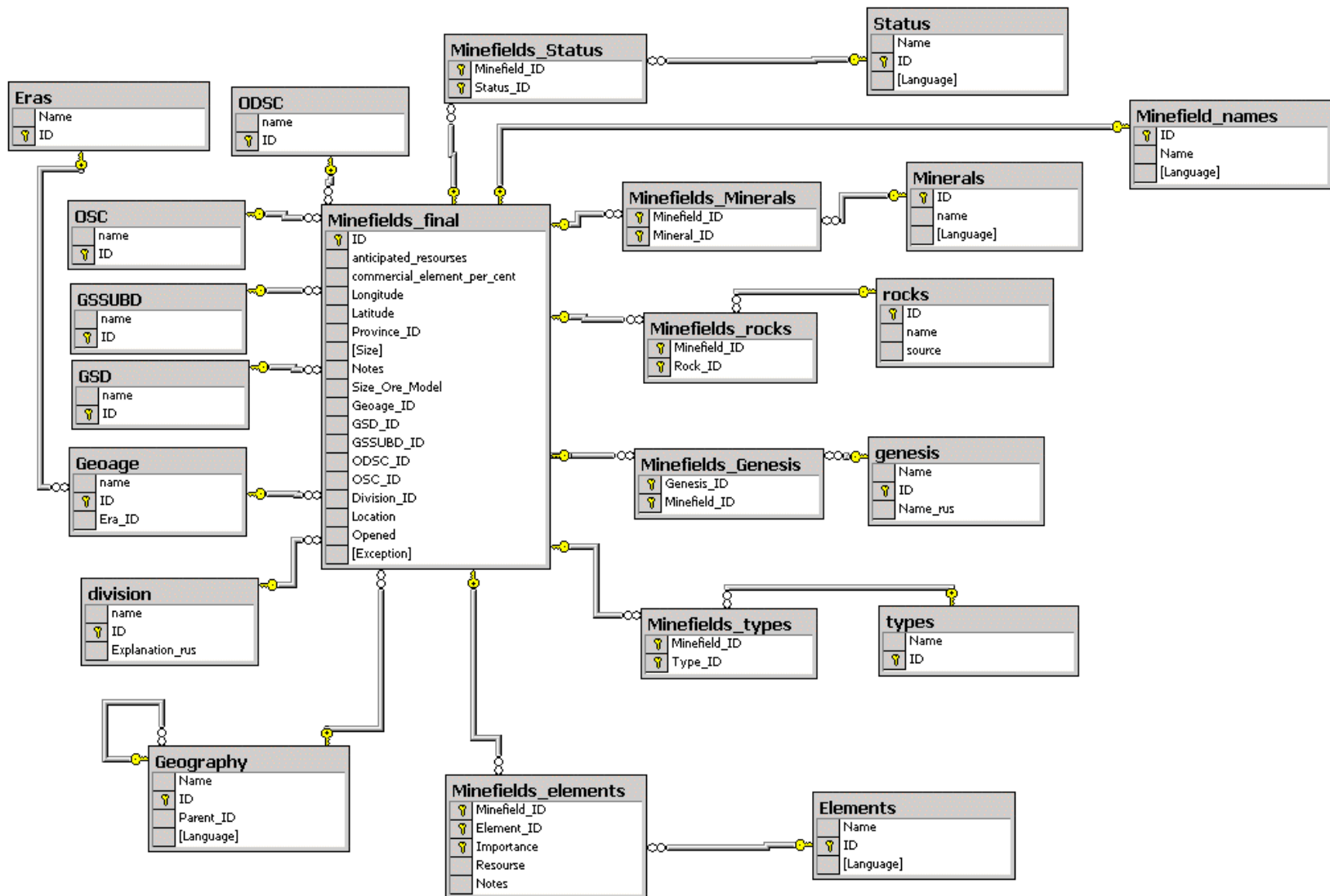
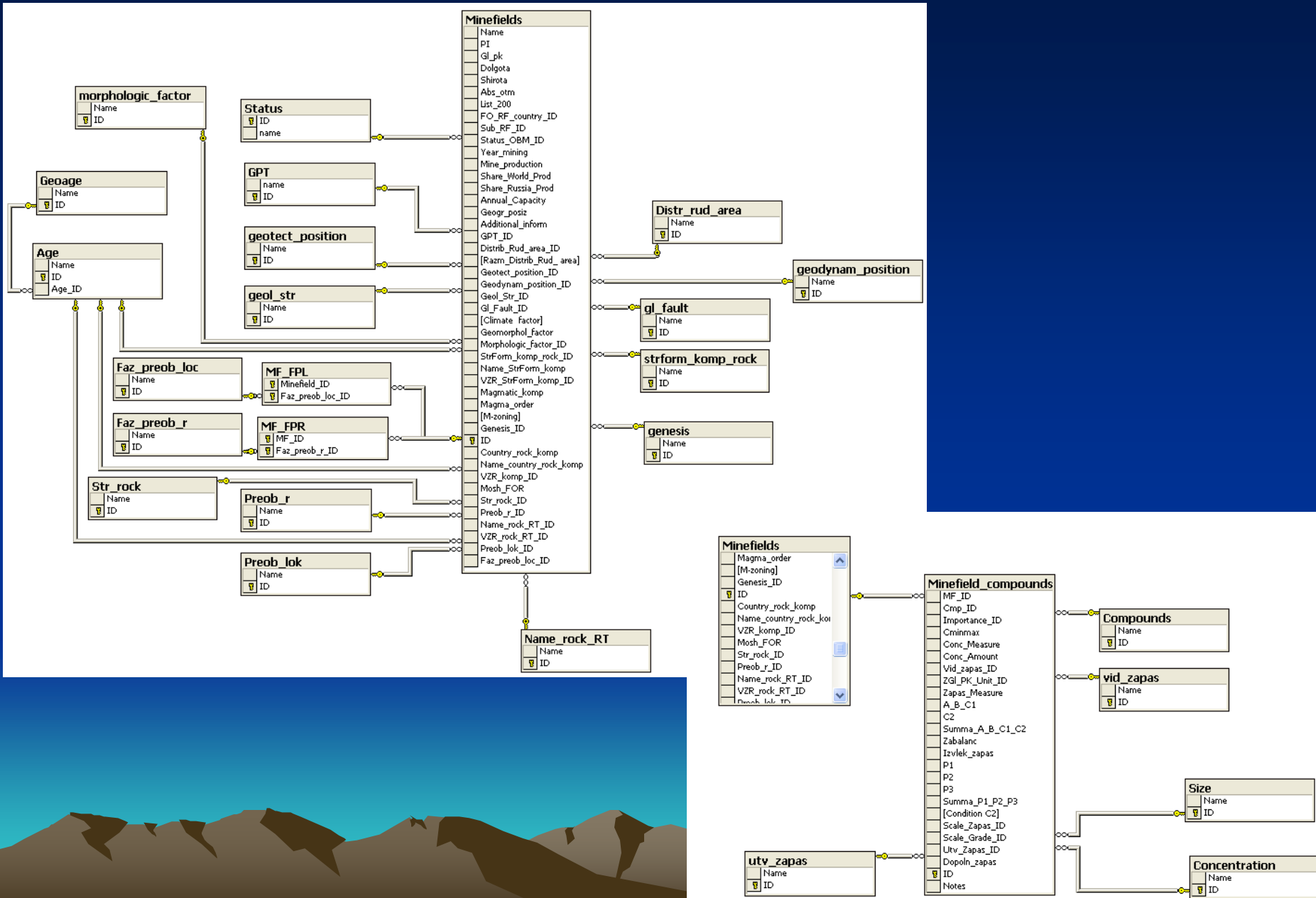


Схема соотношения классов базы данных Усовой Татьяны Юрьевны.



Характеристики Ряховского Владимира Михайловича, описывающие месторождение:

- Геологическая эпоха
- Географическое положение
- Статус месторождения
- Минералы, добываемые на месторождении
- Породы, добываемые на месторождении
- Генезис месторождения
- Геологический тип месторождения
- Химические соединения, добываемые на месторождении
- Географическое положение
- региональных геологических структур
- локальных геологических структур
- рудоконтролирующей структуры
- областной рудоконтролирующей структуры

Связи экземпляров классов «месторождение» и «генезис»

