

django-rdflib and PostgreSQL the best of both worlds

Ştefan Talpalaru *<http://od-eon.com/blogs/stefan/>*

About me

- Python/Django user for the last 3 years
- Linux user from the 20th century
- Gentoo user since first contact
- founder and CTO at Odeon Consulting Group Pte Ltd

Outline

- RDF
- full-text search
- django-fts-odeon
- django-rdflib
- BAMS

RDF

divide et impera

the triple:
subject - predicate - object

the abbreviation for "terminal nerve" in the Swanson 2004 nomenclature for rat is 'tn'

"terminal nerve" - "nomenclature" - "Swanson 2004"

"terminal nerve" - "species" - "rat"

"terminal nerve" - "abbreviation" - "tn"

"terminal nerve" - "abbreviation" - "tn"

http://brancusi1.usc.edu/brain_parts/terminal-nerve-3/

<http://brancusi1.usc.edu/RDF/abbreviation/>

namespaces are syntactic sugar
bams = <http://brancusi1.usc.edu/RDF/>
bams:abbreviation

http://brancusi1.usc.edu/brain_parts/terminal-nerve-3/
<http://brancusi1.usc.edu/RDF/abbreviation/>
"tn"

simple model, simple storage schemes -
triplestores

the quad:
subject - predicate - object - context

labeled graphs
marking subsets of the data

further triple/quad store complications

optimizing frequent operations
(like fetching subjects with a certain rdf:type)

indexing single elements and pairs
(to speed up lookups)

Full-text search

because string matching is not good enough

stemming -> stem

stemming algorithms

PostgreSQL

Snowball version of the Porter algorithm

spelling dictionaries

Ispell, Hunspell

thesaurus - map phrases to words

flexibility in search term ordering

bar foo
foo bar

ranking by relevance

highlighting results

django-fts-odeon

fork of django-fts

focus on the postgresql back-end

better Django integration

- custom manager - the fts methods now survive chaining
- management command to update indices

result highlighting support

```
from django.db import models
import fts

class Literals(fts.SearchableModel):
    lexical = models.TextField()
    search_objects =
fts.SearchManager(fields=('lexical',))
```

```
Literals.search_objects.search('foo bar',  
rank_field='rank')
```

django-rdflib

- pure Python library
- place it in a git repository and have it work on various setups

dependencies:

- Django
- PostgreSQL
- South
- pyparsing for the SPARQL parser
- django-fts-odeon

forked from rdflib

rdflib looked poorly maintained at the time

developers started dropping features
instead of fixing the bugs

first they came for SPARQL (now in a library
called "rdfextras")

then they came for most of the storage back-ends
(MySQL and PostgreSQL included)

why stick with it?

nice pythonic query API

no better alternative

easier to fix/enhance than to rewrite

Django integration

the easy part: reusing the db connection

```
from rdflib.store.MySQL import SQL
from django.db import connection,
transaction
```

```
class PostgreSQL(SQL):
```

```
...
```

```
    def _connect(self, db=None):
        return connection
```

```
    def commit(self):
```

```
transaction.commit_unless_managed()
```

```
    def rollback(self):
```

```
transaction.rollback_unless_managed()
```

the interesting part:

- using a South migration to create the db tables
- creating (unmanaged) models for rdflib's tables

```
from rdflib.term import Literal, URIRef,
BNode, Variable
from django_rdflib.utils import
get_rdflib_store_graph
store, graph = get_rdflib_store_graph()
```

```
triple1 = (URIRef('http://foo.com/subject1'),  
URIRef('http://bar.com/pred1'),  
Literal('obj1'))  
triple2 = (URIRef('http://foo.com/subject2'),  
URIRef('http://bar.com/pred2'),  
Literal('obj2'))  
graph.add(triple1)  
graph.add(triple2)  
graph.commit()
```



```
quad1 = (URIRef('http://foo.com/subject1'),  
None, None, None )  
quad2 = (URIRef('http://foo.com/subject2'),  
None, None, None )  
graph.removeN([quad1, quad2])
```

```
lit_str = """foo  
Bar"""
```

```
q = """  
SELECT ?s ?p WHERE {  
  ?s ?p \"""%s\""" .  
}  
""" % (lit_str, )
```

```
for (s, p) in graph.query(q):  
    pprint((s, p))
```

BAMS - brancusi1.usc.edu

Brain Architecture Management System

Mihail Bota - Associate Professor (Research)
of Biological Sciences, University of
Southern California

collating neurobiological data from scientific publications

PHP/MySQL -> Python/Django/PostgreSQL/RDF

BAMS₂

Brain Architecture Management System

[Overview](#)[FMC](#)[Connectome](#)[Search](#)[Download](#)[Publications](#)[Links](#)[Workspace](#)[Nervous system parts](#)[Thesaurus](#)

Nervous system part:

?

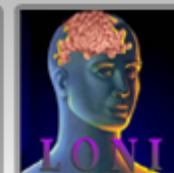
Species:



Nomenclature:



Search in My Workspace:

☐[Policies](#)[BAC](#)[Contact Us](#)[News](#)

BAMS₂

Brain Architecture Management System

Overview **FMC** Connectome **Search** Download Publications Links Workspace

Nervous system parts | Thesaurus

Matches by: **Name (6)** | Description (0) | Abbreviation (0)

Displaying 1 - 6 out of 6 results for: "terminal nerve"

[Do another search](#)

NERVOUS SYSTEM PART (ABBREVIATION)	TYPE	NOMENCLATURE	SPECIES	GENERAL DESCRIPTION	COLLATOR	BAMS ASSOCIATED INFORMATION
Terminal Nerve (tn)	fiber tract	Swanson-1992	Rat	According to Bojsen-Moller 1975, Schwanzel-Fukuda et al. 1985, Demski and Schwanzel-Fukuda 1987.	Mihail Bota	Associated info
Terminal Nerve (tn)	fiber tract	Swanson-1998	Rat	According to Bojsen-Moller 1975, Schwanzel-Fukuda et al. 1985, Demski and Schwanzel-Fukuda 1987.	Mihail Bota	Associated info
Terminal Nerve (tn)	fiber tract	Dong-2007	Mouse	No description provided.	Mihail Bota	Associated info
Terminal Nerve (tn)	fiber tract	Swanson-2004	Rat	According to Bojsen-Moller 1975, Schwanzel-Fukuda et al. 1985, Demski and Schwanzel-Fukuda 1987.	Mihail Bota	Associated info
Terminal Nerve (trn)	fiber tract	Bowden-Macaca-2002	Macaca fascicularis	No description provided. The original URL is provided:	Mihail Bota	Associated info
Terminal Nerve (trn)	fiber tract	Bowden-Human-2002	Human	No description provided. The original URL is provided:	Mihail Bota	Associated info

[Policies](#)

[BAC](#)

[Contact Us](#)

[News](#)



Brain part

Terminal Nerve

Hierarchy level in atlas is 4: 3 superstructures include it.

abbreviation: tn

► Details ?

► Same term found in other nomenclatures ?

▼ Tree of Terminal Nerve ?

☐ Central Nervous System

☐ Central nervous system white matter

☐ Lateral forebrain bundle system

☐ Extrapyrmidal fiber systems

☐ Medial forebrain bundle system

☐ Cranial Nerves, Spinal Nerves, and Related

☐ vagus nerve

glossopharyngeal nerve

☐ Vestibulocochlear nerve

☐ facial nerve

☐ dorsal roots

FMC | Connectome

[Brain Part](#) > [Edit](#)

Edit "Brain Part" object

Fields added by BAMS Administrators

Name: *	<input type="text" value="Cortex Stefan"/>
Abbreviation:	<input type="text"/>
Collation Date:	<input type="text"/>
Collator Argument:	<input type="text" value="0"/>
Collator Involvement:	<input type="text" value="-----Select-----"/>
Description:	<input type="text" value="Test Brain Part for workspace demo"/>
Endorsed:	<input type="checkbox"/>
Extra Info:	<input type="text"/>
Gross Constituent:	<input type="text" value="-----Select-----"/>
Nomenclature:	<input type="text" value="-----Select-----"/>
Old_id:	<input type="text"/>
Reference:	<input type="text" value="-----Select-----"/> + Add New
Species:	<input type="text" value="-----Select-----"/>
Thesaurus:	<input type="text" value="-----Select-----"/> + Add New

Actions

[new "Brain Part" object](#)
[delete "Brain Part" object](#)
[add new field](#)
[predicates list](#)

RDF representation

[Cortex-Stefan](#) - rdf:type - <http://brancusi1.usc.edu/RDF/brainPart>
[Link](#)

[Cortex-Stefan](#) - bams:name - Cortex Stefan
[Link](#)

[Cortex-Stefan](#) - bams:workspace - 1
[Link](#)

[Cortex-Stefan](#) - bams:collatorArgument - 0
[Link](#)

[Cortex-Stefan](#) - bams:description - Test Brain Part for workspace demo
[Link](#)

[Cortex-Stefan](#) - bams:endorsed - False
[Link](#)

Actions

- new "Brain Part" object
- delete "Brain Part" object
- add new field
- predicates list

RDF representation

Cortex-Stefan - rdf:type -
<http://brancusi1.usc.edu/RDF/brainPart>
Link

BAMS₂

Brain Architecture Management System

[Overview](#)[FMC](#)[Connectome](#)[Search](#)[Download](#)[Publications](#)[Links](#)[Workspace](#)[FMC](#)[Connectome](#)[Brain Part](#) > [Cortex-Stefan](#)

Add New Field to Cortex-Stefan

Name: *

Value type: *

Value: *

[+ Add Another](#)[Policies](#)[BAC](#)[Contact Us](#)[News](#)

Foundational Model of Connectivity (Swanson & Bota, 2010)

Overview

FMC

Connectome

Search

Download

Publications

Links

Workspace

Nervous system parts

Thesaurus

Matches by:

Entry (0)

Abbreviation (0)

Reference (0)

Definition (14)

Displaying 1 - 10 out of 14 term results for: "complete"

<< previous

1

2

next >>

Do another search

ENTRY (ABBREVIATION)	REFERENCE	DEFINITION
Connectome	Bota and Swanson, 2010	For the Foundational Model of Connectivity, the complete structural connection matrix of nodes forming the nervous system (Monro, 1783) of ... show more
Neural Network		The structural arrangement of all neurons (Waldeyer, 1891) and their connections in a nervous system (Monro, 1783). A neural network ... show more
Topographic Description Of Body	Swanson & Bota, 2010	One of two common orthogonal ways of describing completely the body; the other is the systems description of body. Topography ... show more
Systems Description Of Body	Swanson & Bota, 2010	One of two common orthogonal ways of describing completely the body; the other is the topographic description of body. An ... show more
Circuitry		A vague term referring to all or part of the connectome or complete wiring diagram; it should be avoided, see ... show more
Neural Subnetwork		An arbitrary subset of a complete neural network, often distinguished on functional grounds. A closely related term is neural subsystem.
Neural Subsystem		A subset of the complete nervous system (Monro, 1783) defined on functional grounds; for example, the visual system or the ... show more

BAMS₂

Brain Architecture Management System

Overview **FMC** Connectome Search Download Publications Links Workspace

Nervous system parts | **Thesaurus**

Matches by: Entry (0) | Abbreviation (0) | Reference (0) | **Definition (1)**

Displaying 1 - 1 out of 1 term result for: "**nucleus lateral**"

[Do another search](#)

ENTRY (ABBREVIATION)	REFERENCE	DEFINITION
Gray Matter Nucleus		<p>A term applied to many gray matter regions in the cerebrospinal axis (Meckel, 1817), usually though not always (e.g., the human dorsal lateral geniculate nucleus) when they have relatively clear borders and are nonlaminated. The first use of the term in this way was by Ludwig (1779, Fig. 2b, p. 36); the cell nucleus, a cytological feature, was named by Brown (1833, p. 710).</p> <p>show less</p>

[Policies](#)

[BAC](#)

[News](#)

[Contact Us](#)

copyright © 2010-2011 The University of Southern California

Page generated in: 4.02s



BAMS₂

Brain Architecture Management System

Overview	FMC	Connectome	Search	Download	Publications	Links	Workspace
--------------------------	------------	----------------------------	------------------------	--------------------------	------------------------------	-----------------------	---------------------------

Principles	Thesaurus	References	Search FMC	Ontology	[administer]
----------------------------	---------------------------	----------------------------	----------------------------	-----------------	------------------------------

RDF/XML	Viewer
-------------------------	------------------------

RDF serialization

[Download it](#)

Documentation

General structure

The XML document is a serialization of RDF triples that describe two types of data: a hierarchical structure expressed using the "OWL DL 2" standard, and additional information about the classes and instances in that hierarchy.

The root tag of the document contains a list of namespace prefixes and their aliases, followed by tags describing unique RDF subjects (in the form of alphanumeric IDs if they are blank nodes, or URIs).

File Edit View Terminal Go Help

bams_ontology_2011-06-... ✕ slides.txt (/src/77_DLD/eur... ✕ mc [stefan@laptop]:/src/7... ✕ mc [stefan@laptop]:~/.0s... ✕ stefan@laptop:~/scripts ✕ stefan@laptop:~/scripts ✕

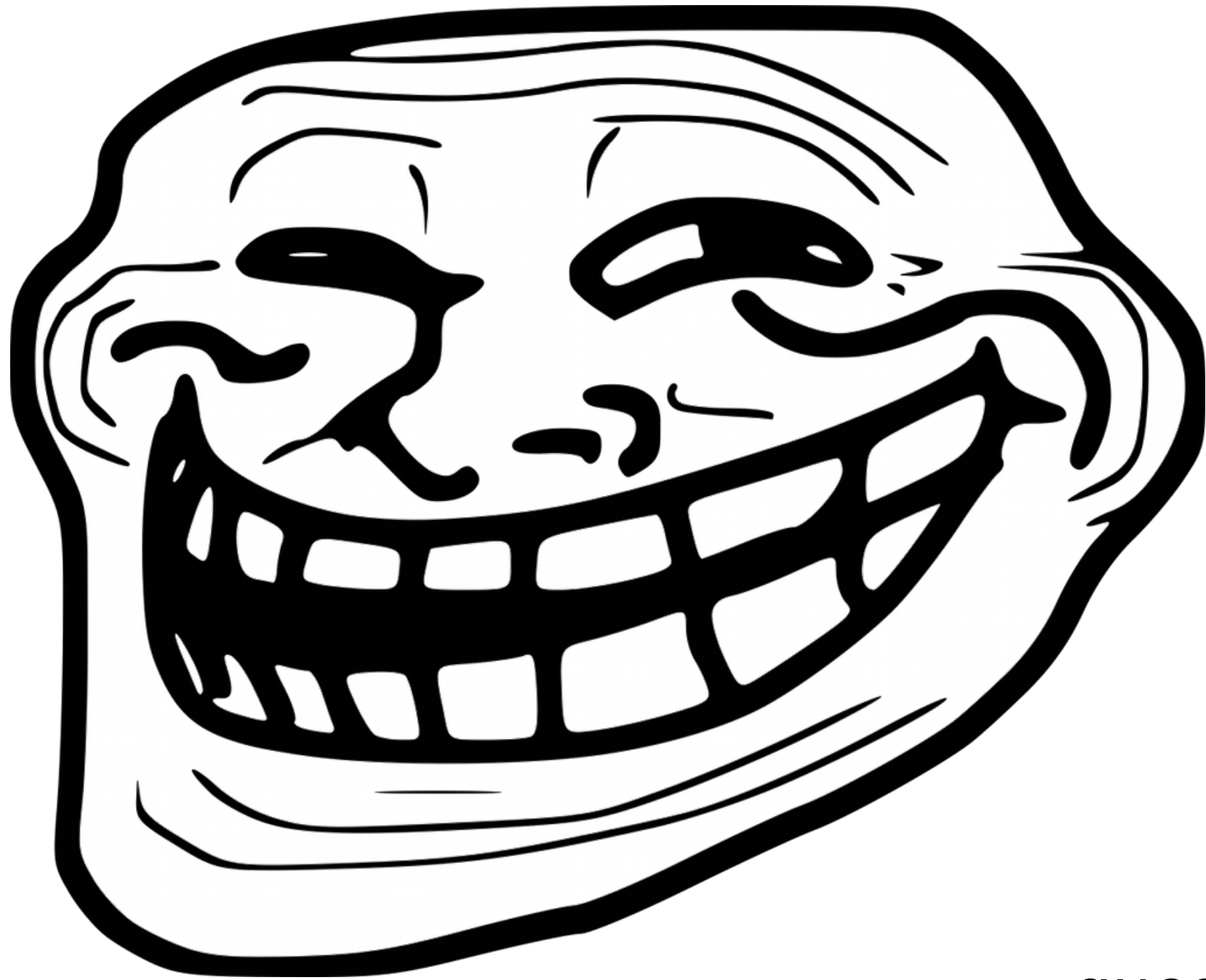
```

</rdf:Description>
<rdf:Description rdf:about="http://brancusi1.usc.edu/brain_parts/terminal-ganglion/">
  <rdf:type rdf:resource="http://brancusi1.usc.edu/RDF/brainPart"/>
  <bams:abbreviation>GTE</bams:abbreviation>
  <bams:collationDate>2004-04-15</bams:collationDate>
  <bams:collator>476</bams:collator>
  <bams:description>No description provided</bams:description>
  <bams:grossConstituent rdf:resource="http://brancusi1.usc.edu/RDF/grayMatter"/>
  <bams:name>terminal ganglion</bams:name>
  <bams:nomenclature rdf:resource="http://brancusi1.usc.edu/rdf/nomenclature/Swanson-2004/">
  <bams:reference rdf:nodeID="auBw0opM362"/>
  <bams:species rdf:resource="http://brancusi1.usc.edu/RDF/rat"/>
  <bams:workspace>0</bams:workspace>
</rdf:Description>
<rdf:Description rdf:about="http://brancusi1.usc.edu/brain_parts/terminal-nerve-3/">
  <rdf:type rdf:resource="http://brancusi1.usc.edu/RDF/brainPart"/>
  <bams:abbreviation>tn</bams:abbreviation>
  <bams:collationDate>2004-11-04</bams:collationDate>
  <bams:collator>476</bams:collator>
  <bams:description>According to Bojsen-Moller 1975, Schwanzel-Fukuda et al. 1985, Demski and Schwanzel-Fukuda 1
987.</bams:description>
  <bams:grossConstituent rdf:resource="http://brancusi1.usc.edu/RDF/fiberTract"/>
  <bams:name>terminal nerve</bams:name>
  <bams:nomenclature rdf:resource="http://brancusi1.usc.edu/rdf/nomenclature/Swanson-2004/">
  <bams:reference rdf:nodeID="auBw0opM362"/>
  <bams:species rdf:resource="http://brancusi1.usc.edu/RDF/rat"/>
  <bams:workspace>0</bams:workspace>
</rdf:Description>
<rdf:Description rdf:about="http://brancusi1.usc.edu/brain_parts/thalamic-peduncles/">

```

/src/77_DLD/bams_ontology_2011-06-20_08-02-50.xml [xml] [line:128922 col:0001] [62%] [lines:206714]

"bams_ontology_2011-06-20_08-02-50.xml" 206714L, 14741010C



questions?

<https://github.com/odeoncg/django-fts-odeon>
<https://github.com/odeoncg/django-rdfliib>