

Подсистема работы с текстовыми данными на географической карте «Волхв-ГЕО»

Актуальность темы: В последние несколько лет объёмы новостных потоков значительно увеличились. остро встаёт вопрос анализа данных и представления их в удобном для человека виде.

Новостные данные отличаются от иных текстовых данных тем, что они обычно привязаны к определённой местности. Географический анализ новостей становится всё более актуальным.

Тем не менее, объёмы новостны потков не всегда позволяют человеку оперативно оценивать и прогнозировать изменение обстановки. Появляется запрос на прогнозирующую систему, способную предсказывать активность по тем или иным темам.

Цель: Разработать подсистему которая позволит проводить анализ и отображение на географической карте новостной информации в текстовом виде. Подсистема предоставляет спектр методик автоматического анализа новостной динамики и настраиваемую методику интеграции оценок. Подсистема самостоятельно реализует геопривязку новостей с помощью полнотекстовых запросов.

Сравнение аналогов и прототипов с учётом весовых коэффициентов

Критерий	α	liveuamap	militarymaps	Волхв-Гео
Пользовательский интерфейс	0.1	0.75	0.5	0.75
Геопривязка	0.2	0.75	0.75	0.25
Автоматизация	0.25	0.5	0.5	0.75
Интеграция	0.15	0	0.5	0.25
Прогноз	0.2	0.5	0	0.25
Сценарии	0.1	0.5	0.25	0.75
Итого	1	0.3625	0.9375	1.25

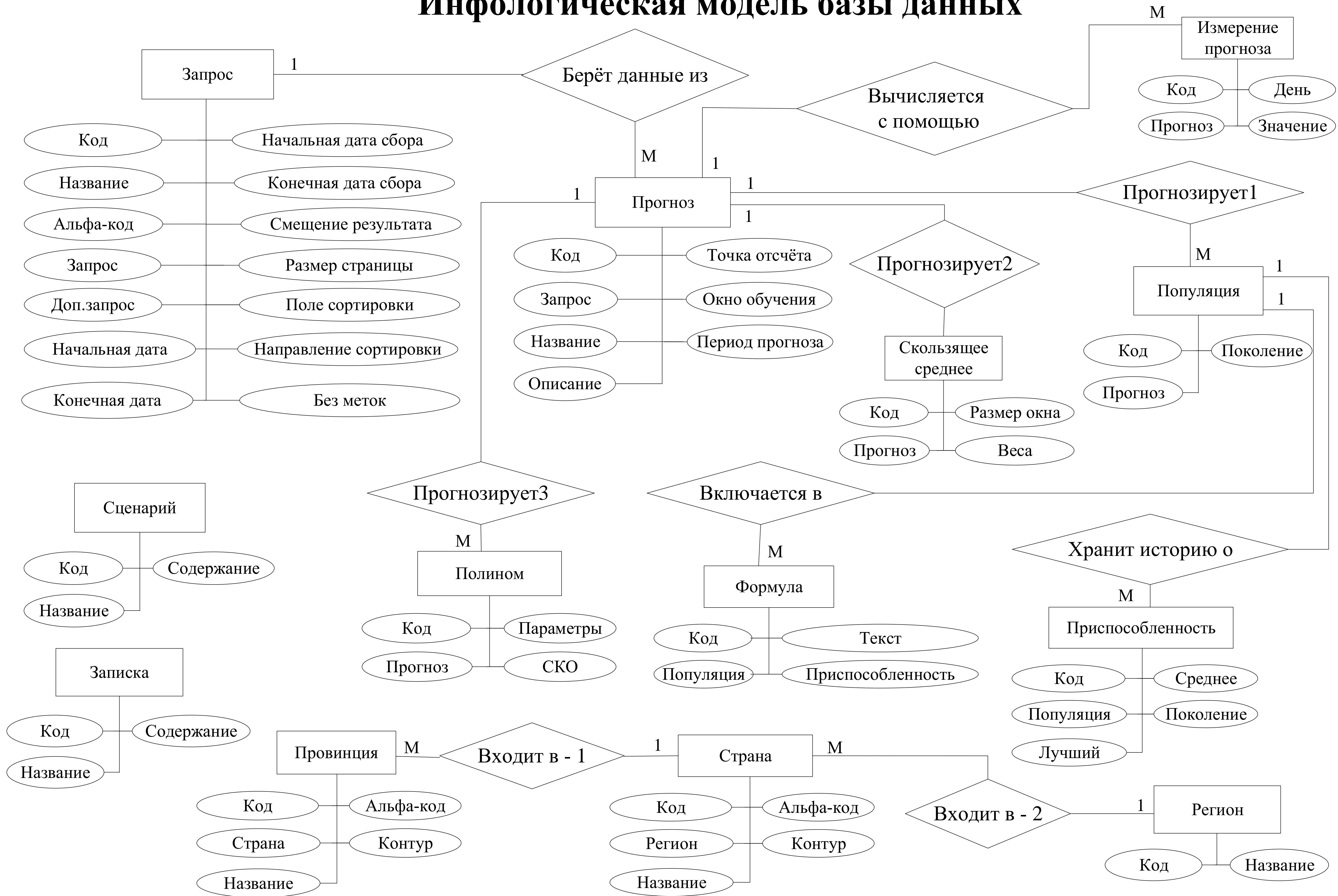
Перечень задач, решенных в процессе проектирования:

- Исследование и анализ предметной области;
- Анализ и определение критериев качества;
- Определение функциональных требований разрабатываемой системы;
- Разработка структуры модулей системы, с выделением функциональности для каждого модуля;
- Проектирование базы данных: инфологическая, даталогическая модели;
- Проектирование эволюционного алгоритма прогнозирования;
- Рассмотрение и обоснование архитектуры системы;
- Выбор программных библиотек для реализации модулей;
- Разработка интерфейса взаимодействия пользователя с программой;
- Реализация графа диалога;
- Написание и отладка программного кода модулей системы;
- Разработка технической документации.

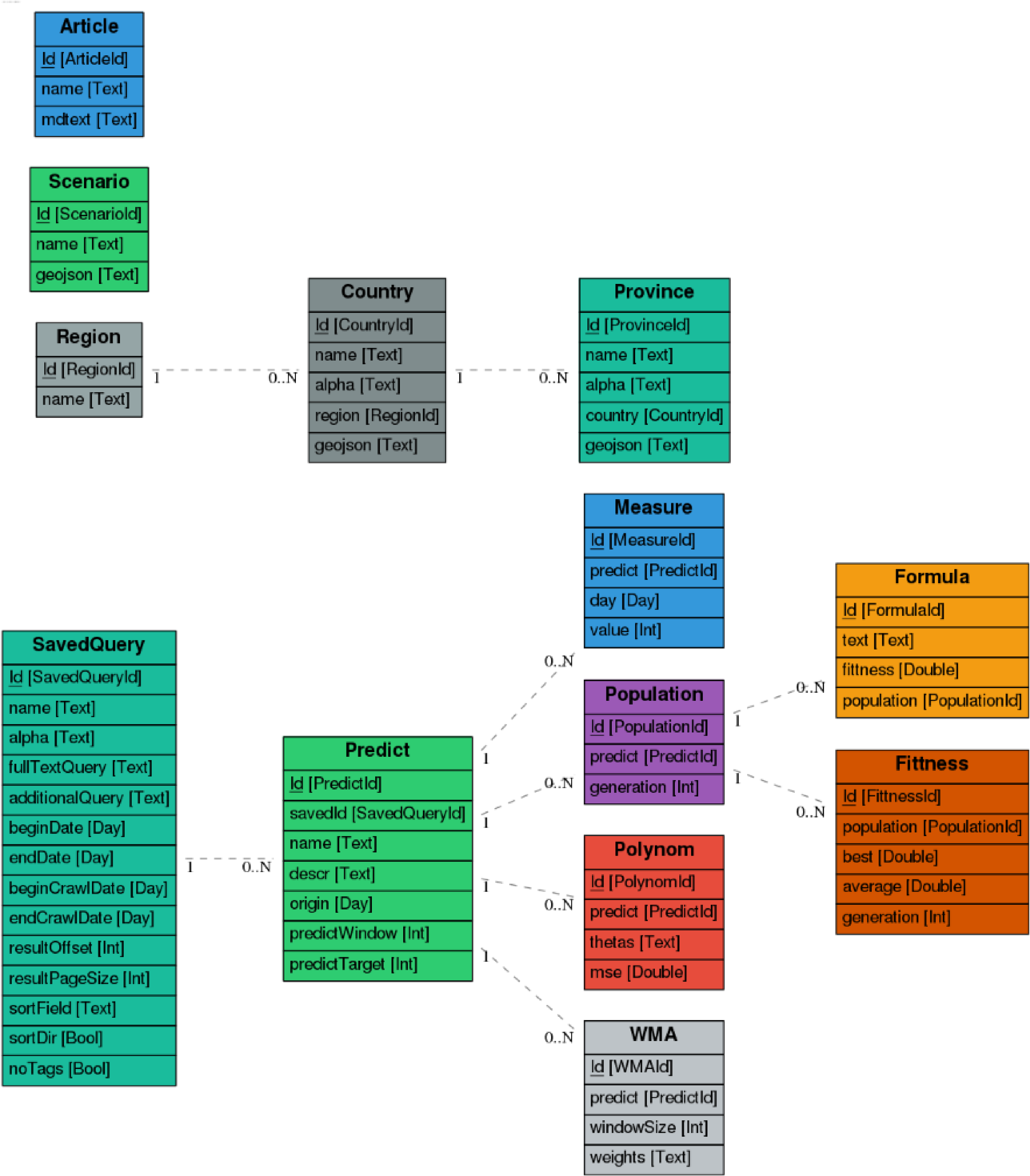
Диаграмма предметной области



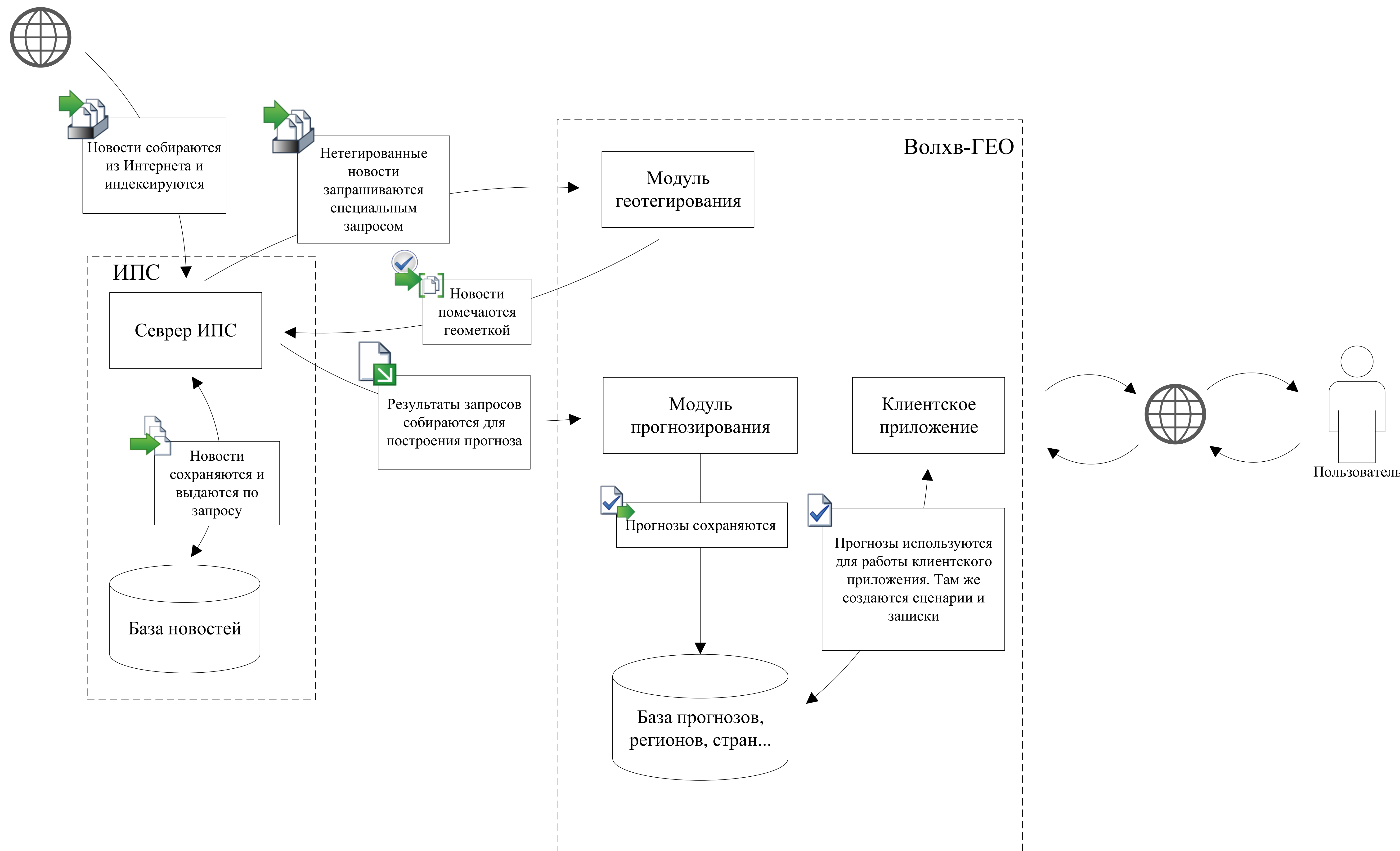
Инфологическая модель базы данных



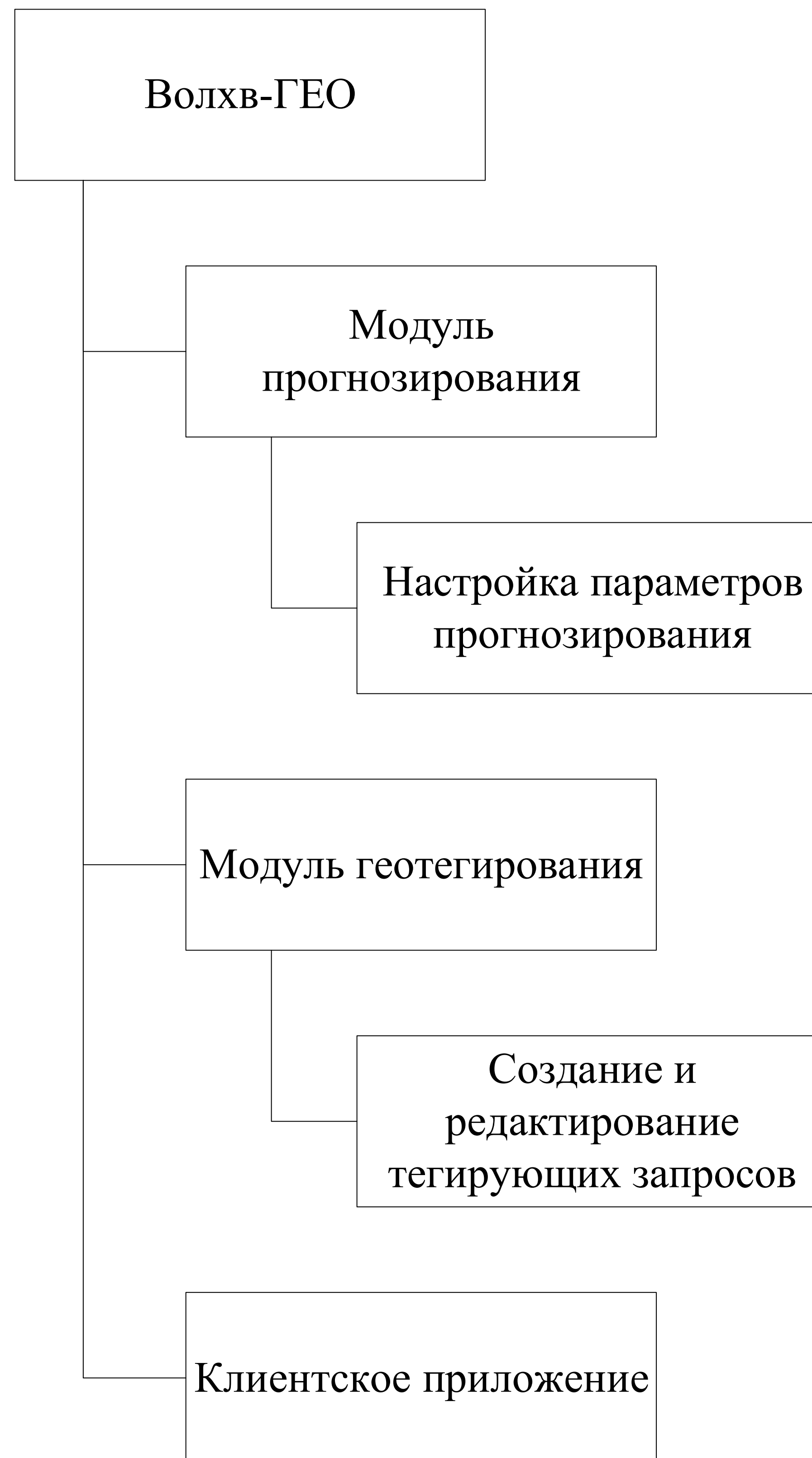
Даталогическая модель базы данных



Бизнес-процесс подсистемы

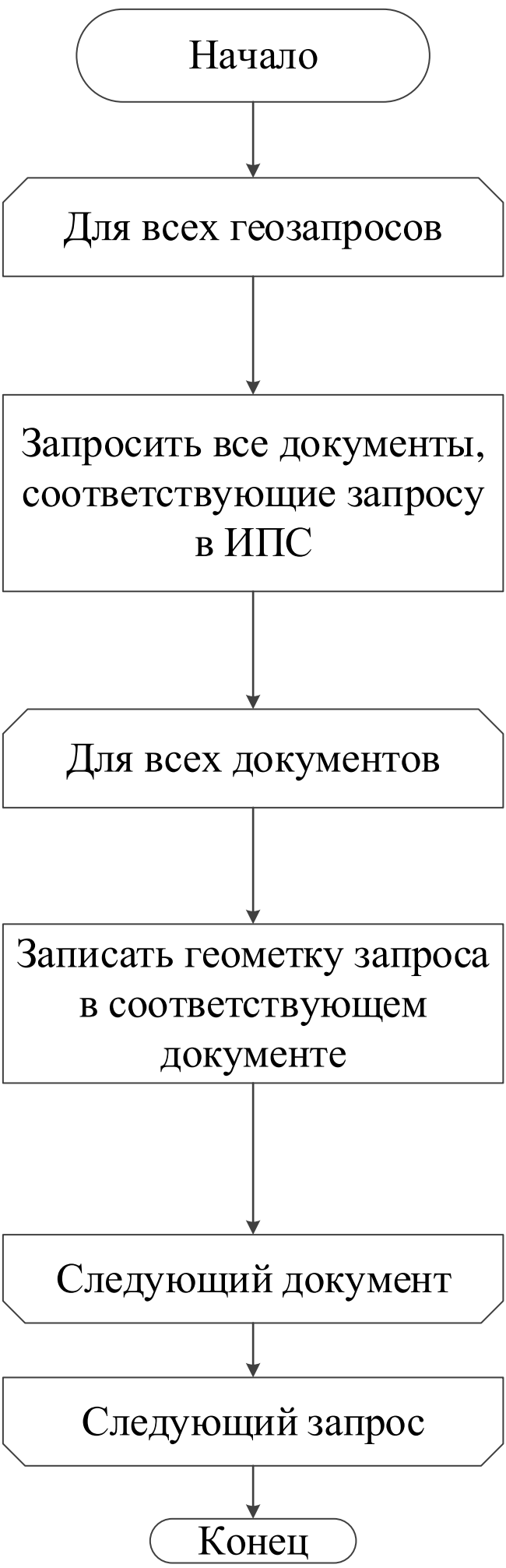


Структура подсистемы



Основные алгоритмы подсистемы

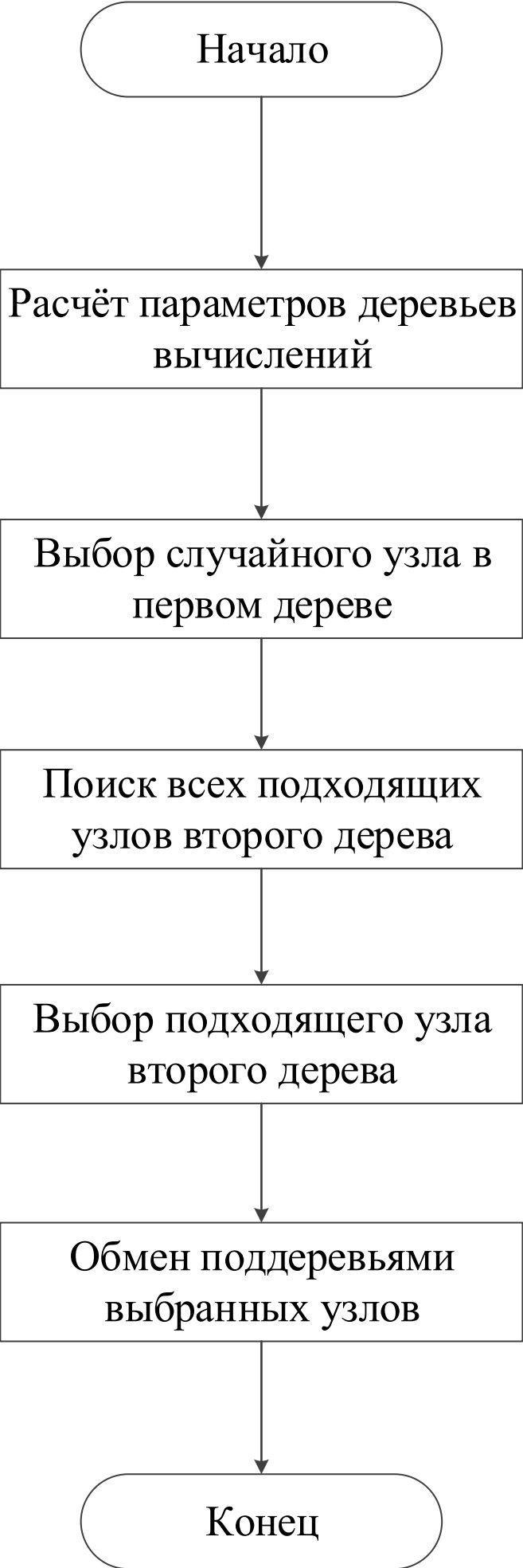
Алгоритм геотегирования



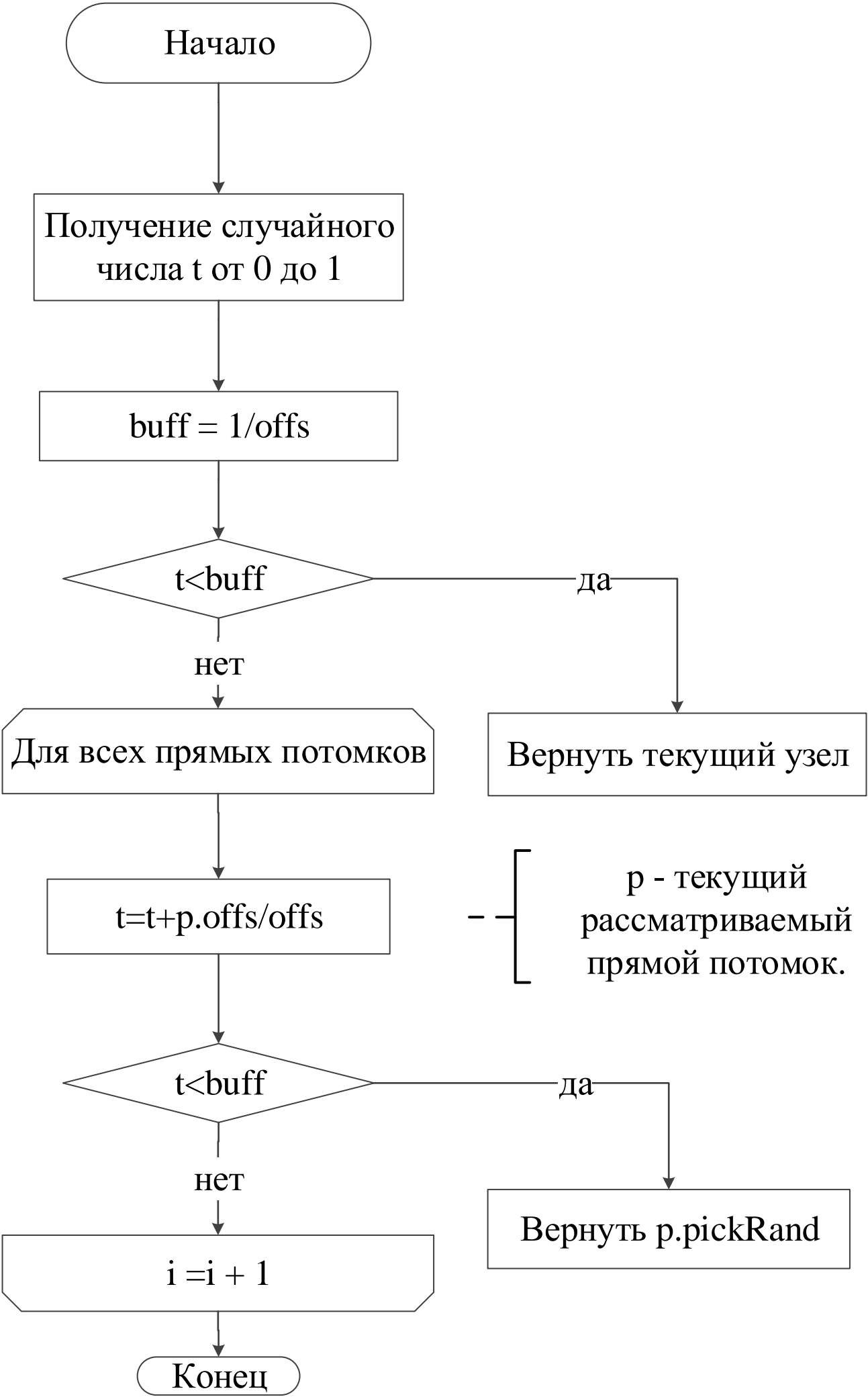
Алгоритм репродукции



Алгоритм кроссинговера



Алгоритм выбора случайного узла



Исследовательская часть

Постановка задачи

Обеспечить прогнозирование развития ситуации с учётом географического распределения новостей

Входные данные для прогноза

- Количество новостей по темам в рассматриваемой провинции в текущем и прошлых периодах.
- Количество новостей по темам в соседних провинциях в текущем и прошлых периодах.

Выходные данные прогноза

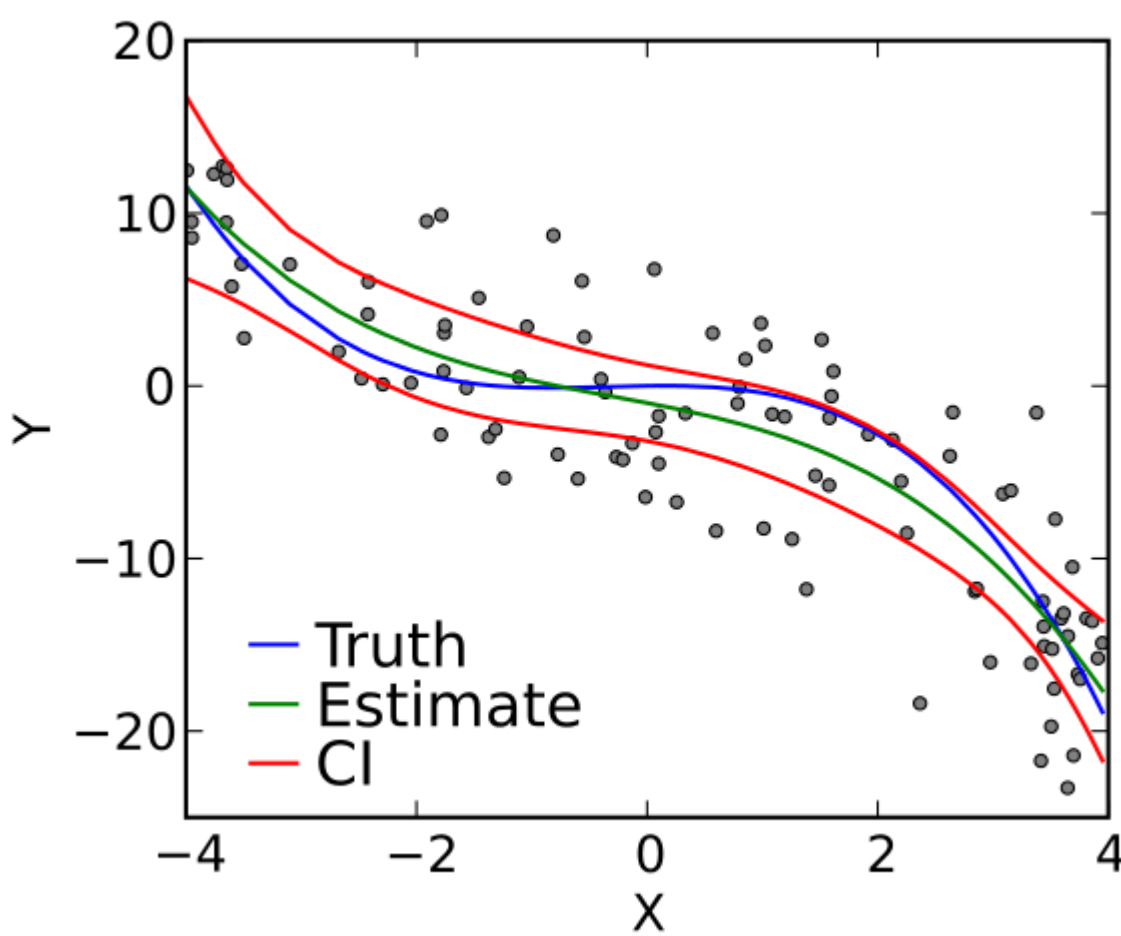
- Прогнозируемое количество новостей по темам через указанный период времени

Критерий	Эволюци-онный алгоритм	Полиноми-нальная регрессия	Скользящее среднее	Комбини-рованный метод
Быстродействие	3	4	4	1
Точность	4	3	3	5
Стабильность	3	4	5	4
Настраиваемость	4	1	1	5
Итого	14	12	13	15

Сравнение метод прогнозирования

Варианты

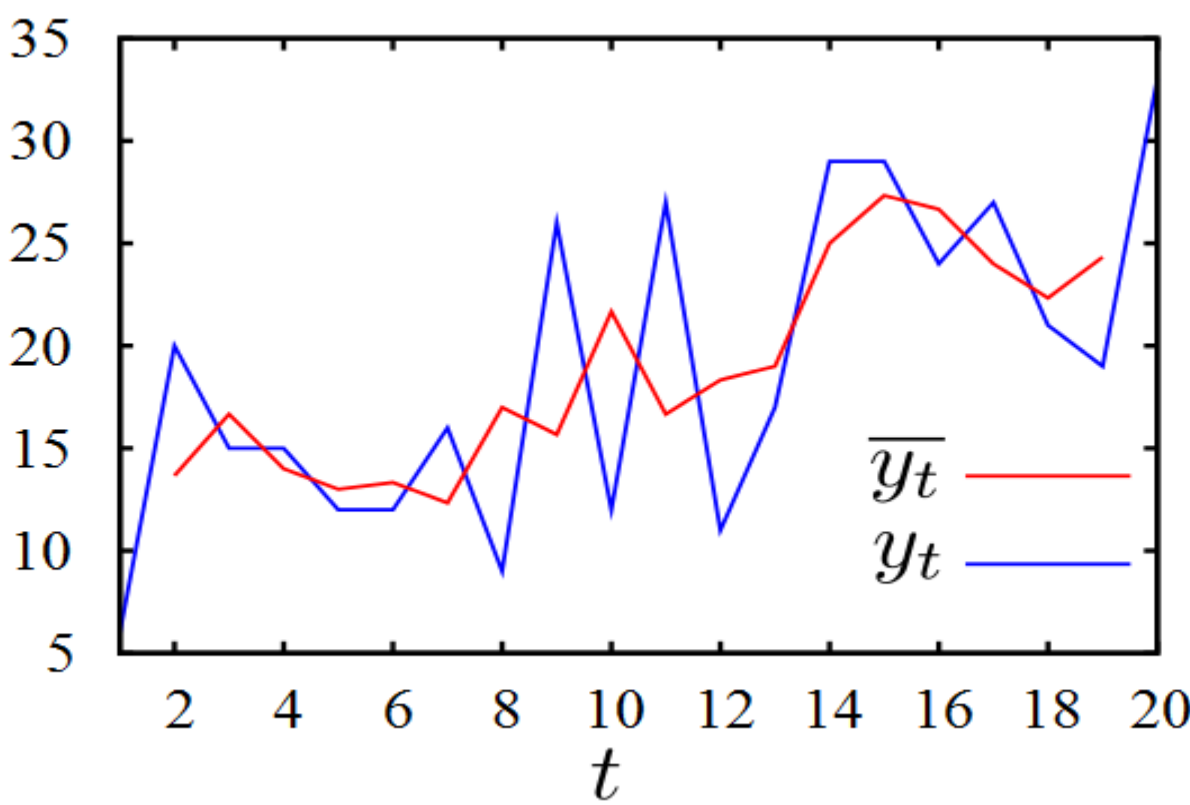
Полиномиальная регрессия – целевая функция рассматривается как полином n-ной степени от входных переменных



Пример полиномиальной регрессии

Эволюционный алгоритм – представлен на отдельном листе

Скользящая средняя – прогнозируемым значением функции является её среднее значение за предыдущий период – окно.

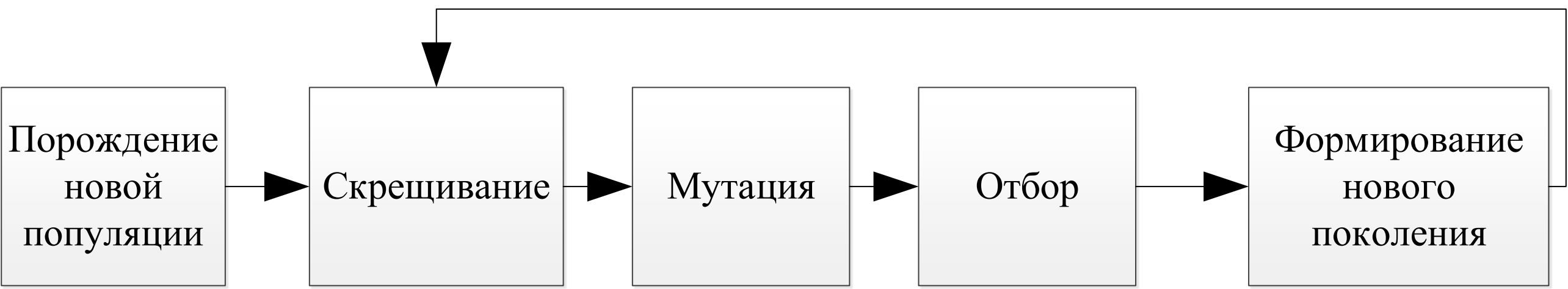


Пример анализа скользящим средним

Комбинированный метод – используем прогнозируемые значения вышеописанных методик. Итоговый прогноз рассчитывается как средневзвешенная сумма методик.

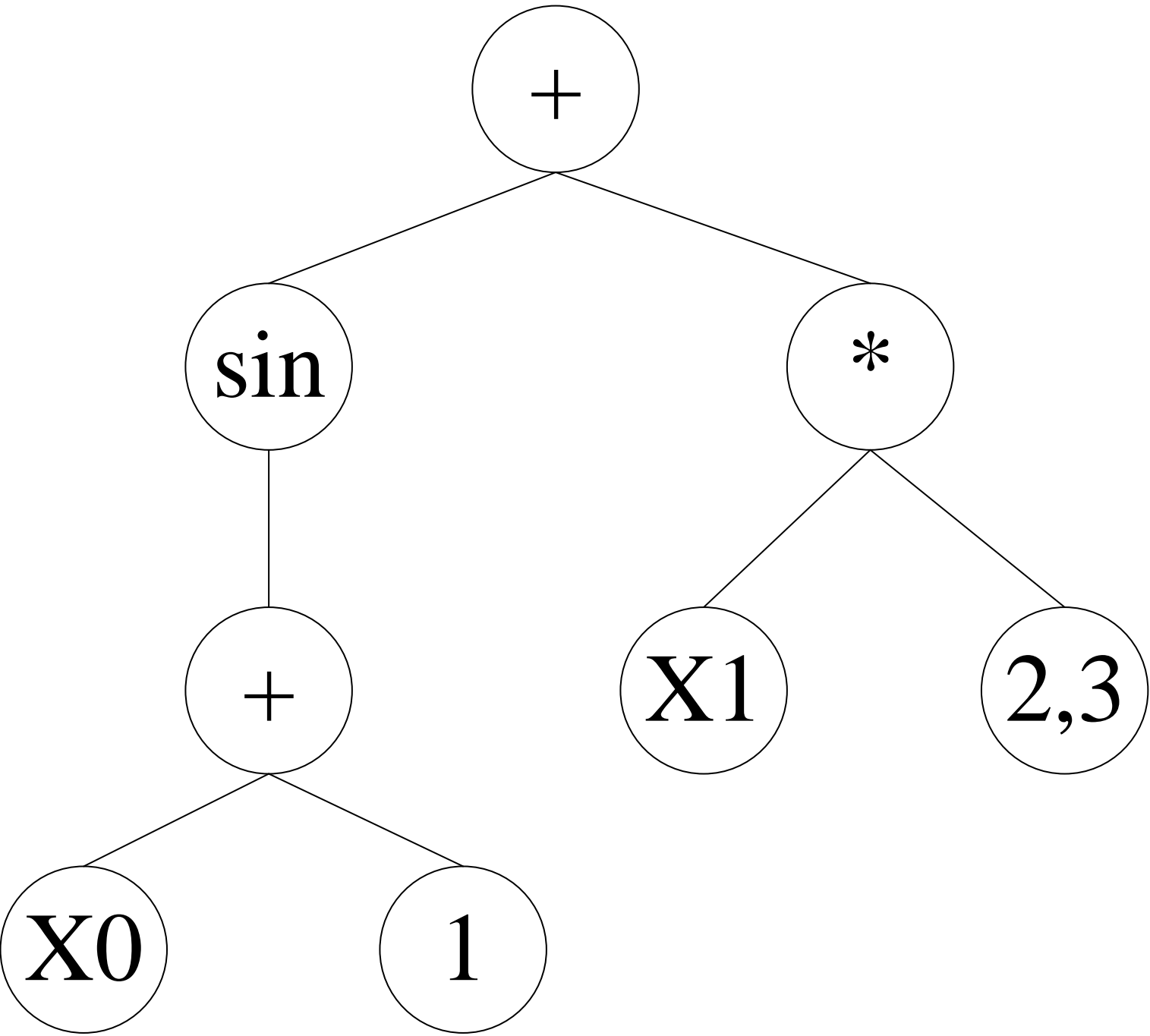
Исследовательская часть 2

Общее представление эволюционного алгоритма



Эволюционирующий объект – дерево вычислений

- Дерево строится по следующим правилам:
- В узлах дерева находятся операции.
 - Потомками узла являются аргументы для той или иной операции.
 - Листами дерева являются константы и переменные.
 - Корень дерева считается находящимся на глубине 1.
 - Глубина дерева не может превышать задаваемой пользователем глубины.



Пример: дерево, описывающее формулу $F(X0,X1) = \sin(X0+1) + X1*2,3$

Порождение случайного дерева

Новое дерево создаётся вызовом функции `getRandExpr(int depth)`, которая принимает максимальную глубину дерева и возвращает ссылку на корень дерева. Функция создаёт корневой узел и исходя из его типа, рекурсивно создаёт нужное количество узлов-потомков.

Стратегия скрещивания

Одноточечное скрещивание в виде обмена поддеревьями. Так же необходимо следить чтобы при скрещивании деревья-потомки не превосходили максимально допустимую глубину деревьев, так как это приведёт к утечке памяти и зависанию программы

Стратегия мутаций

`mutationRate` определяет какой процент популяции подвергнется мутации. Мутация не затрагивает «элитных» индивидов. Случайным образом выбирается узел дерева. Вероятность выбора для всех узлов одинаковая. Выбранный узел заменяется на новое, случайно сгенерированное дерево.

Стратегия отбора

Функцией приспособленности обратно пропорциональна квадрату среднеквадратичной ошибки индивида на обучающей выборке. Вероятность оставить потомство пропорциональна приспособленности индивида

Стратегия отбора

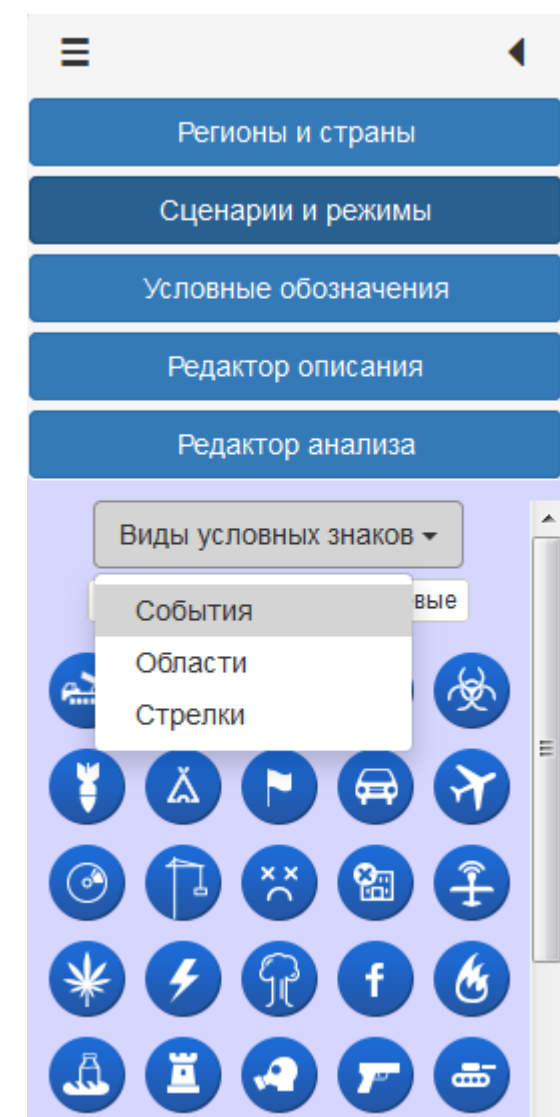
При условии что выбранный узел первого родителя имеет глубину d и высоту h , «хорошие» узлы второго родителя должны удовлетворять условию: $(depth \leq MAX_DEPTH - h) \wedge (height \leq MAX_DEPTH - d)$, где $depth$ и $height$ – глубина и высота узла второго родителя соответственно и MAX_DEPTH - максимальная допустимая глубина дерева.

Сравнение стратегий

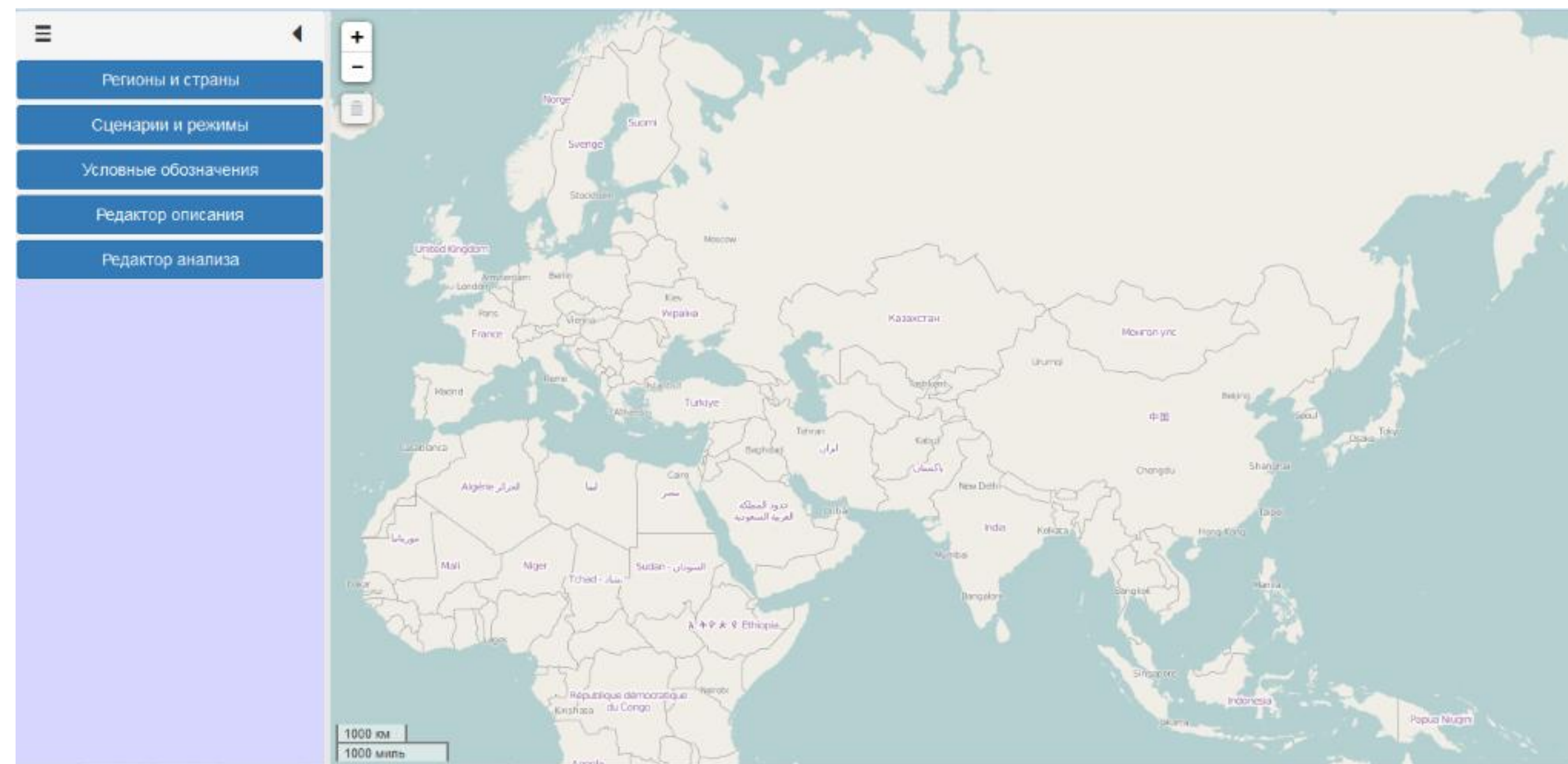
Стратегия \ Критерий	Пропорциональный	«Состязание»
Вычислительная сложность	2	1
Память	1.5	1.5
Скорость схождения	1.5	3
Σ	6.5	7

Выбор стратегии скрещивания

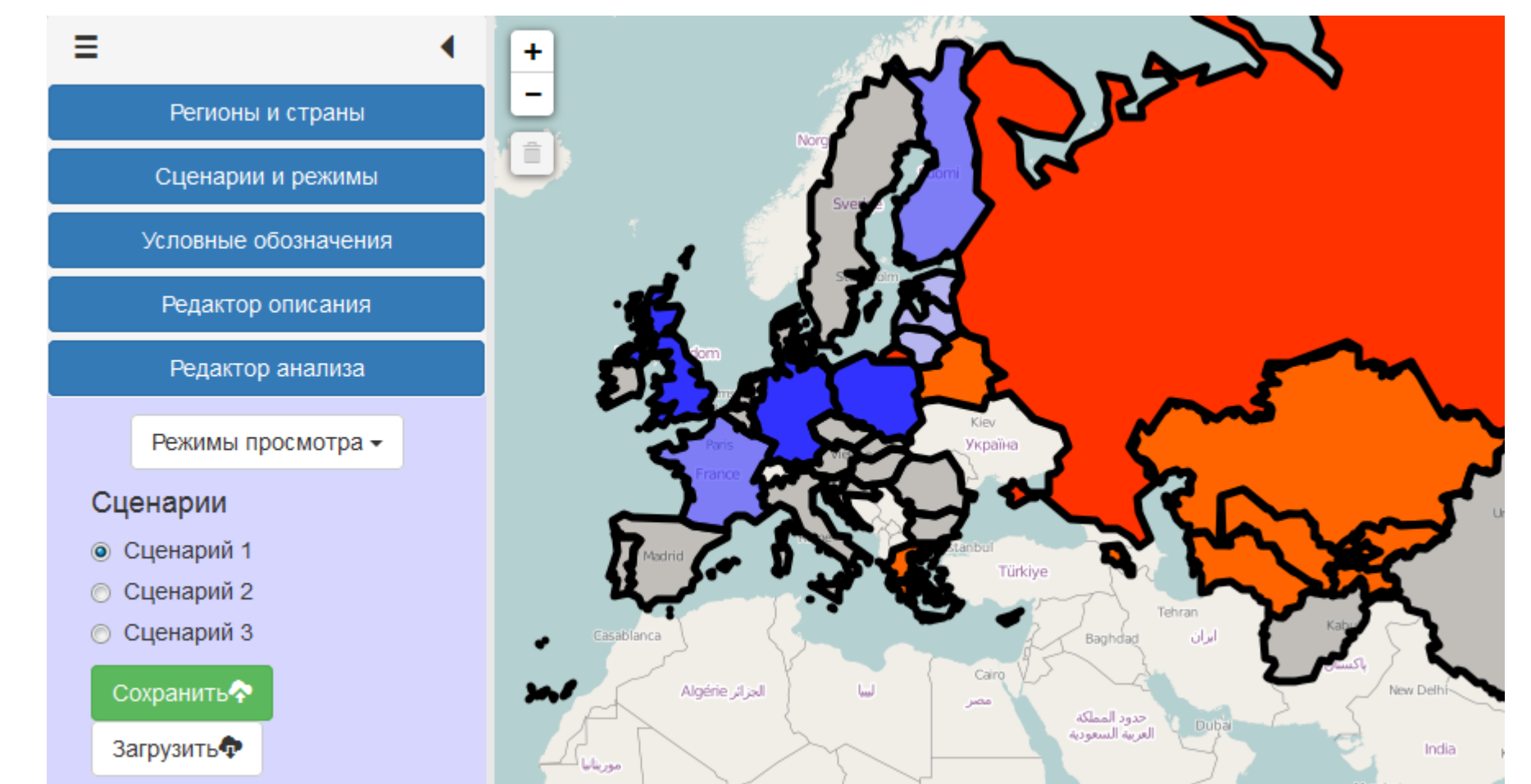
Экранные формы



Виды
условных
знаков



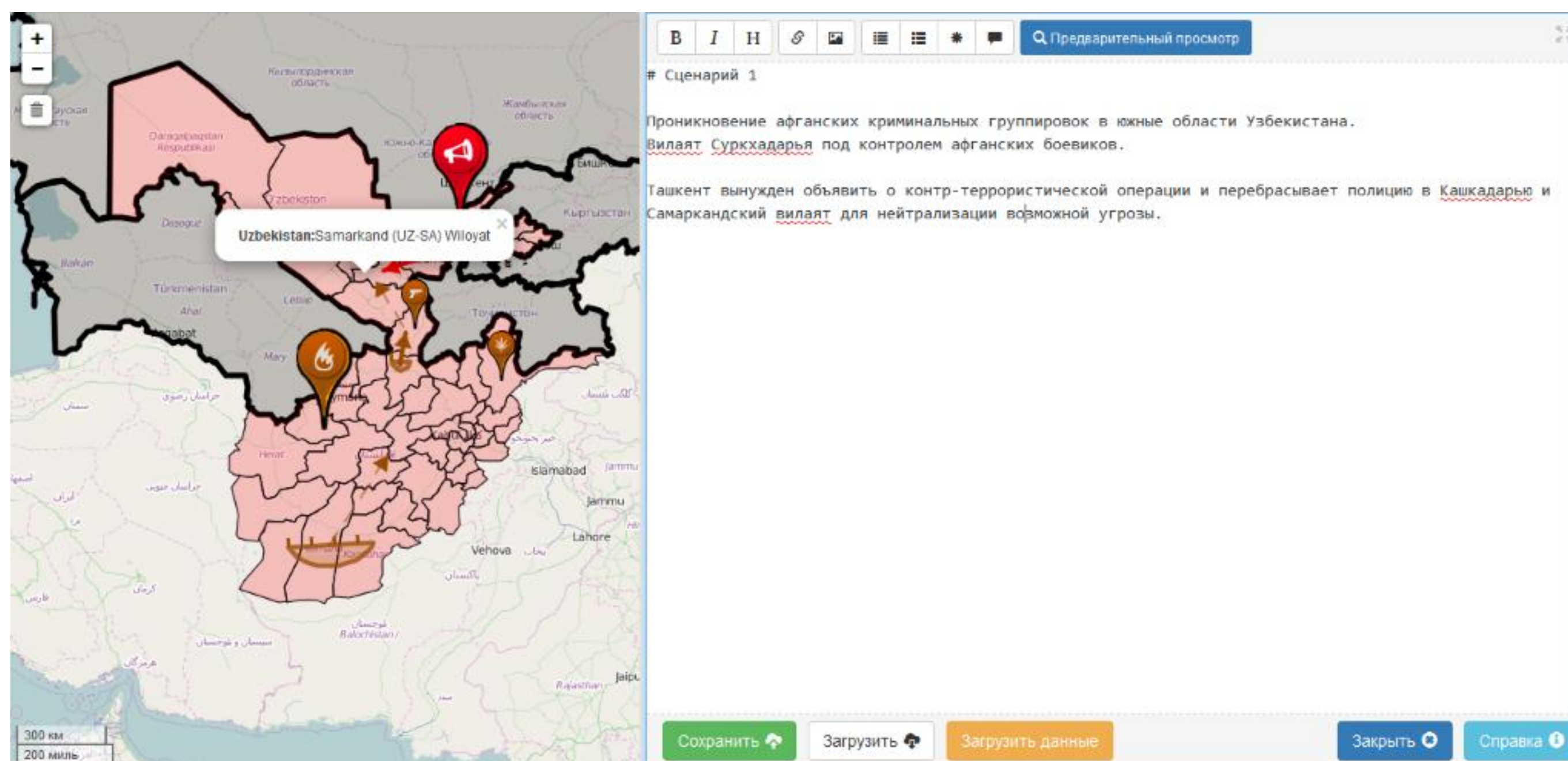
Стартовый экран



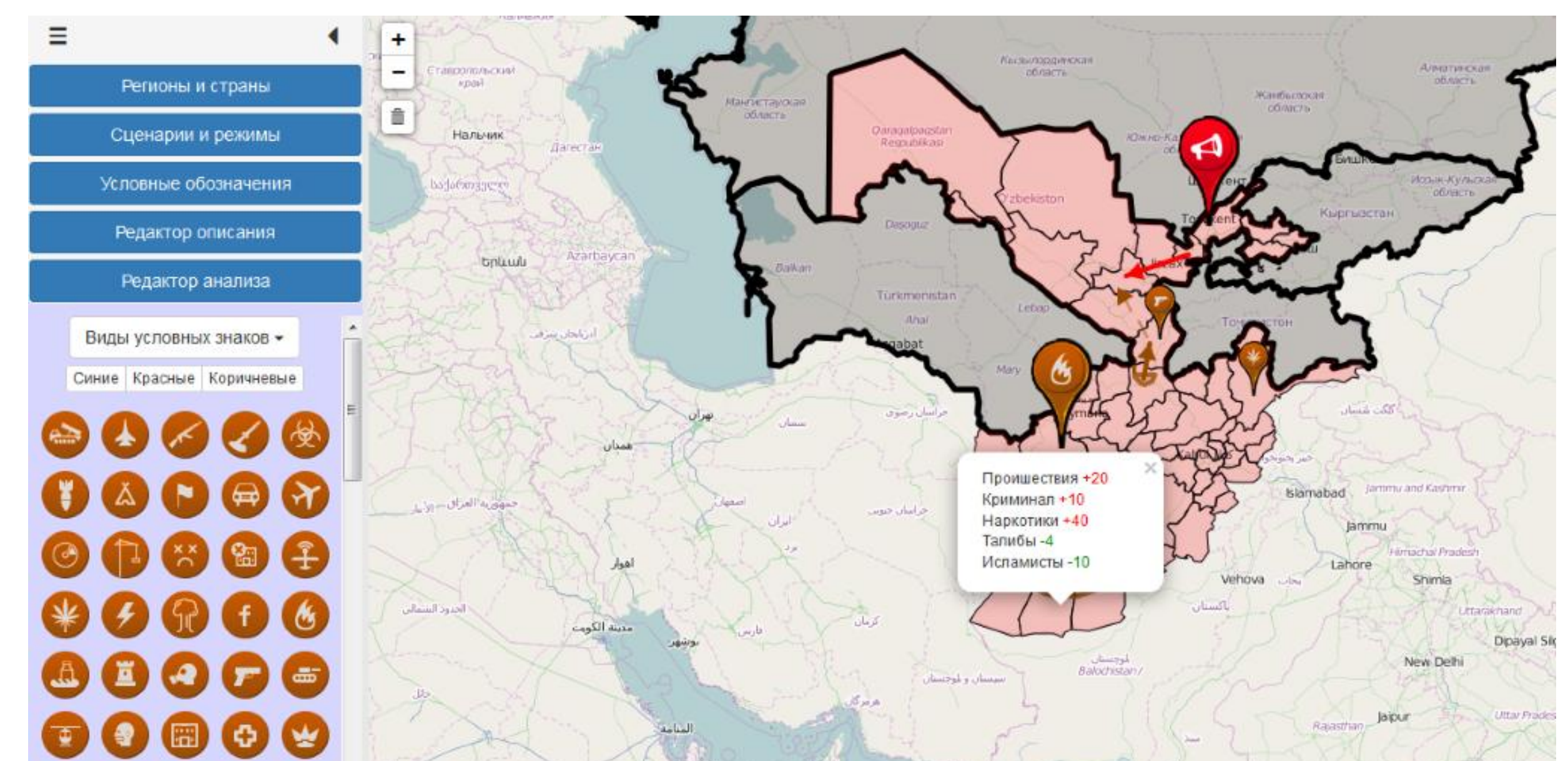
Отображение активности тем по
странам



Пример вывода прогноза



Редактор аналитической записки



Пример сценария