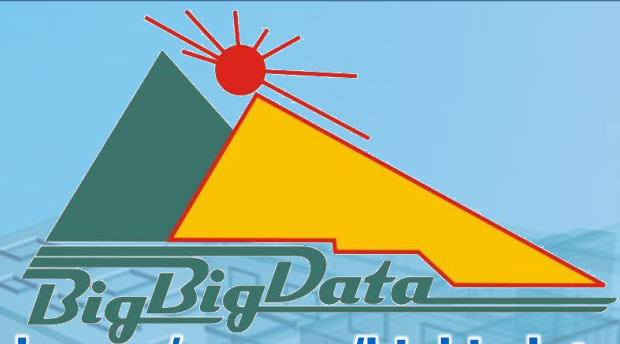


PCWEEK RUSSIAN EDITION

Russian Open Source Summit (ROSS) 2016

5 апреля, 2016



<http://facebook.com/groups/bigbigdata>

Open Source базы данных побеждают проприетарные на примере Multi-Model БД **OrientDB**

Юрий Петров

<https://www.facebook.com/groups/yurypetrov.com.nosql>

Big Data это зонтичный термин и он настолько широк, что бессмысленно придумывать для него короткое определение

Ключевая цель — **обработка разнородных данных любого объема с использованием алгоритмов любой сложности в реальном времени**

Почему пришла Big Data на смену традиционной инфраструктуре

Непреодолимые **ограничения традиционных систем хранения** и обработки данных давно привели к пониманию того, что **задача анализа любого объема разнородных данных** со скоростью близкой к реальному времени **не будет решена никогда**



Большая

революция **Open Source**

Благодаря появлению

прорывных моделей монетизации

был совершен **гигантский**

технологический рывок



Тихая

революция

MULTI-MODEL базы данных



- Document-oriented + Graph-oriented
- Распределенная архитектура
- Multi Master репликация
- Автошардирование между кластерами
- **Apache 2**

Преимущества документо-ориентированных БД над реляционными

1. Отсутствует реляционной поиск
2. Scheme-less
3. Многомерное табличное пространство, которое описывается JSON-схемой
4. Развитые механизмы шардирования, распределения по ключу, распределение таблицы только на указанном кластере "Из коробки"
5. Shared nothing и MPP архитектура
6. Возможность хранения в атрибуте физической ссылки на другую коллекцию атрибутов в формате JSON

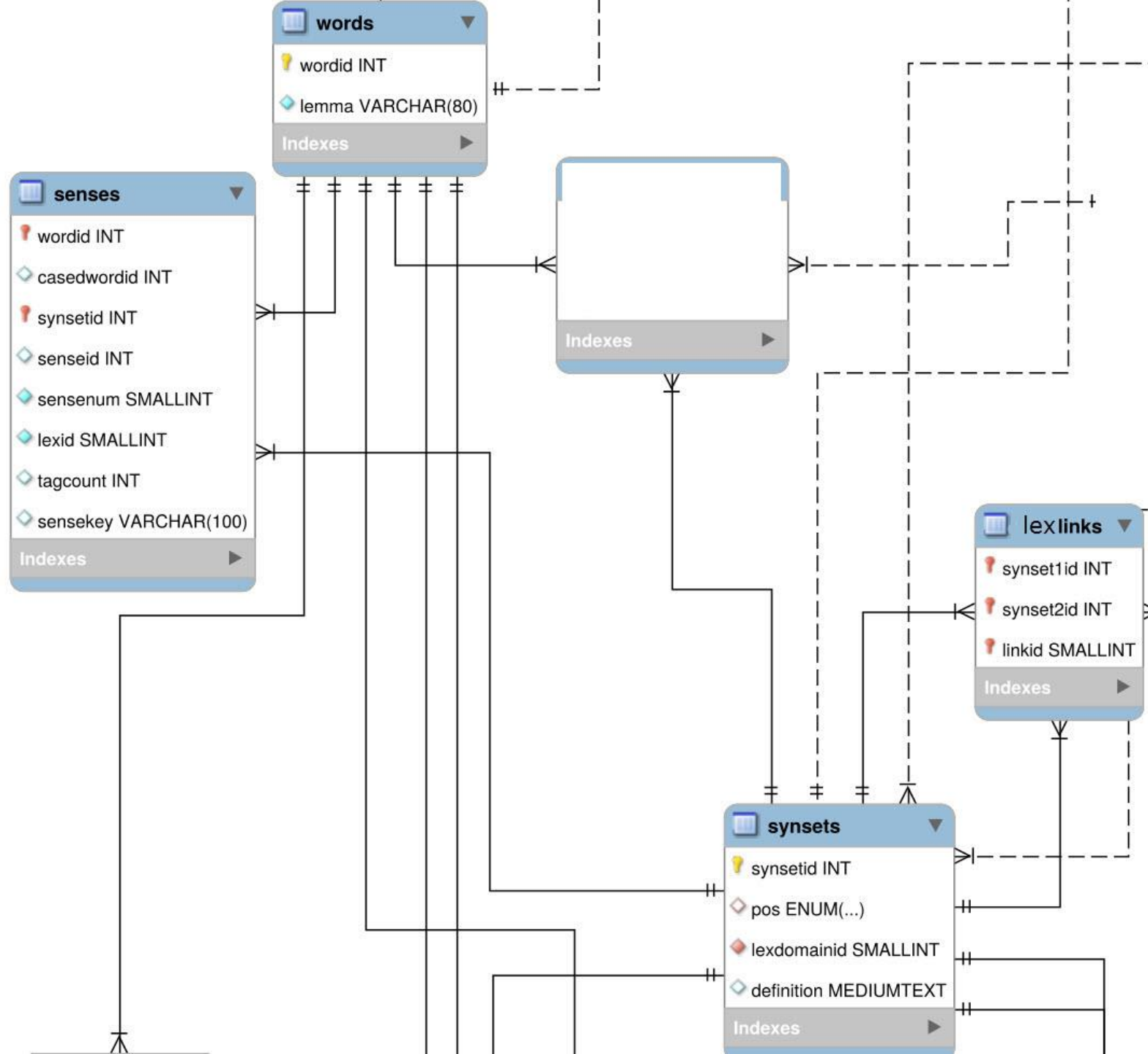
Дополнительные преимущества графовых БД

1. Вершина является JSON-документом
2. Все вершины связаны между собой ребрами
3. Ребра имеют 2 направления и тоже являются JSON-документом
4. Вершины могут иметь неограниченное количество ребер
5. 1 вершина = 1 документ = 1 строка в таблице
6. Вершины и ребра имеют физические адреса, которые однозначно указывают на местоположение данных
7. Легкое построение любых иерархических структур



Электронная тезаурус/семантическая сеть английского языка,
разработанная в **Принстонском университете**
под **некопилефтной свободной лицензией**

Реляционный WORDNET



Реляционный WORDNET: Определяем антоним

```
select distinct
w.wordid
,w.lemma
,sy.synsetid
,w2.wordid
,w2.lemma
,lt.link
,sy.definition
from words w
join senses s1 on s1.wordid = w.wordid
left join synsets sy on sy.synsetid = s1.synsetid
join lexlinks sem on sem.synsetlid = s1.synsetid
join senses s2 on s2.synsetid = sem.synset2id
left join linktypes lt on lt.linkid = sem.linkid
join words w2 on w2.wordid = s2.wordid
WHERE w.lemma = 'cold' and lt.link = 'antonym'
ORDER BY s1.sensenum
```

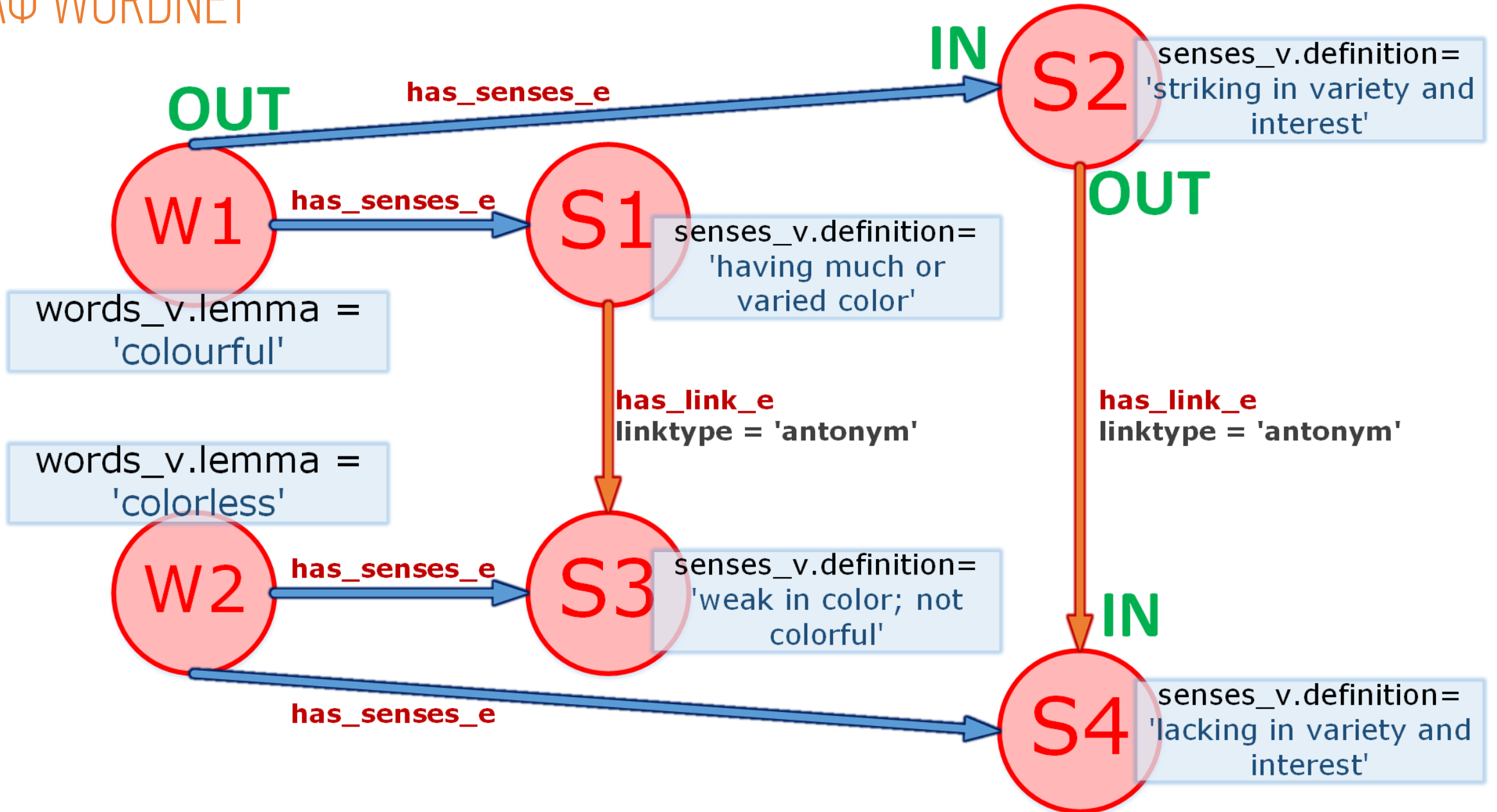
17 | 21 : 0 | INS | PC | [31.03.2016 1:47:44] Сценарий выполнен - Нет ошибок [Время: 1с]

select distinct w.wordid,w...

5 запись(и) [Извлечь (fetch) данные: 2мс] | Сумма:

wordid	lemma	synsetid	wordid	lemma	link	definition
27063	cold	301254201	65472	hot	antonym	having a low or inadequate temperature or feeling a sensation of coldness or havir
27063	cold	105022301	62731	heat	antonym	the absence of heat
27063	cold	105022301	64000	high tempera	antonym	the absence of heat
27063	cold	105022301	65553	hotness	antonym	the absence of heat
27063	cold	301260684	65472	hot	antonym	extended meanings; especially of psychological coldness; without human warmth

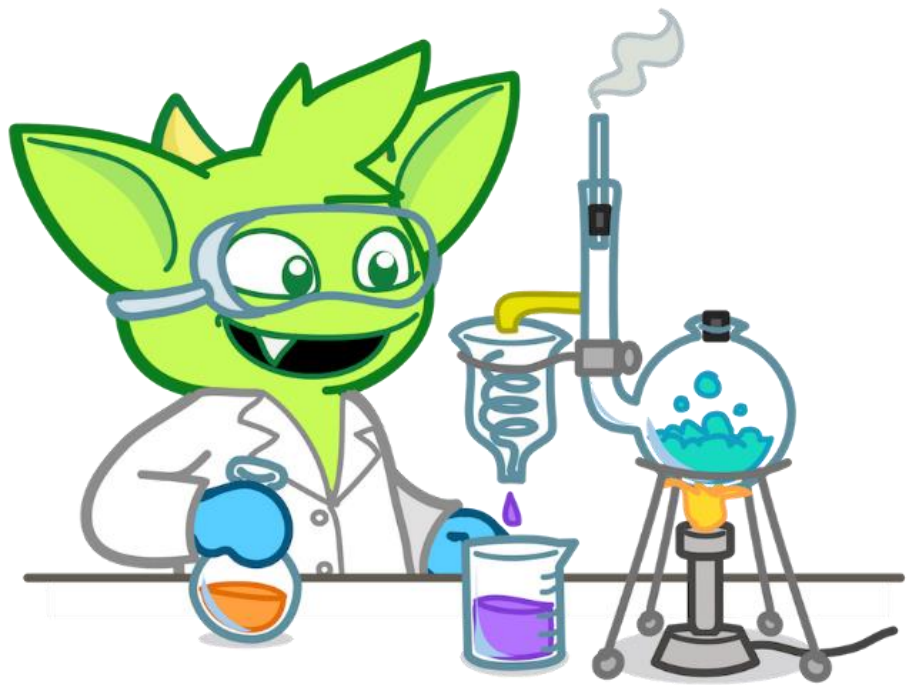
ΓΡΑΦ WORDNET



ГРАФ WORDNET:

Создаем вершины и ребра

```
CREATE CLASS words_v EXTENDS V  
CREATE CLASS senses_v EXTENDS V  
CREATE CLASS has_link_e EXTENDS E  
CREATE CLASS has_senses_e EXTENDS E
```



Язык запросов **GREMLIN**
является стандартом
для **графовых** баз данных

Разрабатывается
в рамках проекта

Apache TinkerPop

<http://tinkerpop.apache.org/>



ГРАФ WORDNET:

Примеры запросов на языке GREMLIN #1

```
1 select out.lemma, in.definition from has_senses_e
2 where out.lemma.length()>3
```

Run: Ctrl + Return | **Undo:** Ctrl/Cmd + Z | **Redo:** Ctrl/Cmd + Shift + Z

Search : Ctrl/Cmd + F | **Toggle Comment:** Ctrl/Cmd + / | **Autocomplete:** Ctrl + Space

PROPERTIES

out	in
ab initio	at the beginning
aba transit number	an identification number consisting of a two-part code assigned to banks and savings associations. The first part identifies the bank's location and the second identifies the bank itself
abaca	a kind of hemp obtained from the abaca plant in the Philippines
abaca	Philippine banana tree having leafstalks that yield Manila hemp used for rope and paper etc

ГРАФ WORDNET:

Примеры запросов на языке GREMLIN #2

```
1 select
2     out.in('has_senses_e').lemma as lemma
3     ,out.inE('has_senses_e').outV('words_v').lemma as same_lemma
4     ,out.definition as definition
5 from has_link_e
6 where linktype = 'antonym'
```

Run: Ctrl + Return | **Undo:** Ctrl/Cmd + Z | **Redo:** Ctrl/Cmd + Shift + Z

Search : Ctrl/Cmd + F | **Toggle Comment:** Ctrl/Cmd + / | **Autocomplete:** Ctrl + Space

PROPERTIES

lemma	same_lemma	definition
["aboveboard"]	["aboveboard"]	in an honest manner
["absolute"]	["absolute"]	perfect or complete or pure
["absorb"]	["absorb"]	suck or take up or in

ГРАФ WORDNET:

Примеры запросов на языке GREMLIN #3

```
1 select
2   out.in('has_senses_e').lemma as left_lemma
3   ,out.definition as left_definition
4   ,in.in('has_senses_e').lemma as right_lemma
5   ,in.definition as right_definition
6 from has_link_e
7 where linktype = 'antonym' and out.in('has_senses_e').lemma = ['colourful']
```

Run: Ctrl + Return | Undo: Ctrl/Cmd + Z | Redo: Ctrl/Cmd + Shift + Z

Search : Ctrl/Cmd + F | Toggle Comment: Ctrl/Cmd + / | Autocomplete: Ctrl + Space



Run

Exp

PROPERTIES

left_lemma	left_definition	right_lemma	right_definition
["colourful"]	having much or varied color	["colorless"]	weak in color; not colorful
["colourful"]	striking in variety and interest	["colorless"]	lacking in variety and interest
["colourful"]	having much or varied color	["colourless"]	weak in color; not colorful

Спасибо за внимание!

Петров Юрий Владимирович

Евангелист, тренер, консультант
(BIG DATA и DWH)

Мои бесплатные материалы:

<http://yurypetrov.com>

Контакты:

WhatsApp, Viber, Telegram: +7-926-5872119

Email: petrov@gobigdata.info

<http://twitter.com/gobigdatainfo>

<http://facebook.com/gobigdata.info>

<https://www.facebook.com/groups/yurypetrov.com.nosql/>

