

Problem Solving using Semantic Search with Unstructured Data

Joe Presbrey

CTO, Qwobl

presbrey@qwobl.com

<<http://presbrey.mit.edu/foaf#presbrey>>



Qwobl
Simply Semantic

Example Problem

Find events in Boston related to my interests and which of my friends may also be interested

- Find all the events in Boston for each of my interests
 - Ticketmaster, TicketNetwork, StubHub, etc.
- Find friends with interests that correspond with my own
 - Facebook, copy-paste, Excel?

Example Problem

Find events in Boston related to my interests and which of my friends may also be interested

- Find all the events in Boston for each of my interests
 - Ticketmaster, TicketNetwork, StubHub, etc.
- Find friends with interests that correspond with my own
 - Facebook, copy-paste, Excel?

Linked Data and Semantic Search provide a better solution!

Semantic Search



Find events in Boston related to my interests

- Linked Data



- Select relevant data-sets for the search

Semantic Search



```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix ex: <http://example.com/schema#>.
```

```
<http://presbrey.mit.edu/foaf#presbrey> a foaf:Person ;
  rdfs:label "Joe Presbrey" ;
  foaf:interest
    <http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000c914c7> ,
    <http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000003ad880f> ;
  foaf:knows <http://www.w3.org/People/Berners-Lee/card#i> .
```

```
<http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000c914c7>
  rdfs:label "Taylor Swift" .
<http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000003ad880f>
  rdfs:label "Jay-Z" .
```

```
<http://example.com/event/1> a ex:Event ;
  ex:artist <http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000c914c7> ;
  ex:city <http://geonames.example.com/Boston> .
<http://example.com/event/2> a ex:Event ;
  ex:artist <http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000003ad880f> ;
  ex:city <http://geonames.example.com/Boston> .
```

```
<http://geonames.example.com/Boston> rdfs:label "Boston" ;
```

Semantic Search



Find events in Boston related to my interests

- Linked Data



- Select relevant data-sets for the search

- FOAF, Interests: Freebase+Wikipedia, Events, Geonames

- Query Language+Engine



- Ontological knowledge of how resources relate

- Data-specific representation of search constraints

Semantic Search



Find events in Boston related to my interests

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX ex: <http://example.com/schema#>
SELECT ?event WHERE {
  <http://presbrey.mit.edu/foaf#presbrey> foaf:interest ?i .
  ?event a ex:Event ;
    ?p ?i ;
    ex:city ?location ;
  ?location rdfs:label "Boston" .
}
```

Semantic Search



Find events in Boston related to my interests

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX ex: <http://example.com/schema#>
SELECT ?event WHERE {
  <http://presbrey.mit.edu/foaf#presbrey> foaf:interest ?i .
  ?event a ex:Event ;
    ?p ?i ;
    ex:city ?location ;
    ?location rdfs:label "Boston" .
}
```

```
$ roqet -r simple -D example0.ttl example0.rq
roqet: Querying from file example0.rq
roqet: Query has a variable bindings result
result: [event=uri<http://example.com/event/1>]
result: [event=uri<http://example.com/event/2>]
roqet: Query returned 2 results
```


Live Demo

stub.ly

<http://beta.stub.ly/search?what=presbrey+interests+concerts&where=boston&when=>

Live Demo

stub.ly

[Home](#) | [Sign in](#)

presbrey interests concerts

boston

what: Taylor Swift, Red Sox..

where: New York, 02139..

when: Today, October 2010..

Search

What

presbrey

interests

concerts

Where

boston

When

anytime

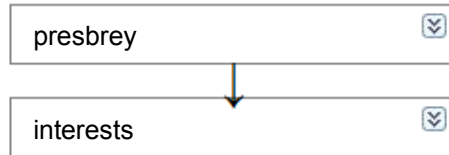
Result 1-1 of 1

Event	Location	Date	Performer
Jay-Z	South End, Boston, MA, Massachusetts	Mar 11 2010	Jay-Z

```
PREFIX qs: <http://rdf.qwobl.com/schema#>
PREFIX es: <http://events.rdf.qwobl.com/schema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT * WHERE {
  ?t0 rdfs:label "presbrey" .
  { { ?t0 ?p0_1 ?t1 } UNION { ?t1 ?p0_1 ?t0 }
    ?t1 qs:class ?t1c . ?t1c rdfs:label "interest" }
  UNION { ?t0 ?p0_1 ?t1 . ?p0_1 rdfs:label "interest" }
  { { ?t1 ?p1_2 ?t2 } UNION { ?t2 ?p1_2 ?t1 }
    ?t2 qs:class ?t1c . ?t1c rdfs:label "concert" }
  UNION { ?t1 ?p1_2 ?t2 . ?p1_2 rdfs:label "concert" }
  ?t2 qs:class ?t2f . ?t2f rdfs:label "event" .
  { ?t2 es:city ?city . ?city rdfs:label "boston" . }
  UNION { ?t2 es:state ?state . ?state rdfs:label "boston" . }
  UNION { ?t2 es:zip ?zip . ?zip rdfs:label "boston" . }
}
```

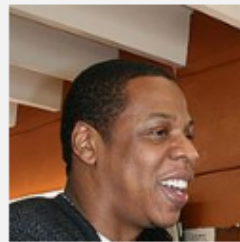
Simple Semantics

Q: presbrey interests Ask



Result 1-4 of 4

[Joe Presbrey](#) Unknown



[Jay-Z](#)

Actor | Artist | Award Nominee | Award Winner | Celebrity | Company Founder | Person | Producer | Writer
Shawn Corey Carter (born December 4, 1969), better known by his stage name, Jay-Z, is an American hip hop artist and businessman. He is the former CEO of Def Jam Recordings and one of the three founde...

[Albums](#) | [Associated Acts](#) | [Concerts](#)

[why](#) [sources](#)



[Taylor Swift](#)

Artist | Award Nominee | Guitarist | Person

Taylor Alison Swift (born December 13, 1989) is an American country pop singer-songwriter, guitarist and actress. In 2006, she released her debut single "Tim McGraw", which peaked at number six on the...

[Albums](#) | [Concerts](#)
[YouTube](#) | [IMDB](#) | [MySpace](#) | [Twitter](#)

[why](#) [sources](#)



[Switchfoot](#)

Artist | Musical Group

Switchfoot is an alternative rock band from San Diego, California. The band's members

Qwobl resolves keywords to URIs, infers intended relations, and executes SPARQL

Simple Semantics

Q: presbrey interests album Ask



Result 1-10 of 45

[Joe Presbrey](#) Unknown

[Jay-Z](#) Person | Celebrity | Artist | Producer | Writer...



[Nas](#)

Album | Composition | Song

Nasir bin Olu Dara Jones (born September 14, 1973), who performs under the mononym Nas ({{IPA-en|u02C8nu0251u02D0z}}), formerly Nasty Nas, is an American rapper and actor. The son of jazz musician Olu ...

[Associated Acts](#)

[why](#) [sources](#)

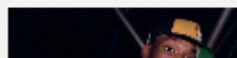


[Infinity On High](#)

Album | Composition | Song

Infinity on High is the fourth studio album by American rock band Fall Out Boy. It was the follow up to their 2005 album From Under the Cork Tree. Originally planned for release at the end of 2006, it...

[why](#) [sources](#)



[Big L](#)

Simple Semantics

Q: seinfeld episode writers Ask

seinfeld episode

↓

writers

Result 1-10 of 239

[The Trip \(Seinfeld episode\)](#) Tv Series Episode



[Larry Charles](#)

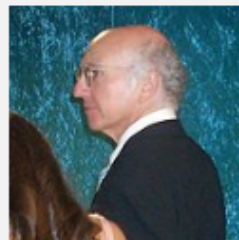
Comedy Writer | Director | Person | Tv Director | Tv Writer | Writer

Larry Charles is an American writer, director, and producer. He is best known as a staff writer for the American sitcom Seinfeld, contributing some of the show's darkest and most absurd storylines. He...

why sources

No information available

[The Pilot \(Seinfeld episode\)](#) Tv Series Episode



[Larry David](#)

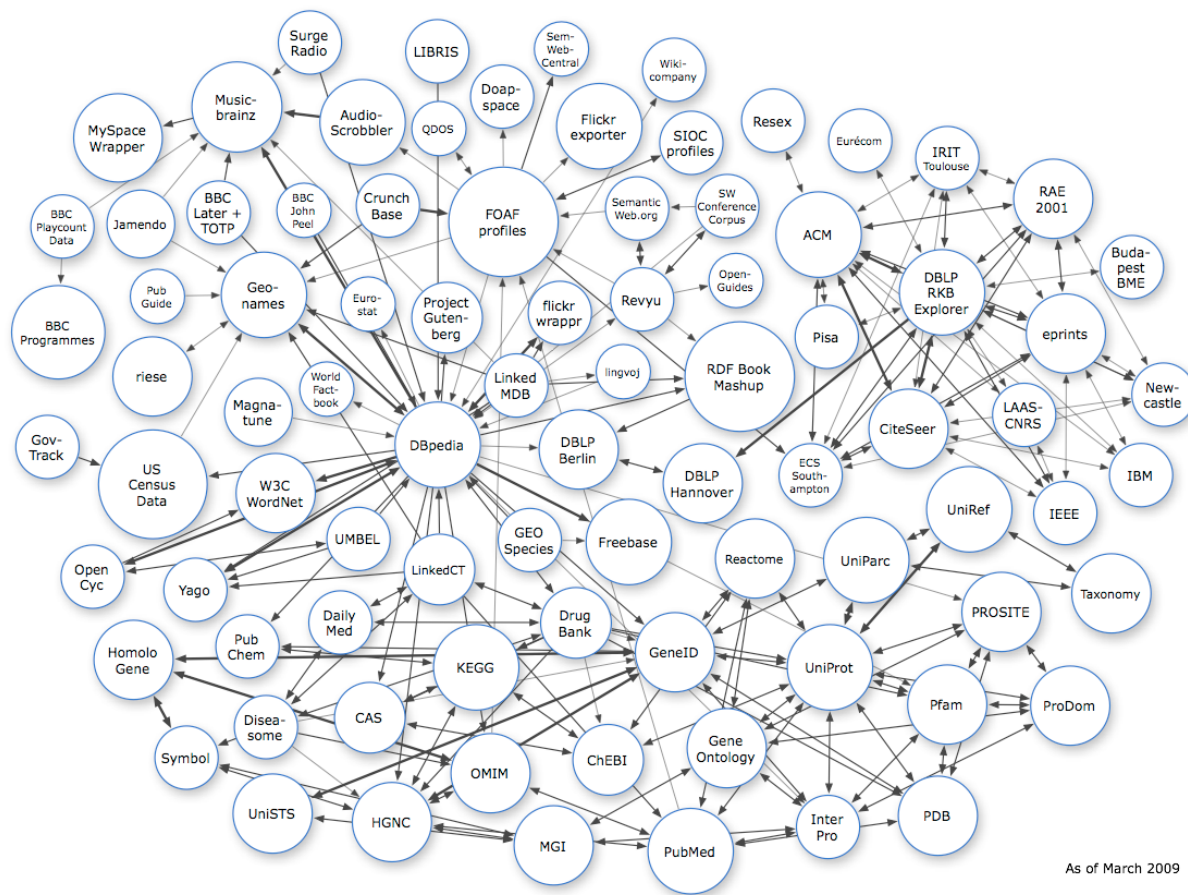
Authoress | Award Nominee | Comedian | Comedy Writer | Composition | Fictional Character Creator | Person | Song | Tv Actor | Tv Producer | Tv Program Creator | Tv Writer

Lawrence Gene "Larry" David (born July 2, 1947) is an American actor, writer, comedian, producer, and film director. David is the co-creator and producer of two successful television comedies, Seinfeld...

why sources

*No RDF,
ontology, or
SPARQL
experience
required*

Linked Data



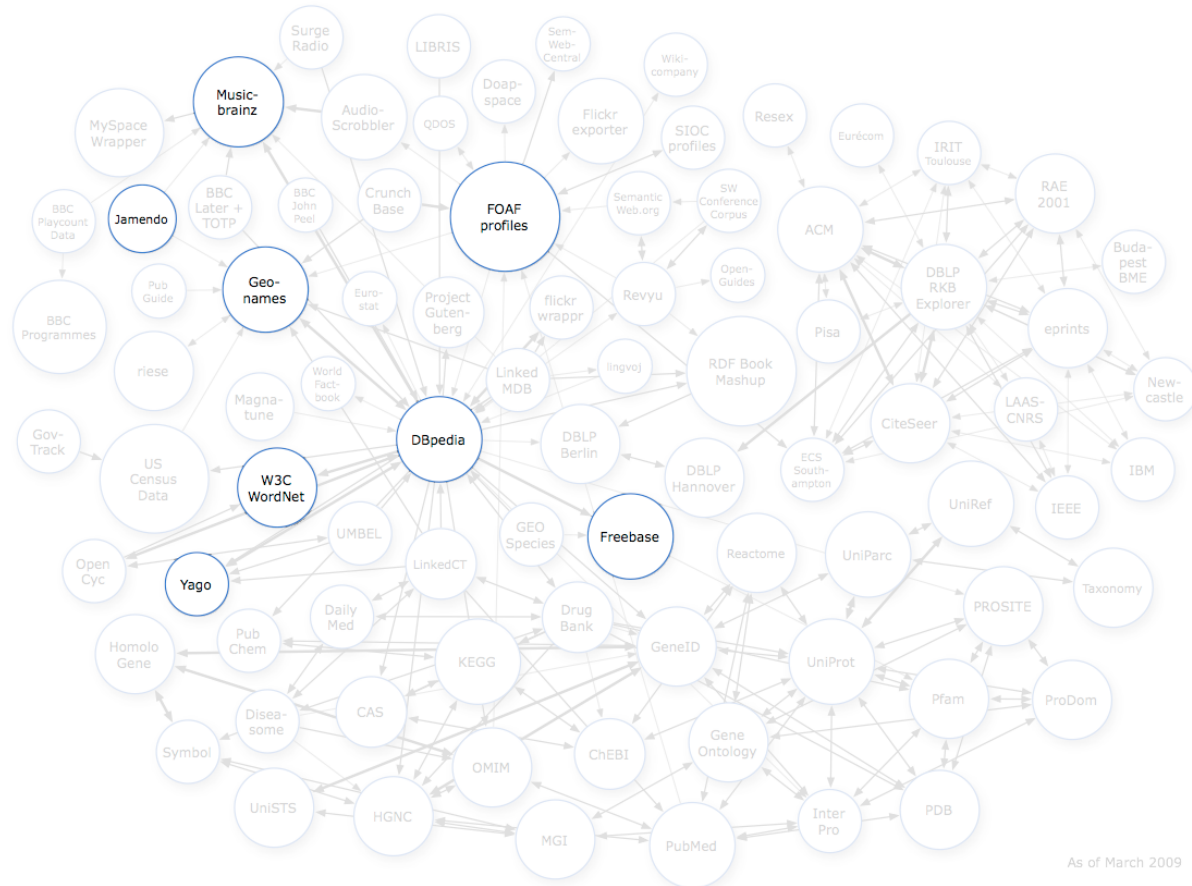
Linked Data

presbrey's interests' concerts in Boston

“interests”:
FOAF profiles

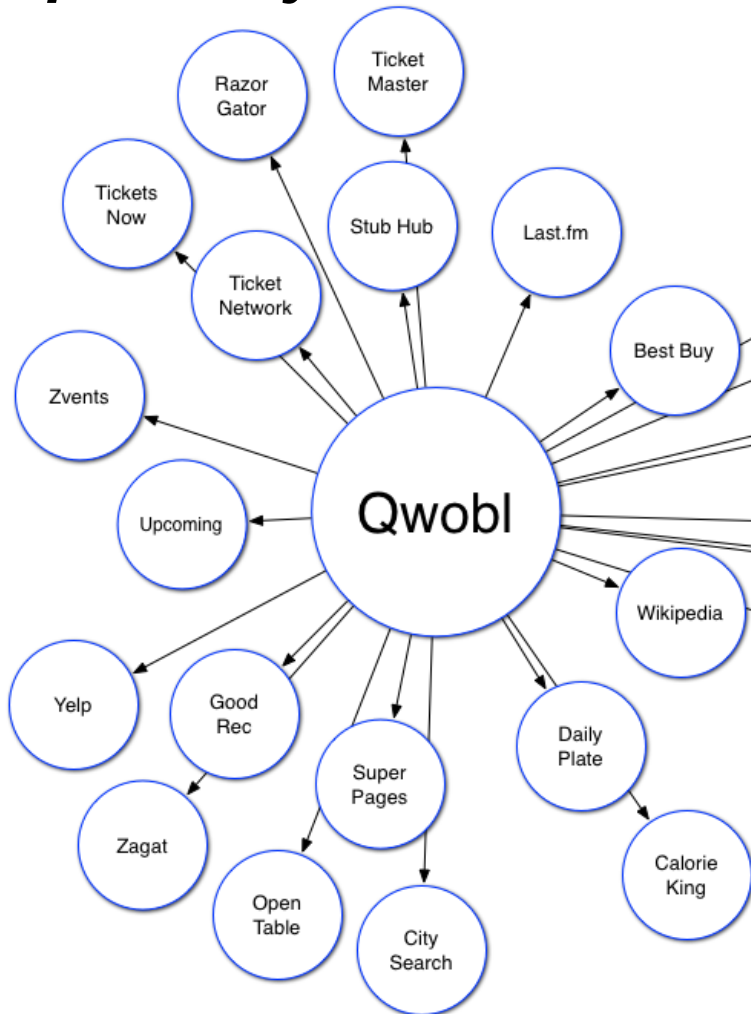
“in Boston”:
Geonames

Lexical inference:
Wordnet, Yago



Qwobl Linked Data

presbrey's interests' concerts in Boston



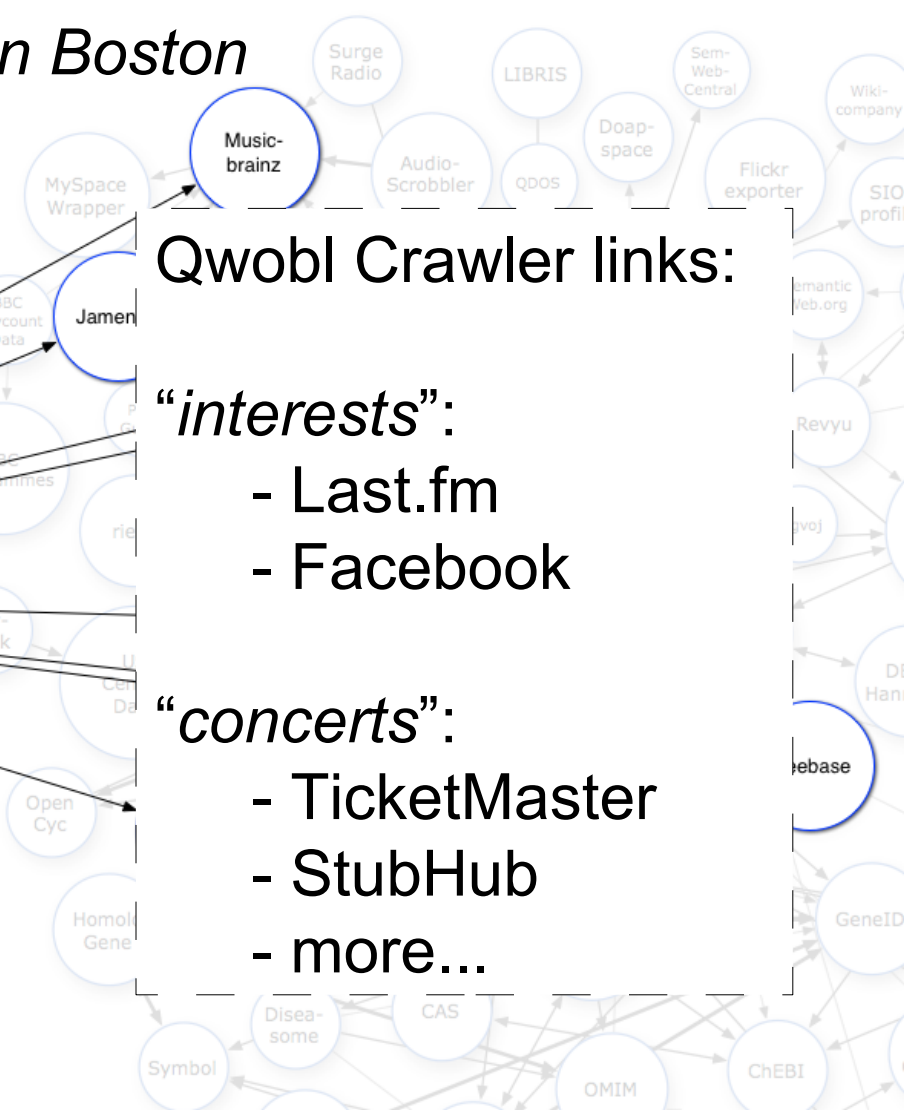
Qwobl Crawler links:

“interests”:

- Last.fm
- Facebook

“concerts”:

- TicketMaster
- StubHub
- more...



Qwobl SPARQL Engine

- Auto-follow owl:sameAs

Equating equivalent entities across data-sets provides for property corroboration and strengthened resource identities but can be significant overhead for the language and/or engine

```
SELECT ?event WHERE {  
  <http://presbrey.mit.edu/foaf#presbrey> foaf:interest ?i .  
  ?event a ex:Event; ?p ?i ; ex:location ?location ; ?location  
  rdfs:label "Boston" . }
```

Qwobl SPARQL Engine

- Auto-follow owl:sameAs

Equating equivalent entities across data-sets provides for property corroboration and strengthened resource identities but can be significant overhead for the language and/or engine

```
SELECT ?event WHERE {  
  <http://presbrey.mit.edu/foaf#presbrey> foaf:interest ?i .  
  ?event a ex:Event; ?p ?i ; ex:location ?location ; ?location  
  rdfs:label "Boston" . }
```

```
SELECT DISTINCT ?event WHERE {  
  { <http://presbrey.mit.edu/foaf#presbrey> foaf:interest ?i }  
  UNION  
  { <http://presbrey.mit.edu/foaf#presbrey> foaf:interest ?i0 .  
    ?i0 owl:sameAs ?i }  
  UNION  
  { <http://presbrey.mit.edu/foaf#presbrey> owl:sameAs ?presbrey .  
    ?presbrey foaf:interest ?i }  
  UNION  
  { <http://presbrey.mit.edu/foaf#presbrey> owl:sameAs ?presbrey .  
    ?presbrey foaf:interest ?i0 . ?i0 owl:sameAs ?i } ... }
```

Qwobl SPARQL Engine

- High-performance full-text rdfs:label's

Replace any rdfs:label variable FILTER(regex) constraints with calls to a full-text index and use matching URIs to prepopulate SPARQL variable bindings

```
SELECT * WHERE {
  ?t0 rdfs:label ?t0l FILTER regex(?t0l, "presbrey")
  ?t0 ?p0_1 ?t1 . ?p0_1 rdfs:label ?p0_1l
  FILTER regex(?p0_1l, "interests")
  { { ?t1 ?p1_2 ?t2 } UNION { ?t2 ?p1_2 ?t1 }
    ?t2 qs:class ?t2c . ?t2c rdfs:label ?t2cl
    FILTER regex(?t2cl, "concerts") }
  UNION { ?t1 ?p1_2 ?t2 . ?p1_2 rdfs:label ?p1_2l
    FILTER regex(?p1_2l, "concerts") }
  { ?t2 es:city ?city . ?city rdfs:label ?cityl
    FILTER regex(?cityl, "boston") }
  UNION { ?t2 es:state ?state . ?state rdfs:label ?statel
    FILTER regex(?statel, "boston") }
  UNION { ?t2 es:zip ?zip . ?zip rdfs:label ?zipl
    FILTER regex(?zipl, "boston") }
}
```

Qwobl SPARQL Engine

- High-performance full-text rdfs:label's

Replace any rdfs:label variable FILTER(regex) constraints with calls to a full-text index and use matching URIs to prepopulate SPARQL variable bindings

```
SELECT * WHERE {
  ?t0 rdfs:label "presbrey" .
  ?t0 ?p0_1 ?t1 . ?p0_1 rdfs:label "interests" .
  { { ?t1 ?p1_2 ?t2 } UNION { ?t2 ?p1_2 ?t1 }
    ?t2 qs:class ?t2c . ?t2c rdfs:label "concerts" }
  UNION { ?t1 ?p1_2 ?t2 . ?p1_2 rdfs:label "concerts" }
  { ?t2 es:city ?city . ?city rdfs:label "boston" }
  UNION { ?t2 es:state ?state . ?state rdfs:label "boston" }
  UNION { ?t2 es:zip ?zip . ?zip rdfs:label "boston" }
}
```