

Об использовании ретроспективного геокодирования для географического поиска в электронных библиотеках

© Д.М. Скачков, О.Л. Жижимов

Институт вычислительных технологий Сибирского отделения РАН,
г. Новосибирск
danil.skachkov@gmail.com, zhizhim@sbras.ru

Аннотация

Обсуждаются вопросы, связанные с построением интероперабельного тезауруса географических названий, включающего геометрические данные географических объектов, в том числе и ретроспективные. Приводятся основные требования к подобному тезаурусу, формулируются основные позиции соответствующего профиля для организации доступа к тезаурусу, приводится пример записей разработанного тезауруса.

1. Введение

В настоящее время в связи с возрастающей потребностью общества в информационном обеспечении, в том числе и связанным с географическим аспектом информации, все большую актуальность приобретают разработки, направленные на интеграцию «негеографических» информационных систем с информационными системами, изначально ориентированных на обработку географической информации (ГИС). Это касается в том числе и электронных библиотек. Добавление географического аспекта к информации, хранящейся в таких системах, как электронные библиотеки, позволяет существенно повысить функциональность навигационных, поисковых и визуализационных сервисов этих систем. Например, позволяет находить информацию, которая относится к конкретному географическому региону [13].

Следует заметить, что существующие в настоящее время программные комплексы для организации электронных библиотек не содержат необходимой функциональности по хранению и обработке географических данных. Наделение же их требуемой функциональностью осложняется отсутствием единых стандартов на поиск и представление данных, связанных с географическим

аспектом, которые бы сопрягались с существующими геоинформационными системами, т.е. с системами, для которых географический аспект информации является основным [12].

Таким образом, разработка технологии, обеспечивающей обработку географического аспекта информации в «негеографических» информационных системах общего назначения, в частности, в электронных библиотеках, является актуальной и перспективной.

2. Необходимость ретроспективного геокодирования

Далее мы будем считать, что географический аспект информации может быть зафиксирован на уровне метаданных, описывающих контент и контекст. При этом «географические» метаданные объекта могут быть заданы двумя способами:

- с помощью количественного геометрического описания географического объекта на основе координат;
- с помощью ссылки на элемент некоторого тезауруса, включающего географические названия соответствующих объектов.

Первый вариант является предпочтительным, так как исключает неоднозначные толкования, но, в то же время, он не очень удобен по причине необходимости внесения существенных изменений в уже существующие информационные системы. Второй вариант не является однозначным, но может быть реализован на базе существующих парадигм информационных систем при условии их небольшой модернизации. Поэтому ниже речь пойдет только о втором варианте.

Так как термин «тезаурус» может употребляться в различных значениях, в данной работе будем говорить об информационно-поисковом тезаурусе. [15]

Существует множество тезаурусов географических наименований, но сложность их использования заключается в том, что географический аспект объектов, хранящиеся в электронных библиотеках, зачастую относится не к текущему моменту времени, а к моментам времени прошедшим. В то время как большинство тезаурусов содержит информацию, относящуюся

только к текущему моменту времени. Но с течением времени могут изменяться как географические названия, так и границы географических объектов, что препятствует использованию существующих тезаурусов географических наименований в подобных системах.

Следует заметить, что любые изменения географических названий и геометрических объектов, ассоциированных с ними, как правило, привязываются к какому-либо нормативному документу.

Более того, в существующих тезаурусах координаты географического объекта чаще всего задаются в виде точки, в то время как реальные координаты объекта представляют собой далеко не точку, а, в общем случае, некоторую область. Что, конечно же, также уменьшает полезность таких тезаурусов при проведении поиска. Поэтому более предпочтительным будет тезаурус, где положение объектов задано с помощью координат границ области, занимаемой объектом.

Также, для задач поиска, полезными будут данные о том, как географические объекты расположены относительно друг друга. Например, если производится поиск по некоему региону, целесообразно считать релевантными также и элементы, относящиеся к географическим объектам, лежащим в целевом регионе.

Таким образом, для использования в информационных системах общего назначения (в электронных библиотеках в частности) географического аспекта в его любом виде необходим справочный аппарат (тезаурус), который бы включал в себя не только географический аспект информации, но и ее временной (исторический) аспект.

В данной работе сделана попытка сформулировать основные требования к подобному тезаурусу географических названий, который бы мог удовлетворить потребности существующих информационных систем по обработке географического и исторического аспекта информации. В работе приводится обзор некоторых схем существующих тезаурусов, анализ их сильных и слабых сторон (в контексте применения в задаче привязки географических метаданных к объектам информационных систем) и формулировка требований к тезаурусу географических наименований подходящему для использования в электронных библиотеках. Также приводятся примеры записей разработанного тезауруса, для иллюстрации рассмотренных особенностей тезауруса.

3. Обзор существующих решений

Рассмотрим существующие на данный момент схемы представления данных и существующие тезаурусы, которые могут представлять интерес в рамках этой задачи.

При рассмотрении будем обращать внимание на следующие свойства:

1. Наличие ретроспективных данных. Возможность извлечь данные, относящиеся к прошлому.

2. Наличие связей с нормативными документами. Возможность определить, согласно какому документу было изменено название или координаты объекта.

3. Описание координат географического объекта согласно его форме. Представление географического объекта не только в виде точки, а также в виде замкнутого контура, линии, композиции примитивов.

4. Наличие связей, отражающих относительное расположение географических объектов.

В первую очередь, рассмотрим существующие схемы данных. И, затем, существующие тезаурусы географических наименований. [5, 17]

3.1 ГОСТ Р 52573-2006

Национальный стандарт Российской Федерации "Географическая информация. Метаданные". Стандарт предназначен для специалистов в области информационных технологий, разработчиков геоинформационных систем, баз и банков пространственных данных, а также прикладных информационных систем различного назначения. Стандарт разработан в соответствии с правилами создания профилей, указанными в стандарте ISO 19115 [10].

Данный стандарт содержит рекомендацию к использованию ретроспективных данных (Сущность EX_Extent). Координаты объекта задаются с помощью одной из сущностей:

- EX_BoundingPolygon - Многоугольник (задается множеством точек)
- EX_GeographicBoundingBox - Прямоугольная область (задается координатами углов)
- EX_GeographicDescription - Описание объекта с использованием географического идентификатора.

Есть сведения о документе источнике, но они привязаны к объекту в целом, а не к данным о его координатах и наименовании.

Сведения о связях между объектами отсутствуют в данной схеме.

3.2 CIDOC CRM

"Committee on Documentation" "Conceptual Reference Model" - не является схемой тезауруса, но рассматривается в данной работе так как представляет собой формальную онтологию, предназначенную для улучшения интеграции и обмена гетерогенной информацией по культурному наследию. Более конкретно, CIDOC CRM определяет семантику схем баз данных и структур документов, используемых в культурном наследии

и музейной документации, в терминах формальной онтологии. Модель не определяет терминологию, появляющуюся в конкретных структурах данных, но имеет характерные отношения для ее использования.

Модель может служить как руководством для разработчиков информационных систем, так и общим языком для экспертов предметной области и специалистов по информационным технологиям. Она предназначена для покрытия контекстной информации исторического, географического и теоретического характера об отдельных экспонатах и музейных коллекциях в целом [16].

Структурно CIDOC CRM состоит из иерархии классов и широкого набора свойств (бинарных отношений), связывающих классы между собой. Все концепты (классы и свойства) модели можно разделить на три группы. Первая группа включает классы и отношения, охватывающие наиболее общие понятия окружающего мира: постоянные и временные сущности, отношения участия, зависимости, совпадения во времени. Вторая группа содержит понятия, частично поддерживающие функции управления: приобретение и учет единиц хранения, передача прав собственности на объекты культуры. К третьей группе относятся классы и свойства, используемые для внутренней организации самой онтологии: средства, необходимые для подключения внешних источников терминов, например, тезаурусов по отраслям культуры.

В CIDOC CRM нас интересует сущность E53_Place (Место), которая как раз описывает географические метаданные объекта. Данная сущность является экземпляром E44_Place_Appellation. E44_Place_Appellation содержит данные о координатах, адресе, географическом наименовании. Координаты могут задаваться в любом виде (не только географические). Присутствует возможность задать ссылки на родительский элемент (иерархические связи).

Но в онтологии CIDOC CRM не учтено изменение свойств географических объектов с течением времени, отсутствует связь географических метаданных с нормативными документами.

3.3 Getty Thesaurus of Geographic Names

Тезаурус географических имен Института Getty - Англоязычный тезаурус, содержащий более чем миллион географических имен, информацию о континентах, физических объектах, административных сущностях и нациях современного политического мира, а также сведения об исторически значимых областях [4].

Схеме тезауруса Getty, естественно, присущи как положительные, так и отрицательные черты.

Из отрицательных черт можно отметить отсутствие информации об изменении координат географических объектов с течением времени.

Координаты объекта могут быть либо точкой, либо прямоугольником, что недостаточно для полного описания области на поверхности земли.

В тоже время, в схеме данного тезауруса учтено временное изменение названия объекта с течением времени (Term_Date). Также учтены нормативные документы (Subject_Sources) для данного объекта и для его наименований (Term_Source) [3, 5]. Записи содержат данные об иерархии.

3.4 Российская Государственная Библиотека

Тезаурус географических названий Российской государственной библиотеки. Содержит наименование географических объектов (городов, рек, и т. д.) на территории Российской Федерации [21, 14].

Тезаурус не содержит ретроспективных данных в записях. Невозможно получить ни данных о предыдущих названиях, ни данных о предыдущих координатах объектов.

В записях присутствуют ссылки на нормативные документы, определяющие наименование объекта.

Координаты географических объектов заданы в виде координат точек, что не совсем соответствует действительности.

Из записей могут быть получены данные о иерархических связях с помощью обработки. Явным образом иерархические связи не указаны.

3.5 Служба геокодирования API Карт Google

Позволяет определить координаты объекта, а также найти адрес наиболее близкий к указанным координатам [9].

В записях, предоставляемых данной службой отсутствуют ретроспективные данные. Отсутствуют связи с нормативными документами.

Координаты объектов указаны в виде точки или в виде прямоугольной области.

В записях содержатся иерархические связи.

В то же время стоит отметить, что тезаурус содержит данные не только о крупных географических объектах, но так же и об адресах. Есть возможность произвести обратное геокодирование.

3.6 Служба геокодирования API Яндекс.Карт

Имеет функциональность, аналогичную геокодеру Google [18]. Обладает практически теми же достоинствами и недостатками, но из дополнительных достоинств можно выделить более обширную базу российских наименований географических объектов.

3.7 Сравнительная таблица

Составим сравнительную таблицу рассмотренных схем (Таблица 1).

Таким образом, проанализировав существующие решения, мы приходим к выводу, что схемы тезауруса с необходимой нам функциональностью нет.

Таблица 1. Сравнительная таблица схем тезаурусов

Схема/тезаурус	Содержит ретроспективные сведения	Содержит ссылки на документные источники	Координаты географических объектов заданы соответственно их размерам и форме	Наличие иерархических связей
ГОСТ Р 52573-2006	+	+	±	-
CIDOC CRM	-	-	±	+
Getty	+	+	±	+
Российская государственная библиотека	-	+	-	±
Геокодер Карт Google	-	-	±	+
Геокодер Яндекс.Карт	-	-	±	+

Но есть достаточно близкие схемы, которые можно изменить для реализации желаемой функциональности. Наиболее подходящей схемой является схема тезауруса географических наименований Getty. Её и возьмем за основу.

4. Ретроспективный тезаурус

4.1 Онтология тезауруса

Сформулируем список требований тезаурусу, подходящему для использования при внедрении в электронные библиотеки.

1. Структура тезауруса должна позволять эффективно решать следующие задачи:
 - a. Прямое и обратное геокодирование
 - b. Ретроспективное прямое и обратное геокодирование
 - c. Позволять включать информацию в технологию поиска в существующих информационных массивах
2. Содержать внутренние связи:
 - a. по географическим объектам
 - b. по временным характеристикам
 - c. по документам
3. Тезаурус должен быть представлен в схеме, максимально приближенной к какой-либо стандартной
4. Он должен однозначно отображаться на другие схемы тезаурусов, в частности необходимо однозначное соответствие профилю Z-Thes [8], быть может, расширенному, для интеграции с существующими информационными системами.

Для упрощения задачи проектирования схемы тезауруса рассмотрим основные сценарии использования тезауруса географических наименований.

Работа с тезаурусом включает два основных сценария: запрос координат геометрического примитива объекта по имени этого объекта и запрос всех имен объектов по заданному координатно привязанному геометрическому примитиву. Обычно это называется прямым и обратным геокодированием. Для электронных библиотек, которые потенциально могут содержать ретроспективную информацию, отличительной особенностью становится необходимость указания времени, для которого соответствующее геокодирование будет актуальным. При этом отсутствие задания момента времени может служить указанием на использование текущего момента времени в качестве параметра запроса.

Существенным моментом прямого геокодирования, является тот факт, что заданные в запросе имя и время могут быть взаимно противоречивы. Например, запрос на координаты объекта (Новосибирск, 19200520) должен возвращать ответ (Новониколаевск, {координаты геометрического примитива}, 19200520). Запросы обратного геокодирования в этом смысле более просты, т.к. задаваемые в запросе координаты не связаны с действующей топонимикой.

Заметим, что правильно организованный тезаурус географических названий может служить основой и для получения и другой информации. В частности,

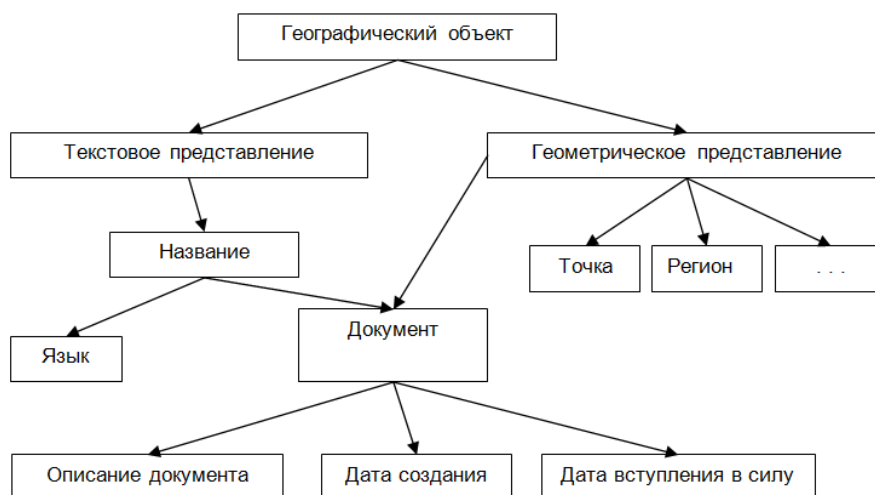


Рисунок 1. Онтология тезауруса

- информации о документах, связанных с конкретным географическим объектом
- информации о времени актуальности названий объектов
- информации о времени актуальности координат объектов
- а также о временных характеристиках производных параметров

Следует также заметить, что любой тезаурус является лишь дополнительной базой данных, которая может быть задействована при обработке запросов к различным информационным массивам. Ретроспективный тезаурус географических названий может быть задействован при обработке запросов, включающих ретроспективные географические названия.

Можно выделить три вида условий в запросе к тезаурусу:

1. По имени
2. По координатам
3. По времени

Данные условия могут комбинироваться друг с другом.

На основании приведенных выше данных, была построена онтология тезауруса, отвечающего приведенным требованиям (Рисунок 1) [20, 14].

На основании данной онтологии и схемы тезауруса географических названий Getty была построена схема тезауруса, подходящая для использования при внедрении в электронные библиотеки.

4.2 Профиль тезауруса

Для интеграции с существующими информационными системами и обеспечения интероперабельности необходимо зафиксировать профиль доступа к обсуждаемому тезаурусу (RGeoThes). Этот профиль, несомненно, должен являться расширением профиля ZThes [8] для доступа к тезаурусам по протоколам Z39.50 и SRW/SRU и включать необходимые компоненты

для временного и географического поиска. При этом профиль должен определять

- схему данных
- структуру записи и наборы элементов
- обязательные и дополнительные индексы (точки доступа)
- синтаксис поисковых запросов и поисковые атрибуты
- форматы представления данных
- протоколы доступа к ресурсу

Протоколы доступа

Для обеспечения интероперабельности доступ к RGeoThes должен обеспечиваться по протоколам

- Z39.50
- HTTP/XML/SOAP/SRW
- HTTP/SRU

Каждый из указанных способов доступа имеет свои специфические особенности, которые должны быть определены общим профилем.

Форматы представления данных

В качестве основного обязательного формата представления записи RGeoThes для всех способов доступа является формат XML. Дополнительным необязательным форматом является HTML. Для доступа по Z39.50 также обязательным форматом является GRS-1. В качестве дополнительных (необязательных) форматов могут использоваться RUSMARC, MARC21 и др.

Схема данных

Схема данных определяется в терминах XML (XSD) и должна соответствовать онтологии, схематично представленной на Рисунке 1.

Индексы и точки доступа

Точками доступа записи RGeoThes должны быть элементы, представленные в таблице 2.

Синтаксис поисковых запросов и поисковые атрибуты

Для доступа по Z39.59 обязательным синтаксисом запросов должен являться RPN-1, необязательным – CQL. Для доступа по SRW/SRU обязательным синтаксисом запросов должен являться CQL, необязательным – RPN-1 (x-pquery).

Таблица 2. Точки доступа записи RGeoThes

Точка доступа	Набор	Тип	Значение
Локальный номер	utility	1	4
Название термина	cross-domain	1	1
Квалификатор термина	zthes-1	1	1
Тип термина	zthes-1	1	2
Статус термина	zthes-1	1	7
Категория термина	zthes-1	1	6
Язык названия	utility	1	3
Дата начала действия названия	cip-1	1	2072
		2	14,15,16,17,18
Дата окончания действия названия	cip-1	1	2073
		2	14,15,16,17,18
Документ, фиксирующий название	cross-domain	1	6
Тип геометрического объекта	cip-1	4	201, 202
Координаты геометрического объекта	cip-1	1	2059, 2060
		2	7,8,9,10
Дата начала действия определения геометрии	cip-1	1	2072
		2	14,15,16,17,18
Дата окончания действия определения геометрии	cip-1	1	2073
		2	14,15,16,17,18
Документ, фиксирующий определения геометрии	cross-domain	1	6
Комментарий	cross-domain	1	4
Идентификатор связанного термина	zthes-1	1	4

Поисковые атрибуты RPN для доступа по Z39.50 для обеспечения интероперабельности должны соответствовать поисковым атрибутам профиля Z-Thes из наборов zthes-1, utility, cross-domain (xd-1). Для поиска по времени и координатам должны использоваться атрибуты из набора cip-1. Аналогичное требование справедливо и для запросов CQL.

Соответствие поисковых атрибутов точкам доступа приведено в Таблице 2 [1].

Таким образом, реализация тезауруса для ретроспективного геокодирования в соответствии с профилем RGeoThes позволила бы решать все перечисленные выше задачи, оставаясь в рамках ограниченной интероперабельности с существующими системами.

5. Примеры записей

Для хранения данных о том, как географические объекты соотносятся друг с другом будем использовать иерархическую структуру записей тезауруса, т. е. каждая запись (кроме корневой) содержит одного родителя (географический объект, внутри которого расположен указанный объект).

Каждая запись содержит первичное название, являющееся основным названием данного объекта на локальном языке. Также запись может содержать множество дополнительных названий объекта. [6]

Так как необходимо будет ссылаться на элементы тезауруса из других систем, каждая запись должна иметь уникальный квалификатор.

Наиболее подходящим для данной задачи представляется использование UUID идентификатора [7].

Данный квалификатор уникально идентифицирует запись тезауруса о географическом объекте и может быть использован при внедрении географической информации в информационную систему.

Рассмотрим подробнее элемент, представляющий наименование географического объекта. Каждый элемент наименования содержит собственно название, тип объекта, код языка [2], и ссылки на нормативные документы, вводящие в действие данное название и отменяющее его. Привязка нормативных документов к наименованиям географических объектов обеспечивает ретроспективную составляющую названий.

Запись, представляющая документ содержит поля описания документа, его URI, дату создания документа и дату вступления изменений в силу.

Так как координаты географических объектов также изменяются с течением времени, в элементы, представляющие географические координаты объектов, также добавлены ретроспективные составляющие в виде ссылок на нормативные документы.

В завершение приведем пример нескольких связанных записей, данные для которых были извлечены из Реестра географических названий [19], и немного дополнены для большей наглядности:

```

Record {
  "qualifier":
    "8c26bc72-31b3-4cfb-bbae-6d731057bcd1",
  "previous": {
    "qualifier": ". . . "
    "document": {
      "description": ". . . "
    }
  },
  "primaryName":
  {
    "name": "Агафониха",
    "type": "деревня",
    "language": "ru",
    "beginDocument": {
      "description": "unknown document"
    },
    "endDocument": {
      "description": "Решение Новосибирского
        областного Совета депутатов №4503
        от 20/09/2002.",
      "date": "Sep 20, 2002 "
    }
  },
  "names": [ {
    "name": "Агафониха1",
    "type": "деревня",
    "language": "ru",
    "beginDocument": {
      "description": "unknown document"
    },
    "endDocument": {
      "description": "unknown document"
    }
  },
  {
    "name": "Агафониха2",
    "type": "деревня",
    "language": "ru",
    "beginDocument": {
      "description": "unknown document"
    },
    "endDocument": {
      "description": "unknown document"
    }
  }
],
  "locations": [{
    point1: {
      "latitude": 54.933333,
      "longitude": 84.133333,
    },
    point2: {
      "latitude": 55.533333,
      "longitude": 85.933333,
    },
    "beginDocument": {
      "description": "unknown document"
    },
    "endDocument": {
      "description": "unknown document"
    }
  }],
  "primaryParent":
    "3c42885f-b842-49d4-acc2-dfe1a6220227",
  "contain": [{
    "qualifier": ". . . "
    "document": {
      "description": ". . . "
    }
  },],
  "belongsTo": [{
    "qualifier": ". . . "
    "document": {
      "description": ". . . "
    }
  }
],]
}

Record {
  "qualifier":
    "3c42885f-b842-49d4-acc2-dfe1a6220227",
  "previous": {
    "qualifier":
      "8c26bc72-31b3-4cfb-bbae-6d731057bcd1"
    "endDocument": {
      "description": "Решение Новосибирского
        областного Совета депутатов
        №4503 от 20/09/2002.",
      "date": "Sep 20, 2002 "
    }
  },
  "primaryName": {
    "name": "Агафончиха",
    "type": "район",
    "language": "ru",
    "beginDocument": {
      "description": "unknown document"
    },
    "endDocument": {
      "description": "unknown document"
    }
  },
  "names": [],
  "locations": [{
    "latitude": 55.091189,
    "longitude": 84.265925,
    "beginDocument": {
      "description": "unknown document"
    },
    "endDocument": {
      "description": "unknown document"
    }
  }],
  "primaryParent":
    "318517ed-ffcf-4125-b576-98184a19f162",
  "contain": [{
    "qualifier":
      "8c26bc72-31b3-4cfb-bbae-6d731057bcd1"
    "document": {
      "description": "..."
    }
  },],
  "belongsTo": [{
    "qualifier": "..."
    "document": {
      "description": "..."
    }
  }
],]
}

Record {
  "qualifier":
    "318517ed-ffcf-4125-b576-98184a19f162",
  "primaryName": {
    "name": "Новосибирская область",
    "type": "область",
    "language": "ru",
    "beginDocument": {
      "description": "unknown document"
    },
    "endDocument": {
      "description": "unknown document"
    }
  }
}

Record {
  "qualifier":
    "3c42885f-b842-49d4-acc2-dfe1a6220227",
  "previous": {
    "qualifier":
      "8c26bc72-31b3-4cfb-bbae-6d731057bcd1"
    "endDocument": {
      "description": "Решение Новосибирского
        областного Совета депутатов
        №4503 от 20/09/2002.",
      "date": "Sep 20, 2002 "
    }
  },
  "primaryName": {
    "name": "Агафончиха",
    "type": "район",
    "language": "ru",
    "beginDocument": {
      "description": "unknown document"
    },
    "endDocument": {
      "description": "unknown document"
    }
  },
  "names": [],
  "locations": [{
    "latitude": 55.091189,
    "longitude": 84.265925,
    "beginDocument": {
      "description": "unknown document"
    },
    "endDocument": {
      "description": "unknown document"
    }
  }],
  "primaryParent":
    "318517ed-ffcf-4125-b576-98184a19f162",
  "contain": [{
    "qualifier":
      "8c26bc72-31b3-4cfb-bbae-6d731057bcd1"
    "document": {
      "description": "..."
    }
  },],
  "belongsTo": [{
    "qualifier": "..."
    "document": {
      "description": "..."
    }
  }
],]
}

Record {
  "qualifier":
    "318517ed-ffcf-4125-b576-98184a19f162",
  "primaryName": {
    "name": "Новосибирская область",
    "type": "область",
    "language": "ru",
    "beginDocument": {
      "description": "unknown document"
    },
    "endDocument": {
      "description": "unknown document"
    }
  }
}

```

6. Заключение

Изложенные выше основные положения организации ретроспективного геокодирования и соответствующего тезауруса географических названий будут в дальнейшем использованы для построения модели информационной системы и возможностями геометрического и ретроспективного поиска информации на основе

картографических интерфейсов в соответствии с описанным выше профилем.

Работа выполнена при поддержке РФФИ, грант № 10-07-00302-а.

Литература

- [1] Catalogue Interoperability Protocol (CIP) Specification - Release B // CEOS/WGISS/ICS/CIP-B, Issue 2.4.75. - April 2005.
- [2] Codes for the Representation of Names of Languages. - http://www.loc.gov/standards/iso639-2/php/code_list.php
- [3] Contribute to the Getty Vocabularies. - <http://www.getty.edu/research/tools/vocabularies/contribute.html>
- [4] Getty Thesaurus of Geographic Names® Online. - <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>
- [5] Introduction To Metadata. - http://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html
- [6] ISO 5964:1985. Guidelines for the establishment and development of multilingual thesauri. - http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?ics1=01&ics2=140&ics3=20&csnumber=12159
- [7] RFC 4122 - A Universally Unique Identifier (UUID) URN Namespace. - <http://tools.ietf.org/html/rfc4122>
- [8] The Zthes specifications for thesaurus representation, access and navigation. - <http://zthes.z3950.org/>
- [9] Геокодирование - Службы API Карт Google. - <http://code.google.com/intl/ru/apis/maps/documentation/geocoding/>
- [10] ГОСТ Р 7.24-2007. Тезаурус Информационно-поисковый многоязычный. Состав, структура и основные требования к построению. // Москва. Стандартинформ, 2006.
- [11] Жижимов О. Л., Мазов Л. А. Использование географических координат при информационном поиске НТИ в библиографических базах данных.
- [12] Жижимов О. Л., Мазов Л. А. Проблемы географической привязки цифровых объектов в электронных библиотеках. // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Труды XII Всероссийской научной конференции RCDL'2010; Казань, Россия, 2010 г.
- [13] Жижимов О.Л., Мазов Н.А. География и стандарты метаданных для электронных библиотек: содержание, применение, проблемы [Электронный ресурс] // Электронные библиотеки, 2009.- т.12, №1 - <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2009/part1/ZM>
- [14] Лаврёнова О.А. Многоязычный доступ к данным на основе тезауруса географических названий // Сборник тезисов постерных докладов 9-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2007 (Переславль-Залесский, Россия, 15-18 октября 2007 г), Переславль-Залесский: изд-во "Университет города Переславля", 2007. - с.57-62 - http://rcdl.ru/doc/2007/paper_56_v1.pdf
- [15] Лукашевич Н.В. Тезаурусы в задачах информационного поиска. - М.: Издательство Московского университета, 2011. - 512 с.
- [16] Онтология в области документации в сфере культурного наследия: CIDOC CRM. - <http://www.intuit.ru/department/expert/ontoth/5/>
- [17] Перечень классификаторов ИР сферы образования. - <http://doc.unicor.ru/classifiers/classifierslist.htm>
- [18] Поиск по карте - Яндекс.Карты. - <http://api.yandex.ru/maps/geocoder/>
- [19] Реестр нормализованных названий ранее существовавших географических объектов, зарегистрированных в АГКГН на 18/11/2010 Новосибирская область. - http://www.rosreestr.ru/upload/www/files/Новосибирская_обл_3.pdf
- [20] Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В. Онтологии и тезаурусы. Учебное пособие. // Казань, Москва, 2006.
- [21] Тезаурус РГБ. - http://aleph.rsl.ru/F/?func=file&file_name=find-b&local_base=tst11

On Using of Retrospective Geocoding for Geographical Search in Digital Libraries

© O.L. Zhizhimov, D.M. Skachkov

The issues related to construction of the thesaurus of place names, containing geometrical data of geographical objects including retrospective, are discussed. Basic requirements to the thesaurus are described and the basic positions of appropriate profile for providing access to the thesaurus are formulated. An example of description from the developed thesaurus is described.