

Институт вычислительных технологий СО РАН, г. Новосибирск

# Модуль географического поиска информации для платформы ZooSPACE

**Скачков Д. М., Жижимов О. Л.**

# ZooSPACE

- Платформа ZooSPACE разрабатывалась в рамках проекта «Разработка принципов и программных средств виртуальной интеграции распределённых источников данных на основе международных стандартов для создания масштабных информационных инфраструктур» (шифр «2012-1.4-07-514-0022-004»). Работа по проекту была завершена несколько месяцев назад.
- Подробнее: *О.Л. Жижимов, В.С. Никульцев, Е. В. Никульцева, А. М. Федотов, Ю. И. Шокин* «Технологическая платформа интеграции разнородных распределенных данных ZooSPACE» - Материалы конференции «Крым-2013»

# ZooSPACE. Интерфейс выбора источников для поиска

ZooSPACE Home x

elib.sbras.ru/ZooSPACE/

Распределенная система интеграции данных ZooSPACE

Версия 0.1, ИВТ СО РАН, 2012

undefined

Простой Расширенный Эксперт **Географический** Рубрикаторы

Отметьте нужные базы данных и перейдите на требуемый тип поиска

**ИВТ СО РАН (db4.sbras.ru:210)**

<input type="checkbox"/>	<input checked="" type="checkbox"/>	Электронная библиотека ИВТ СО РАН
<input type="checkbox"/>	<input checked="" type="checkbox"/>	ИВТ СО РАН: Тестовая БД для ВИНИТИ МЕКОФ
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Конференции СО РАН
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Материалы конференций СО РАН
<input type="checkbox"/>	<input checked="" type="checkbox"/>	ИВТ СО РАН: Труды сотрудников
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Научные Проекты СО РАН
<input type="checkbox"/>	<input checked="" type="checkbox"/>	БД ЦИТИС
<hr/>		
<input type="checkbox"/>	<input type="checkbox"/>	БД Искусство Сибири
<input type="checkbox"/>	<input type="checkbox"/>	БД Памятники Пскова
<input type="checkbox"/>	<input type="checkbox"/>	БД Живопись музеев г.Рыбинска
<hr/>		
<input type="checkbox"/>	<input type="checkbox"/>	ИБЦ СО РАН: Каталог книг и продолжающихся изданий

? Другой сервер: Порт: 210 База: Формат: grs1 Schema: default

User: Domain: © ИВТ СО РАН, 2012 г.

# Географический поиск в ZooSPACE

ZooSPACE Home

elib.sbras.ru/ZooSPACE/

Распределенная система интеграции данных ZooSPACE

Версия 0.1, ИВТ СО РАН, 2012

undefined

Нормализованное представление  y2 56.3165367; Тезаурус: Google

Набор атрибутов: Bib-1 x1 79.7167968; Покази карту x2 90.703125 Для: 29 - Subject-Local

Тип поиска: Области перекрываются y1 52.7495937; Поиск

Результат поиска по запросу:  
@attr 1=2060 @attr 2=7 @attr 4=202 {{{(52.74959372674114,79.716796875),(56.31653672211301,90.703125)}}

Электронная библиотека ИВТ СО РАН dspace 4 Просмотр Все

Всего по запросу найдено: 4 записей

db4.sbras.ru:210/zsgw/gis/goMap3.htm?myScript=setMapDataBR&s1=8&src1=document.fa.gcoord.v...

db4.sbras.ru:210/zsgw/gis/goMap3.htm?myScript=setMapDataBR&s1=8&src1=document.fa.gcoord.value&sx1=document.fa.x1.value&sy1=document.fa.y1.value

Правая кнопка ставит маркер, левая удаляет!

Показать Очистить Сохранить

Карта Спутник

Анжеро-Судженск Юрга Кировский р-н Кемерово Ленинск-Кузнецкий Белово Киселев Прокопьевское Новокузнецкое Заринск Новоалтайск Барнаул

Калининский р-н Новосибирск Кировский р-н Советский р-н Искитим

Начать с: 1 порцией 1

© ИВТ СО РАН, 2012 г.

# Фрагмент результатов поиска

Результат поиска по запросу:

@or @attr 1=2060 @attr 2=7 @attr 4=202 {{{(54.47642158429295,82.474365234375),(55.94919982336746,85.111083984375)}}} @attr 6=10 {thes1.php}

Электронная библиотека ИВТ СО РАН	dspace	282	<input type="button" value="Просмотр"/>	<input type="button" value="Все"/>
ИВТ СО РАН: Тестовая БД для ВИНТИ МЕКОФ	t_mkfvn	45	<input type="button" value="Просмотр"/>	<input type="button" value="Все"/>
Конференции СО РАН	konf	0	<input type="button" value="Просмотр"/>	<input type="button" value="Все"/>
Материалы конференций СО РАН	konfrep	3	<input type="button" value="Просмотр"/>	<input type="button" value="Все"/>
ИВТ СО РАН: Труды сотрудников	trnct	3	<input type="button" value="Просмотр"/>	<input type="button" value="Все"/>

Всего по запросу найдено: 333 записей

Новосибирскому заводу "Труд" - 100 лет

Новосибирский мегаполис

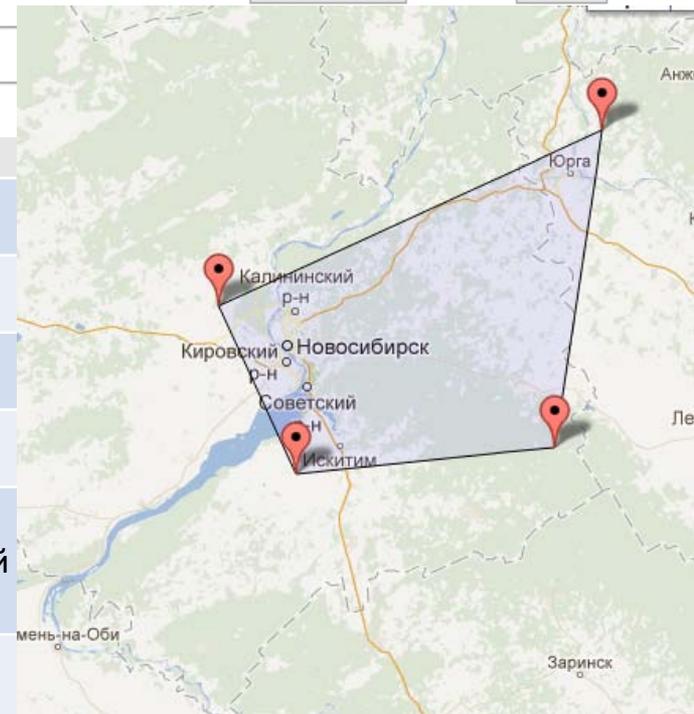
Влияние концентрации аэрозолей на качество атмосферы в г. Новосибирске

Новосибирский ученый удостоен «Глобальной энергии»

Кемеровская область заключила соглашение о создании на ее территории индустрии венчурного инвестирования и развития инновационных отраслей экономики

Кемеровская область занимает второе место по объему инвестиций в Сибирском федеральном округе

....



# Электронные библиотеки и географическая информация

- К ZooSPACE подключено большое количество систем, являющихся электронными библиотеками
  - Полнотекстовые базы данных
  - Реферативные журналы ВИНТИ
  - Базы материалов конференций
- Электронные библиотеки изначально не были предназначены для работы с географической информацией, в том числе и для географического поиска
- Тот факт, что эти системы не были предназначены для работы с географической информацией, еще не говорит, что эта информация там не содержится. Любая статья была где-то написана и опубликована, а тексты зачастую содержат названия географических объектов.

# Пример

## Заголовок

В **Иркутске** состоялась международная конференция "Управление земельными ресурсами с особым акцентом на защиту окружающей среды в районе **озера Байкал**"

Международная конференция "Ультрамафит-мафитовые комплексы складчатых областей докембрия" на **Байкале** п. **Энхалук**, 6-9 сент., 2006

Тенденции изменений климата в г. **Улан-Удэ**

Микробиологическая оценка состояния бассейна рек **Тугнуй-Сухара**

О состоянии изученности разнообразия папоротников **Байкальской Сибири** (на примере **Республики Бурятия**)

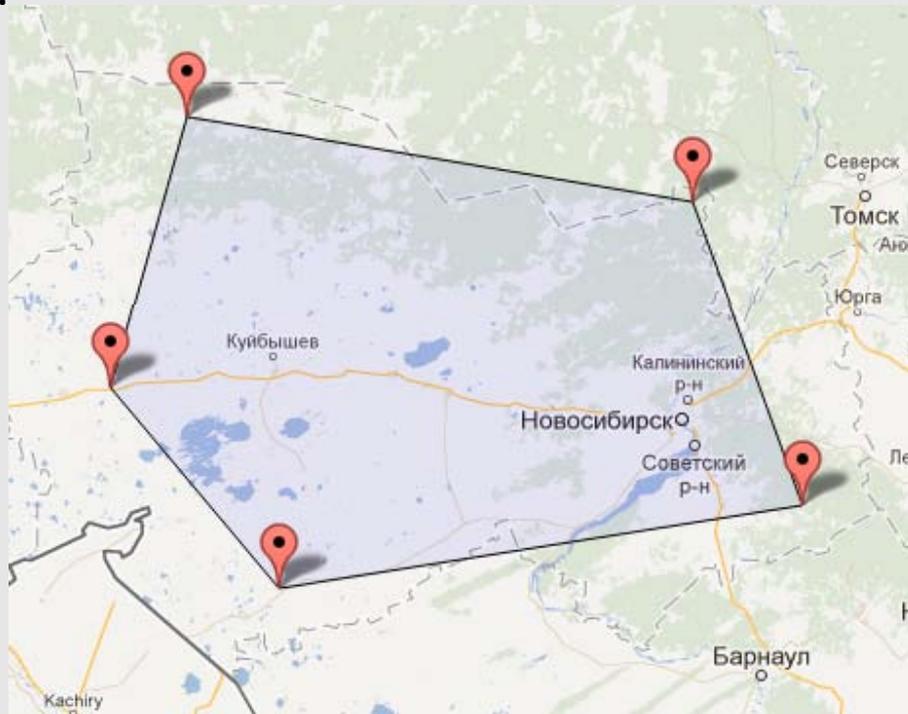
Некоторые результаты исследований динамики плессово-перекатных участков на **р. Оби** в районе **г. Барнаула**

Оценка перспективности минерально-сырьевой базы строительных материалов в связи с ускоренным развитием нефтегазоносных районов **Томской области**

Минерально-вещественный состав пылеаэрозолей на территории г. **Томска**

# Использование географической информации затруднено

- Использовать содержащийся в таких системах географический аспект затруднительно, поскольку, в соответствии с правилами каталогизации (Dublin Core, RUSMARC, МЕКОФ и др.), географическая информация хранится в текстовых полях и пригодна только для простейшего текстового поиска по географическому названию.
- А удобнее было бы искать так:



# Существующие решения

1. GeoNetwork (<http://geonetwork-opensource.org/>)
  2. Среда интеграции пространственных данных «GeoМета» (<http://www.geometa.ru/>)
  3. MultiMeta - Система интеграции пространственных данных и ресурсов электронных библиотек (<http://ceur-ws.org/Vol-803/paper4.pdf>)
- Но библиотека изначально должна использовать такую систему. Чтобы использовать старые массивы информации, их необходимо **конвертировать** и **дополнить** географическими координатами.

# Подходы к географическому поиску

1. Поиск с помощью атрибутивного и полнотекстового поиска по географическому названию.
2. Поиск с использованием координат, непосредственно заданных в записях информационной системы (характерно для ГИС).
3. Поиск с помощью метапоисковой машины, использующей:
  1. **специализированный справочник (тезаурус) географических названий;**
  2. тезаурус географических названий и промежуточную БД связи, организованную с использованием ссылок на записи тезауруса географических названий.

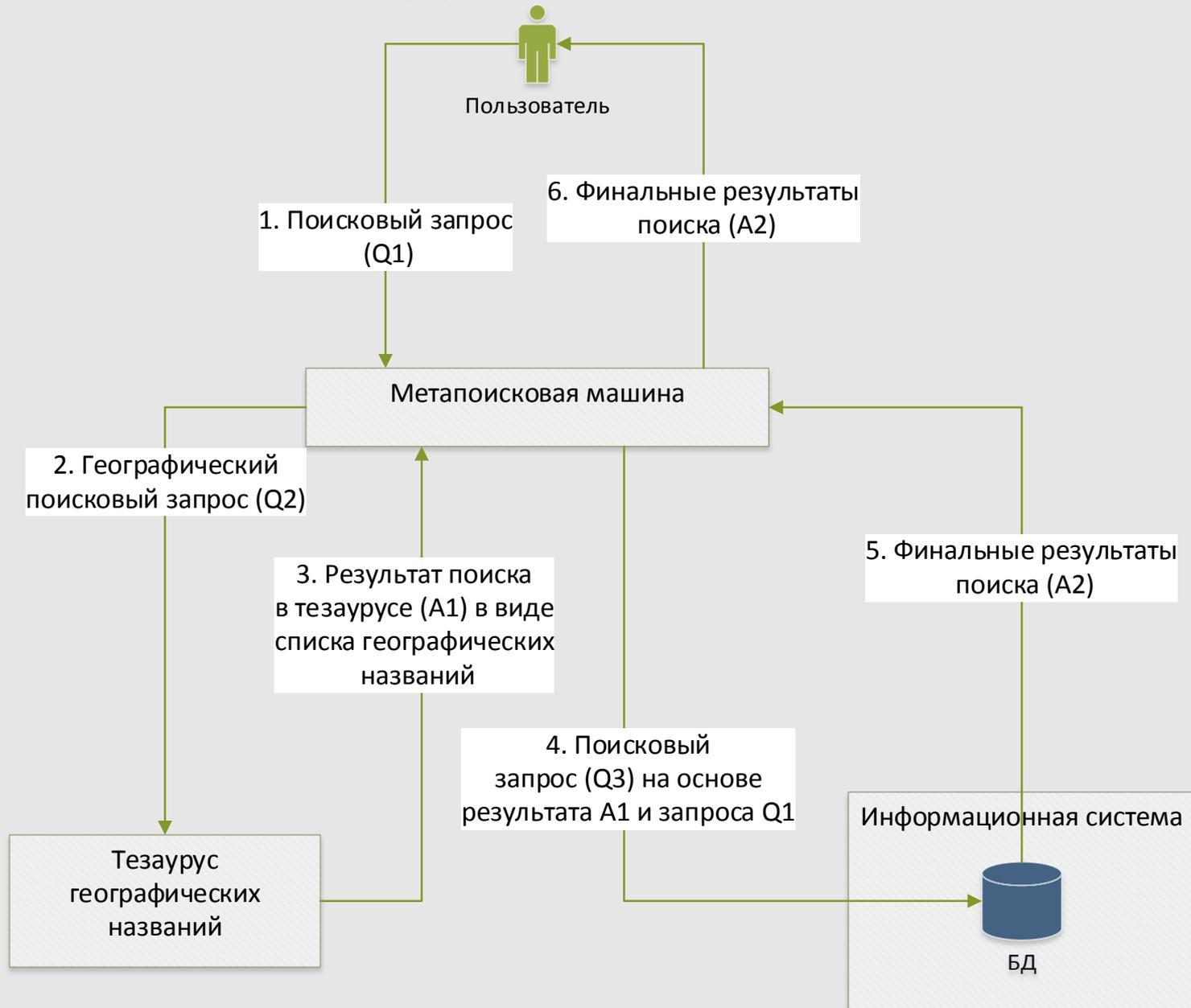
**Метапоисковая машина** – поисковая система, не имеющая собственной базы данных и поискового индекса, и формирующая поисковую выдачу из результатов поиска других поисковых систем.

**Информационно-поисковый тезаурус** – контролируемый словарь терминов на естественном языке, явно указывающий отношения между терминами и предназначенный для информационного поиска.

# Недостатки первых двух способов

1. Полнотекстовый поиск предполагает составление пользователем списка всех географических объектов, лежащих в географическом регионе поиска. С учётом их исторических изменений, альтернативных названий, названия исчезнувших объектов.
2. Использование поиска с использованием географического индекса подразумевает реализацию ГИС функциональности во всех существующих системах.
  1. Не все хранилища данных, на которых построены существующие информационные системы содержат функциональность по обработке и использованию географических координат.
  2. Необходимая поисковая логика должна быть реализована во всех целевых информационных системах по отдельности. Реализация такой логики потребует существенных изменений в структуре информационной системы.

# Поиск с использованием метапоисковой машины и тезауруса географических названий



# Метапоисковая машина

- Основная идея использования метапоисковой машины в преобразовании запроса, содержащего географическую компоненту, в запрос, не содержащий географическую компоненту, чтобы он мог быть обработан целевой информационной системой.
- В роли метапоисковой машины и выступает ZooSPACE

# Пример преобразования запроса

@attr 1=4 @attr 2=7 @attr 4=202 {((52.2, 80.1), (55, 81.0))}



@and @and @and @and @and...

@attr 1=4 @attr 2=3 @attr 4=1 {Баевский район}

@attr 1=4 @attr 2=3 @attr 4=1 {Баевского района}

@attr 1=4 @attr 2=3 @attr 4=1 {Баевскому району}

@attr 1=4 @attr 2=3 @attr 4=1 {Баевском районе}

@attr 1=4 @attr 2=3 @attr 4=1 {Панкрушиха}

@attr 1=4 @attr 2=3 @attr 4=1 {Панкрушихе}

@attr 1=4 @attr 2=3 @attr 4=1 {Панкрушихи}

@attr 1=4 @attr 2=3 @attr 4=1 {Родино} ...

# Требования к информационной системе

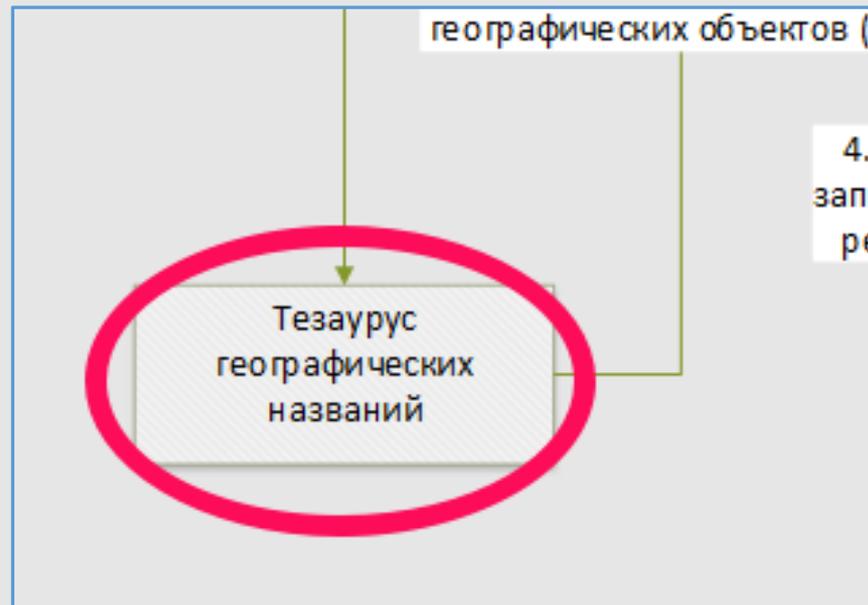
1. Для организации поиска способом 3.1) информационная система должна удовлетворять следующим требованиям:
  1. должна поддерживать поиск данных по протоколу Z39.50 (обязательный синтаксис запросов RPN-1);
  2. атрибуты, в которых может содержаться географическая информация, должны быть доступны для поиска.

Большинство электронных библиотек уже удовлетворяет этим требованиям.

Для оставшихся – реализовать требования сравнительно несложно.

# Тезаурус

- Каким должен быть тезаурус для использования в задаче географического поиска в «негеографических» информационных системах



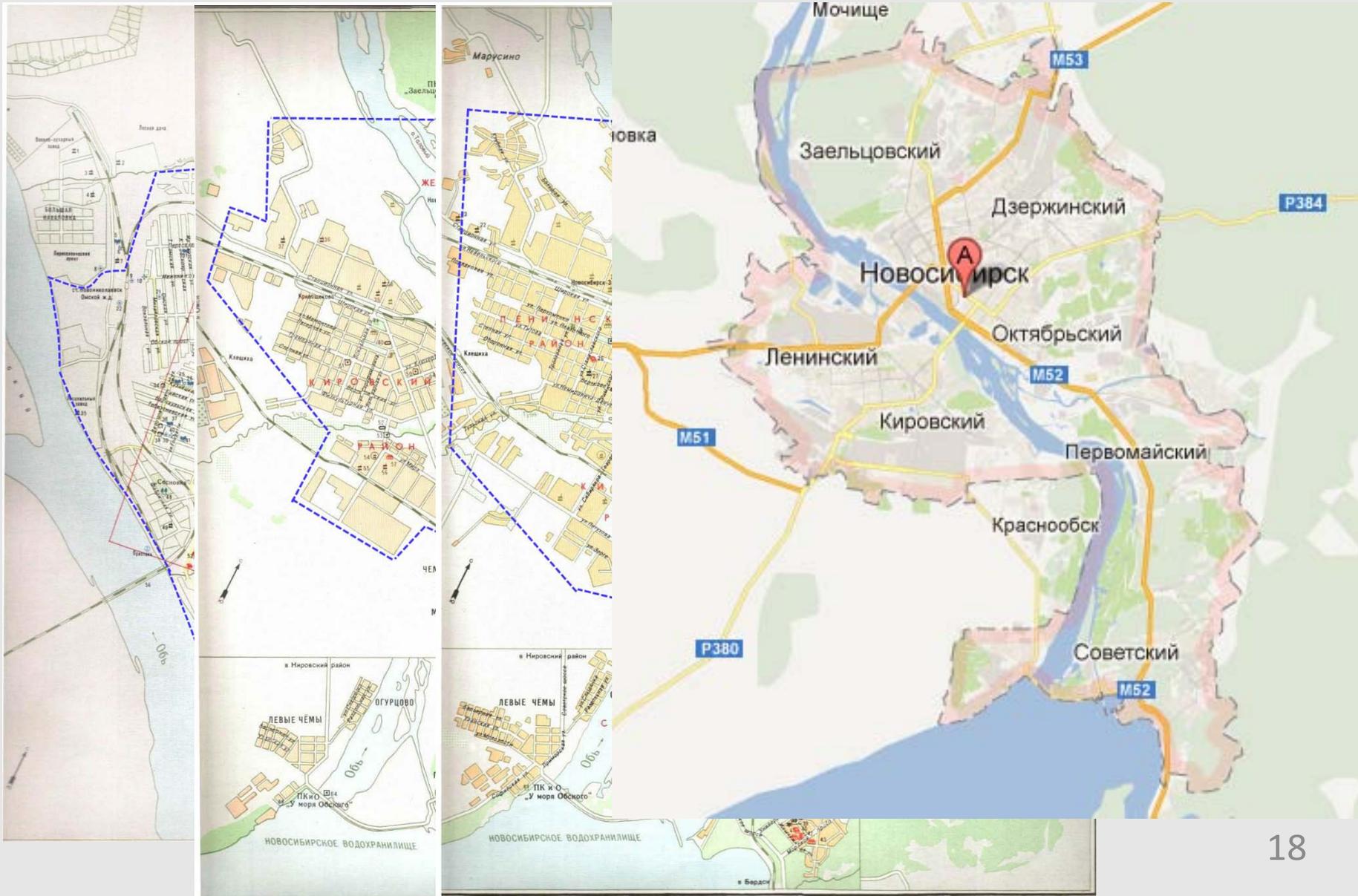
# Геокодирование и ретроспективное геокодирование

- Информация, хранящаяся в информационных системах относится к прошлому. Поэтому необходимо учитывать изменение свойств географических объектов с течением времени:

Исторический период	Название
до 1895	Александровский
1895 - 1903	Ново-Николаевский
1903 - 1917	Ново-Николаевск
1917 - 1926	Новониколаевск
с 1926	Новосибирск

- **Геокодирование** - перевод из терминов географических названий в термины географических координат.
- **Обратное геокодирование** - перевод из терминов географических координат в термины географических названий.
- **Ретроспективное геокодирование** - перевод из терминов географических названий в термины географических координат с учетом исторических изменений свойств географических объектов.

# Изменение геометрии географических объектов



# Доступ к тезаурусу

- Поскольку ZooSPACE построен на основе Z39.50, наиболее удобно организовать доступ к тезаурусу по протоколу Z39.50
- Пример поисковых запросов:
  - Географические объекты в указанном географическом регионе в определенный временной период  
@and @and  
@attr 1=2059 @attr cip 2=7 @attr cip 4=202  
**{{(53.3590,75.2152),(57.2273,85.1248)}}**  
@attr 1=2074 @attr 2=16 @attr cip 4=210 {**2001-10-12**}  
@attr 1=2075 @attr 2=16 @attr cip 4=210 {**2007-01-10**}
  - Географические объекты на территории Новосибирской области в указанном временном периоде  
@and @and  
@attr 1=2 @attr 2=3 @attr 4=108 {**Новосибирская область**}  
@attr 1=2072 @attr 2=16 @attr cip 4=210 {**2001-10-12**}  
@attr 1=2073 @attr 2=16 @attr cip 4=210 {**2007-01-10**}

# Заключение

- Реализован прототип системы, иллюстрирующий технологический подход к организации географического поиска информации в разнородных источниках данных.
- Разработанная технология расширяет поисковые возможности платформы ZooSPACE, и, тем самым, предоставляет пользователю новые возможности по поиску информации.

Благодарю за внимание!