

DjangoProject.com

Full-text Search with PostgreSQL

PAOLO MELCHIORRE



twentytab

Florence 20 April 2018



PyCon
Nove

Paolo Melchiorre



Computer Science Engineer

Python Developer ~ 2006

Django Developer ~ 2011

Senior Developer @ 20Tab

Remote Worker

PostgreSQL user (not DBA)



Outline

- **Basic Search**
- **Full-text search**
- **Django & PostgreSQL**
- **Community**
- **Search in djangoproject.com**

Goal

- **Full-text Search**
- **PostgreSQL**
- **Django**
- **Add/Update search**
- **FLOSS Contribution**

Search Experiences

- **SQL & RDBMS**
- **Plone & ZODB**
- **SQLAlchemy ORM**
- **Django ORM**
- **Psycopg**

Full-text Search Experiences

- **Elastic Search**
- **Apache SOLR**
- **Apache Lucene**
- **Pros**
- **Cons**

Full-text Search

*“... techniques for **Searching**
... computer-stored **Document** ...
... in a **Full-text Database** ...”*

-- [Wikipedia](#)

Document

*“... the **Unit** of searching
in a **Full-text Search System**;
for example, a magazine **Article** ...”*

-- **PostgreSQL**

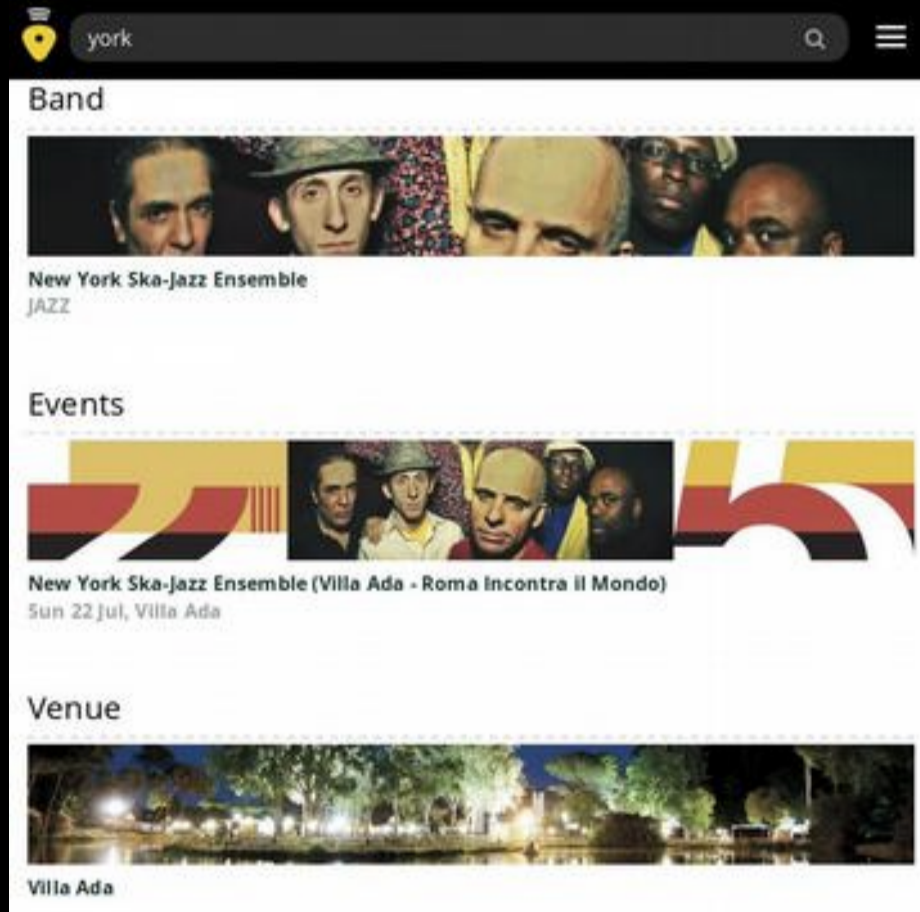
PostgreSQL Full-text Search

- **10-year Support**
- **Dedicated Data Types**
- **Special Indexes**
- **Phrase Search**
- **JSON[b] support**

Full-text Search in Django

- **django.contrib.postgres**
- **Since Django 1.10**
- **BRIN and GIN indexes**
- **Out of the box feature**
- **Quite complete solution**

ConcertiaRoma.com Search



- 15.000 Bands
- 20.000 Events
- 1.000 Venues
- Django 2.0
- Trigram Similarity
- Full-text Search

Full-text Search Talks



- PyCon Otto
- EuroPython 2017
- PGDay.it 2017
- PyRoma Meetup
- paulox.net



twentytab

12



PyCon
Nove

DjangoProject.com Search

- **Elastic-based Search**
- **English-only search**
- **HTML tag results**
- **Complex local environment**
- **Rewrite proposal**

Django Developers Feedback

PROS

- No ext service
- Maintenance
- Setup
- Dogfooding

CONS

- Work
- Features
- Hosting
- Database load

DjangoProject.com Code

- **Django 1.11**
- **PostgreSQL 9.5**
- **Sphinx 1.6**
- **Elasticsearch 5.0**
- **Not self-contained**

EuroPython 2017 Sprints



- Environment test
- Bug reporting
- Elastic removal
- Full-text search
- Working but slow



twentytab

16



PyCon
Nove

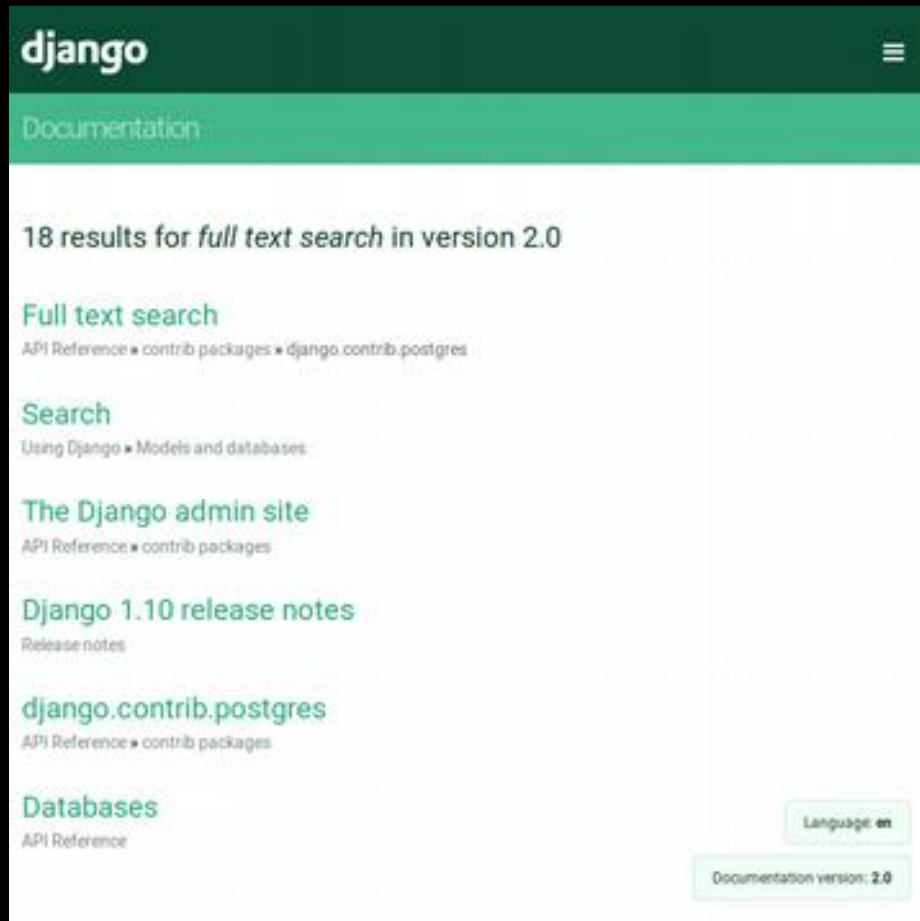
Sprint Code Rewrite

- **Commands restored**
- **Search Vector Field**
- **GIN Index**
- **Trigram similarity**
- **JSON Field**

Pull Request and Deploy

- **Pull Request #809**
- **Tim and Florian review**
- **Tests additions**
- **Migrations squash**
- **Re-index Transaction**
- **Commits clean-up**

DjangoProject.com



- Exact search
- Strip html tag
- Self-contained
- Quite fast
- 7 versions
- 10 languages

Multilingual search

- **15 languages**
- **Default simple**
- **Config field**
- **Query optimizations**
- **Specific tests**

Pull Request & Django 2.0

- **Pull Request #813**
- **Django 2.0 required**
- **Pull Request #794**
- **Blocked by Django-Push**
- **Pull Request #24**

What's next

- **Misspelling support**
- **Search suggestions**
- **Highlighting Results**
- **Searches statistic**
- **PostgreSQL 10**

Conclusions

- **Simplified Infrastructure**
- **Sped up Update**
- **Features preservation**
- **Search improvements**
- **Django/PostgreSQL only**

Full-text Search To-Do list

- 1) PostgreSQL/Django **Versions**
- 2) **Languages** requested
- 3) **Document** to Search Vector
- 4) **Query** to Search Query
- 5) Search Vector **Field & Indexes**

Acknowledgements



twentytab

www.20tab.com



Tim Graham

github.com/timgraham



twentytab

25



PyCon
Nove

Resources

 en.wikipedia.org/wiki/full-text_search

 postgresql.org/docs/current/static/textsearch.html

 docs.djangoproject.com/en/dev/topics/db/search

 stackoverflow.com/tags/full-text-search

 github.com/django

 git.postgresql.org

Thank you

 License (BY SA)

creativecommons.org/licenses/by-sa/4.0/

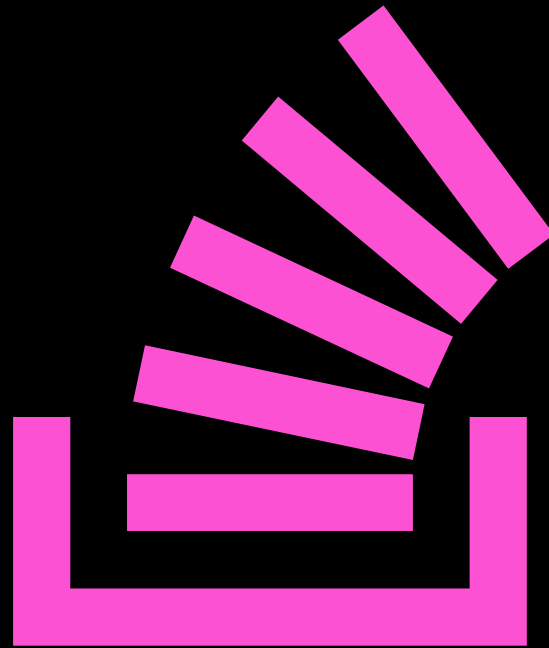
 Source Code

github.com/django/djangoproject.com

 Slides

speakerdeck.com/pauloxnet

Questions



stackoverflow.com/u/755343



twentytab

28



PyCon
Nove

Contacts

 www.paulox.net

 twitter.com/pauloxnet

 [linkedin.com/in/paolomelchiorre](https://www.linkedin.com/in/paolomelchiorre)

 github.com/pauloxnet