Implementando um motor de busca com Django e Elasticsearch

Hello!

Eu sou Sarah Raquel

@sarahraquel

CODE MINER

pyladies

Women Techmakers

GDG Natal

?

O que é um motor de busca ?

Otimização

Textual Matching

Diferentes relevâncias

Características linguísticas

Sinônimos

Introdução ao Elasticsearch

Open source
Baseado no Apache Lucene

Desnormalização?

Performance, Espaço,
Consistência

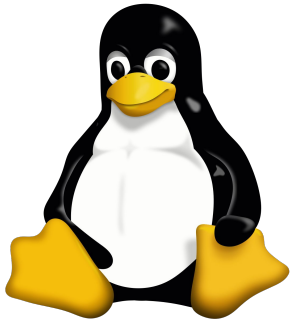
- Write once,
  Read many



Em que pedido meu cliente Paulo comprou um computador Dell X?

elasticsearch

Disponível para:



Dependência:

Download:

https://goo.gl/f9nPhY

http://localhost:9200

TERMINOLOGIA

Index

Operação de adicionar dados

ou

Lugar onde dados são guardados

Sugestões

Podem ser baseadas em termos, em frases ou mesmo sugestões de conclusão da busca

elasticsearch

+

django

Haystack

Modular search for Django

"It features a unified, familiar API that allows you to plug in different search backends (such as **Solr, Elasticsearch, Whoosh, Xapian**, etc.) without having to modify your code"

http://django-haystack.readthedocs.io/en/master/

- **Projeto Exemplo**

- https://github.com/sarahraquel/django-blog

Getting Started with Haystack

```
$ pip install django-haystack
$ pip install elasticsearch
```

- Add "haystack" em INSTALLED_APPS no settings.py

settings.py

```
HAYSTACK_CONNECTIONS = {
    'default': {
        'ENGINE': 'haystack.backends.elasticsearch2_backend.Elasticsearch2SearchEngine',
        'URL': 'http://127.0.0.1:9200/',
        'INDEX_NAME': 'haystack',
    },
}
```

https://goo.gl/iY4Fd1

Lidando com Dados

Branch: **master** ▾ **django-blog** / blog / models / **post.py**

 **reljicd** added __str__() methods to models

1 contributor

18 lines (13 sloc) | 601 Bytes

```
1  from django.contrib.auth.models import User
2  from django.db import models
3  from django.urls import reverse
4
5
6  class Post(models.Model):
7      user = models.ForeignKey(User, on_delete=models.CASCADE)
8      title = models.CharField(max_length=80)
9      body = models.TextField()
10     pub_date = models.DateTimeField('date published', auto_now_add=True)
11
12     def get_absolute_url(self):
13         return reverse('blog:post', kwargs={'pk': self.pk})
14
15     def __str__(self):
16         return '"{title}" by {username}'.format(title=self.title,
17                                                 username=self.user.username)
```

Branch: **master** ▾ **django-blog** / blog / **search_indexes.py**

 **sarahraquel** Add search template

1 contributor

18 lines (14 sloc) | 646 Bytes

```
1  import datetime
2  from haystack import indexes
3  from blog.models.post import Post
4
5
6  class PostIndex(indexes.SearchIndex, indexes.Indexable):
7      text = indexes.CharField(document=True, use_template=True)
8      title = indexes.CharField(model_attr='title')
9      body = indexes.CharField(model_attr='body')
10     author = indexes.CharField(model_attr='user')
11     pub_date = indexes.DateTimeField(model_attr='pub_date')
12
13     def get_model(self):
14         return Post
15
16     def index_queryset(self, using=None):
17         """Used when the entire index for model is updated."""
18         return self.get_model().objects.filter(pub_date__lte=datetime.datetime.now())
```

```
$ python manage.py rebuild_index
```

E quanto a sistemas como Postgres que vem com full-text search, e outras vantagens como transações ACID?

Buscar é mais do que encontrar uma palavra chave em um pedaço de texto: é sobre aplicar conhecimento de domínio específico para implementar bons modelos de relevância, dando um overview de todo o universo do resultado, e fazendo coisas como verificação ortográfica e autocompletion.

Tudo isso sendo *rápido*. 

Thanks!

Perguntas?

Você pode me encontrar em

@sarahraquell

sarahraqueld@gmail.com

Referências

https://www.elastic.co/learn

https://www.elastic.co/blog/found-elasticsearch-as-nosql

https://www.sitepoint.com/search-engine-node-elasticsearch/

http://blog.adnansiddiqi.me/getting-started-with-elasticsearch-in-python/

https://medium.freecodecamp.org/elasticsearch-with-django-the-easy-way-909375bc16cb

... e vários outras que eu não tô lembrada