# Синтаксис языка запросов

На верхнем уровне запрос представляет собой булево выражение с круглыми скобками и операциями (зарезервированными словами) **OR**, **XOR**, **AND**, **NOT**(перечисленными здесь в порядке убывания приоритета).

Нижний синтаксический уровень представляется последовательностями следующих элементов:

* **Термин или строка**

*Термин* – слово, число или знак пунктуации, не играющие роли синтаксических элементов более высоких уровней;

*Строка* – заключенная в двойные кавычки последовательность терминов и экранированных спецсимволов, разделенных пробелами (C-style string).

* **Множество терминов или строк**

Заключенный в фигурные скобки набор строк или терминов, разделенных запятыми: {a1, a2, … aN}

* **Функция**

Синтаксический элемент вида *имя\_функции* (*набор\_аргументов*)

*Имя\_функции* – последовательность символов латиницы;

Аргументы функции разделяются запятыми;

В качестве аргументов могут выступать термины, строки, множества терминов или другие функции.

## **Поддерживаемые функции**

### **Фразовые запросы**

|  |  |  |
| --- | --- | --- |
| Функция | Действие | Прим. |
| OrderedText | Поиск фраз с опционально заданным расстоянием между словами (по умолчанию – без пропусков, 0). Порядок слов важен. | OrderedText(мама, мыла, раму)  OrderedText(мама мыла раму)  OrderedText(10, мама мыла раму) |
| Text | То же, что и OrderedText, но без учета порядка слов.  В скобках на первом месте можно числом задать расстояние между токенами. | Text(город, Москва)  Text(4, город, Москва) Text(выбросить, Org()) |
| OrderedWildcard | То же, что и OrderedText, но трактует аргументы с символами ‘?’ или ‘\*’ как wildcard. | OrderedWildcard (Ивано\*)  OrderedWildcard (Пе?ов) |
| Wildcard | То же, что и Text, но трактует аргументы с символами ‘?’ или ‘\*’ как wildcard. | Wildcard(Ивано\*)  Wildcard(Пе?ов) |
| OrderedTextAnyDistance | То же, что и OrderedText, но расстояние между термами не важно (по умолчанию – большое число). | OrderedTextAnyDistance(мама мыла раму) |
| InSameSentence | Поиск фраз с опционально заданным расстоянием между словами в пределах предложения. Порядок термов неважен. Пропуски разрешены. | InSameSentence(фонд, валюта)  InSameSentence(OrderedTextAnyDistance(валютный, фонд)) – в пределах предложения, порядок важен.  InSameSentence(OrderedText(фонд, валюта)) – в пределах предложения, без пропусков |
| InSameParagraph | Поиск фраз с опционально заданным расстоянием между словами в пределах абзаца. Порядок термов неважен. Пропуски разрешены.  InSameParagraph() без аргументов имеет специальный смысл (см. в таблице примеров) | Аналогично InSameSentence. |
| Exact | Строгий поиск фразы без нормализации. | Exact(OrderedTextAnyDistance(банку огурцов)) |

### **Запросы объектов и размерных величин**

|  |  |  |
| --- | --- | --- |
| Функция | Действие | Прим. |
| Id | Поиск по идентификатору БД | Name(Id(306548)) |
| Metadata | Поиск по идентификатору сущности из БД | Metadata(47571) |
| Name | Функция поиска персон. Может использоваться как отдельно, так и в контексте фразовых запросов. В качестве аргументов принимает либо идентификатор персоны, либо имя, либо имя и роль | Name(7888)  Name(Юрий Гагарин)  Name(Гагарин, космонавт). |
| Org | Функция поиска организаций. Аргументы аналогичны Name, за исключением роли. | Аналогично Name. |
| Location | Функция поиска географических объектов. Аргументы аналогичны Name, за исключением роли. | Location(Москва)  Location(24F7D) |
| Trademark | Поиск всех упоминаний бренда и дочерних брендов. Возможен фразовый поиск. | Trademark(Id(C39066))  Trademark(F000716)  Trademark(red bull) |
| IndustrialSector | Поиск всех упоминаний индустриального сектора и связанных типов событий. Возможен фразовый поиск. | IndustrialSector(Id(F000716))  IndustrialSector(D000061)  IndustrialSector(органы госуправления) |
| IndustrialSectorPositiveTonality | Поиск всех положительных упоминаний индустриального сектора. Фразовый поиск невозможен. | Аналогично IndustrialSector |
| IndustrialSectorNegativeTonality | Поиск всех отрицательных упоминаний индустриального сектора. Фразовый поиск невозможен. | Аналогично IndustrialSector |
| NewsCluster | Поиск по новостному кластеру. | NewsCluster(Id(2C8AFDD))  NewsCluster(6129C4B3) |
| Date | Функция поиска дат или диапазонов дат, встречающихся в документе. Возможен фразовый поиск. | date(19.07.2011) – все документы, содержащие данную дату, или содержащие интервалы, в которую она входит.  date(01.01.1980, 01.07.2011) – все документы, в которых есть даты, попадающие в данный интервал, или интервалы дат, пересекающиеся с данным.  date(>=01.01.1980) – все даты, начиная с первого января 1980 года.  OrderedText(date(01.01.2010, 01.01.2011), произошло) – даты из интервала рядом со словом «произошло». |
| Dim | Функция поиска размерных величин, таких как суммы, массы, время. Размерные величины одного типа приводятся к общему спецификатору, например к миллиграммам для массы.  Возможен фразовый поиск. | dim(150<=тонн)  dim(тонн>150 )  dim(150<=тонн<300)  dim(150<=грамм<300)  dim(150<=рублей<300)  dim(150<=километров<300) |
| AuthorName | Поиск по автору документа или источнику в цитатах/прямой речи (совместно с функцией Speech) | AuthorName(21049559)  AuthorName(саша пятницкая) |

### **Запросы фактов, рубрик и прямой речи**

|  |  |  |
| --- | --- | --- |
| Speech | Функция поиска прямой речи.  В аргументе функции Speech, задающем фрагмент текста, возможно использование фразовых функций – OrderedTextAnyDistance, Text, OrderedText.  Функции InSameSentence и InSameParagraph могут использоваться, если автор не задан или задан идентификатором (набором идентификаторов). | Speech(AuthorName(Id(4AD74)), экономика)  Speech(AuthorName(Id(4AD74)), Text(экономика,рост)) |
| Event | Функция поиска событий заданных типов (тематик). В качестве параметров на данный момент доступны субъекты, объекты и тональность. Тональность бывает негативной и позитивной и может быть задана четырьмя вариантами:  негатив, negative, позитив, positive  Можно искать по типу события ("*action*"):   * Event(EventType(Экономические преступления)) | * Event(Subject(Name(Петров)), EventType(негатив)) – все негативные факты, относящиеся ко всем Петровым. * Event(Subject(Name(91FB))) – все факты, относящиеся к сущности с идентификатором 91FB. * Event(EventType(позитив)) – все известные позитивные факты. * Event("Обыск проводится в организации", Subject(Org(22201694))) - поиск по названию события с указанием субъекта * Event(EventType(Экономические преступления)) - поиск по типу события |
| Subject | Поиск фактов по субъекту. Для применения внутри функции Event(). | Event(Subject(Name(Id(4AD74)))) |
| EventType | Поиск фактов по типу события. Аргумент – строковое название или идентификатор. Для применения внутри функции Event(). | Event(EventType(преступления и криминал))  Event(EventType(Id(825C93)))  Event(EventType(700EE82)) |

### **Запросы источников**

|  |  |  |
| --- | --- | --- |
| Функция | Действие | Прим. |
| Source | Поиск документов источника по идентификатору. | Metadata(Source(800235)) |
| SourceLevel | Поиск документов с уровня источника(региональный, федеральный…) по идентификатору. | Metadata(SourceLevel(1)) |
| SourceCategory | Поиск документов с категорией источника(газеты, блоги…) по идентификатору. | Metadata(SourceCategory(6)) |
| SourceRegion | Поиск документов с регионом источника по идентификатору. | Metadata(Sourceregion(1125A60)) |
| SourceTopic | Поиск документов с тематикой источника по идентификатору. | Metadata(Sourcetopic(8)) |

### **Остальное**

|  |  |  |
| --- | --- | --- |
| Функция | Действие | Прим. |
| Language | Язык документа.  01 – английский  08 – русский | language(01)  language(08) |
| DocumentType | Тип документа.  Digest = 0,  MarketReport = 1,  PressRelease = 2,  Announcement = 3,  News = 4,  Table = 5,  Unknown = 6,  Interview = 7,  Reportage = 8,  AnalitySource = 9,  Historic = A,  Advertisement = B | DocumentType(0)  DocumentType(4) |