

# ABOUT LANGUAGES OF SEARCH QUERIES FOR SPATIAL DATA

*Oleg L. Zhizhimov*

Institute of Computational Technologies SB RAS, Novosibirsk, Russia

## **Abstract**

On a simple example of a query for searching by spatial characteristics, some languages of search queries are considered. Some weak and strong sides of the search query languages are discussed.

*Keywords: search query languages, spatial data, SQL, CQL, SparQL, PQF, RPN*

# ОБ ЯЗЫКАХ ПОИСКОВЫХ ЗАПРОСОВ ДЛЯ ПРОСТРАНСТВЕННЫХ ДАННЫХ

Жижимов О.Л.

Институт вычислительных технологий СО РАН, Новосибирск

На простом примере запроса для поиска по пространственным характеристикам рассмотрены некоторые языки поисковых запросов. Обсуждаются некоторые слабые и сильные стороны языков поисковых запросов.

*Ключевые слова:* языки поисковых запросов, пространственные данные, SQL, CQL, SparQL, PQF, RPN.

**Введение.** Традиционно языки поисковых запросов попадают в две группы: мощные, выразительные языки, не очень легко читаемые и не очень удобные для построения запросов специалистами (например, SQL, PQF и XQuery), или простые и интуитивные языки, недостаточно мощные для выражения сложных понятий (например, CCL и Google). Ниже рассмотрены некоторые языки поисковых запросов, которые содержат функциональность поиска по записям баз данных, содержащих пространственные характеристики. Список рассмотренных языков не является полным. Более того, в нем содержатся только те языки, которые соответствуют некоторым отраслевым или международным стандартам.

Полное описание каждого языка, а также рассмотрение всех возможностей и недостатков, выходит за рамки настоящего доклада. Однако, основные свойства могут быть проиллюстрированы на простых примерах. Поэтому для каждого языка приводится один простой пример – сформулировать запрос для поиска всех записей, имеющих отношение к области, ограниченной поисковым термом. При этом поисковый терм представляет собой прямоугольник.

**SQL.** Язык запросов SQL (Structured Query Language) является стандартным языком запросов для реляционных СУБД. Вполне естественно, что в этот язык были включены расширения, которые позволяют оперировать с пространственными данными. Формально эти расширения основаны на спецификациях OGC [1] и оформлены в виде документов ISO [2]. Однако, реализации этих расширений для различных СУБД различны.

*Пример 1.* Найти все записи, имеющие отношение к области, ограниченной указанным термом:

Для PostgreSQL (без расширений):

```
select * from my_table
      where the_field && box '((52.2,80.3),(50.8,81.0))';
```

Для MySQL:

```
select * from my_table where MBROverlaps(the_geom,
      ST_GeomFromText('LINESTRING (52.2 80.3, 50.8 81.0)');
```

```
select * from my_table where ST_Overlaps(the_geom,
      LINESTRING (52.2 80.3, 50.8 81.0));
```

Для MS SQL Server:

```
select * my_table where
      the_geom.STIntersection(geometry::STGeomFromText(
      'LINESTRING (52.2 80.3, 50.8 81.0)', 0));
```

Несомненно, SQL с учетом расширений предоставляет мощные средства работы с пространственными данными. К сожалению, привязка синтаксиса языка к названиям конкретных полей и таблиц, а также различия в реализациях для различных СУБД, делает этот язык мало пригодным для распределенных разнородных систем, т.к. в этом случае требуется выполнение одного и того же запроса для различных СУБД.

**CQL.** Следует заметить, что под аббревиатурой CQL встречается более десятка различных типов поисковых запросов. В таблице приведены некоторые из них.

Аббревиатура CQL и языки поисковых запросов.

1.	CQL - Contextual Query Language [3]	Contextual Query Language является формальным языком для представления запросов к информационно-поисковым системам (веб-индексы, библиографические каталоги, электронные библиотеки и репозитории, и др.). Язык ориентирован на удобочитаемость и интуитивность при сохранении выразительности более сложных языков запросов.
2.	CQL - Common Query Language [4]	Язык запросов, созданный OGC для сервисов каталогов. Это человекочитаемый язык запросов и его синтаксис очень похож на синтаксис SQL, но с чрезвычайно ограниченными функциями.
3.	CQL - The Corpus Query Language [5]	Специальный язык запросов, используемый в Sketch Engine для поиска сложных грамматических или лексических шаблонов или для использования критериев поиска, которые невозможно установить с помощью стандартного пользовательского интерфейса.
4.	CQL - The Cassandra Query Language [6]	Cassandra Query Language предлагает модель, близкую к SQL в том смысле, что данные помещаются в таблицах, содержащей строку из столбцов.
5.	CQL - Continuous Query Language [7]	Continuous Query Language поддерживается STREAM – прототипом Data Stream Management System в Стэнфорде. CQL – это экспрессивный декларативный язык на основе SQL для регистрации непрерывных запросов для потоков и обновляемых отношений.
6.	CQL - Constraint Query Language [8]	Constraint Query Language – это Prolog интерфейс к SQL базам данных.
7.	CQL - Oracle Continuous Query Language [9]	Oracle Continuous Query Language (Oracle CQL) – язык, основанный на SQL с добавлением конструкций для поддержки потоковых данных.

Здесь и ниже аббревиатура CQL будет сопровождаться указанием индекса по таблице 1. Наиболее универсальным является CQL-1 (Contextual Query Language). Именно этот язык запросов используется в технологиях SRW/SRU [10], которые на сегодняшний день являются максимально адаптированными для использования в распределенных гетерогенных информационных системах.

*Пример 2.* Найти все записи, имеющие отношение к области, ограниченной указанным термом:

```
geo.bounds within/partial/geo.nwse "52.2 80.3 50.8 81.0"
```

В отличие от CQL-1, язык CQL-2 (Common Query Language) был разработан OGC для работы в том числе и с пространственными данными. Несмотря на наличие некоторых общих конструкций с CQL-1, CQL-2 отличается от него способом обращения с пространственными данными, при этом операции отношения представляются в виде специальных операторов (функций):

EQUALS, DISJOINT, INTERSECTS, TOUCHES, CROSSES,  
WITHIN, CONTAINS, OVERLAPS, RELATE, DWITHIN, BEYOND

Существенное отличие CQL-2 (и его расширения ECQL) от CQL-1 – использование в запросах не абстрактных индексов (CQL-1) или поисковых атрибутов (PQF, см. ниже), а названий реальных полей таблиц баз данных, содержащих пространственную информацию. При

этом разработчики языка CQL-2 и ECQL явно указывают на их схожесть с SQL. Следует заметить, что CQL-2 используется в GeoServer [11] при построении cql-фильтров для запросов WMS.

*Пример 3.* Найти все записи, имеющие отношение к области, ограниченной указанным термом:

```
INTERSECTS (the_geom, 52.2, 80.3, 50.8, 81.0)
```

Здесь the\_geom – название поля, содержащего пространственную информацию.

Остальные языки запросов с общим названием CQL из таблицы 1 прямого отношения к пространственным данным не имеют.

**SparQL.** Развитие технологий Semantic Web и хранилищ Open Linked Data (LOD) не могло оставить в стороне возможность использования языка запросов SparQL для поиска информации, связанной с пространственными характеристиками объектов. Разработан документ OGC [12], который регламентирует расширения SparQL для работы с пространственными данными. В качестве иллюстрации ниже приведен пример запроса GeoSparQL.

*Пример 4.* Найти все записи, имеющие отношение к области, ограниченной указанным термом:

```
PREFIX my: <http://example.org/ApplicationSchema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>

SELECT ?f
  WHERE { ?f my:hasPointGeometry ?fGeom .
          ?fGeom geo:asWKT ?fWKT .
  FILTER (geof:sfWithin(?fWKT,
    "<http://www.opengis.net/def/crs/OGC/1.3/CRS84>
    Polygon ((52.2 80.3, 52.2 81.0,
              50.8 81.0, 50.8 80.3,
              52.2 80.3))"^^geo:wktLiteral))
  }
```

Трудно сказать, насколько удобен и выразителен этот язык. Однако есть мнение, что популярность технологий Semantic Web со временем будет возрастать. Список существующих реализаций поддержки GeoSparQL можно найти по ссылке [13].

**PQF.** Язык запросов PQF (Prefix Query Format) [14] разработан для текстового представления поисковых запросов RPN-1 (Reverse Polish Notation), применяемых в ISO-23950 и SIP. В основу языка положены наборы абстрактных поисковых атрибутов, которые отражают семантику поисковых термов, правила сравнения и ограничения области поиска. Каждый набор поисковых атрибутов имеет глобальный идентификатор (OID), регистрируемый в соответствующем агентстве. На сегодняшний день поисковые атрибуты для пространственного поиска включены в наборы атрибутов bib-1 (1.2.840.10003.3.1), gils (1.2.840.10003.3.5), geo (1.2.840.10003.3.9) и sip (1.2.840.10003.3.1000.99.1).

*Пример 5.* Найти все записи, имеющие отношение к области, ограниченной указанным термом – прямоугольной областью, заданной координатами:

```
1) @attrset geo @attr 1=2060
   @attr 2=7 @attr 4=202 ((52.2, 80.3), (50.8, 81.0))
2) @attr 1=1117
   @attr geo 2=7 @attr geo 4=202 ((52.2, 80.3), (50.8, 81.0))
```

Следует заметить, что язык запросов PQF (RPN-1) не привязан ни к конкретной модели данных, ни, тем более, к конкретной структуре данных. Это обеспечивает его высокую универсальность и возможность использования в распределенных гетерогенных информационных системах. Однако в виду высокой сложности реализации сегодня популярность этого языка падает, уступая лидерство более простому CQL-1.

**Заключение.** Следует заметить, что разные языки запросов ориентированы на разные информационные системы. Для некоторых из них (SQL, CQL-2) необходима информация о структуре данных (названия элементов, таблиц, полей), другие (CQL-1, RPN, PQF) ориентированы на работу с абстрактными объектами (поисковыми атрибутами, именованными индексами), которые определяются глобально в соответствующих профилях, идентифицируемых по URI или OID. Функциональность первой группы языков выше, но область применения уже. Для второй группы – наоборот. Для распределенных разнородных информационных систем можно использовать только языки второй группы.

В качестве примера реализации поддержки запросов как первой группы, так и второй, можно привести платформу ZooSPACE [15] с серверами ZooPARK-ZS с провайдерами данных (PostgreSQL, MySQL, MS SQL, Oracle). При этом серверы ZooPARK-ZS являются серверами приложений и обрабатывают запросы RPN, PQF и CQL-1, конвертируя их в соответствующих провайдерах данных в запросы SQL для конкретных СУБД (PostgreSQL, MySQL, MS SQL, Oracle).

*Работа выполнена при частичной поддержке гранта ведущих научных школ НШ-7214.2016.9.*

## ЛИТЕРАТУРА

- [1] OGC - Open Geospatial Consortium. <http://www.opengeospatial.org/> (дата обращения 20.07.2017).
- [2] Geographic information—Simple feature access—Part 2: SQL option, ISO 19125-2:2004, [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=40115](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40115) (дата обращения 20.07.2017).
- [3] CQL - The Contextual Query Language. <https://www.loc.gov/standards/sru/cql/> (дата обращения 20.07.2017).
- [4] CQL and ECQL. [http://docs.geoserver.org/latest/en/user/tutorials/cql/cql\\_tutorial.html](http://docs.geoserver.org/latest/en/user/tutorials/cql/cql_tutorial.html) (дата обращения 20.07.2017).
- [5] CQL – Corpus Query Language. <https://www.sketchengine.co.uk/documentation/corpus-querying/> (дата обращения 20.07.2017).
- [6] The Cassandra Query Language (CQL). <http://cassandra.apache.org/doc/latest/cql/index.html> (дата обращения 20.07.2017).
- [7] Arasu A., Babu S., Widom J. The CQL continuous query language: semantic foundations and query execution // The VLDB Journal - The International Journal on Very Large Data Bases, 2006. v.15, n.2, p.121-142.
- [8] CQL - Constraint Query Language - <http://www.swi-prolog.org/pldoc/man?section=cql> (дата обращения 20.07.2017).
- [9] Oracle CEP CQL Language Reference. 11g Release 1 (11.1.1). [https://docs.oracle.com/cd/E16764\\_01/doc.1111/e12048/intro.htm](https://docs.oracle.com/cd/E16764_01/doc.1111/e12048/intro.htm) (дата обращения 20.07.2017).
- [10] SRU- Search/Retrieve via URL - <http://www.loc.gov/standards/sru/index.html> (дата обращения 20.07.2017).
- [11] GeoServer 2.12.x User Manual / Open Source Geospatial Foundation <http://docs.geoserver.org/latest/en/user/index.html> (дата обращения 20.07.2017).
- [12] OGC GeoSPARQL - A Geographic Query Language for RDF Data / OGC 11-052r4, 2012, [https://portal.opengeospatial.org/files/?artifact\\_id=47664](https://portal.opengeospatial.org/files/?artifact_id=47664) (дата обращения 20.07.2017).
- [13] GeoSPARQL. Wikipedia, <https://en.wikipedia.org/wiki/GeoSPARQL> (дата обращения 20.07.2017).
- [14] Prefix Query Format / Index Data. <http://www.indexdata.com/yaz/doc/tools.html#PQF> (дата обращения 20.07.2017).
- [15] Жижимов О.Л., Федотов А.М., Шокин Ю.И. Платформа ZooSPACE - организация доступа к разнородным распределенным ресурсам // Электронные библиотеки. - 2014. - Т.17. - № 2. - ISSN 1562-5419. - <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2014/part2/ZFS> (дата обращения 20.07.2017).