

Inštalačná príručka

Java, MySQL, ElasticSearch, Tomcat, Webová Služba

Tím č. 4 – aDictIT

Bc. Róbert Horváth
Bc. Peter Jurčík
Bc. Peter Macko
Bc. Vladimír Ruman
Bc. Peter Sládeček
Bc. Maroš Ubreži
Bc. Matúš Vacula

Obsah

| | | |
|---|--|---|
| 1 | Inštalácia Java SE Runtime Enviroment | 3 |
| 2 | Inštalácia ElasticSearch | 3 |
| 3 | Inštalácia MySQL na serveri | 3 |
| 4 | Pridanie kópie databáz na server | 4 |
| 5 | Vytvorenie indexov v ElasticSearch..... | 4 |
| | Použitie CURL API: | 4 |
| | Naplnenie indexu sentences dátami | 6 |
| | Naplnenie indexu synPC dátami | 6 |
| | Naplnenie indexu synthesaurus dátami..... | 7 |
| | Naplnenie indexu shingleindex dátami | 8 |
| 6 | Inštalácia Webového Servera Tomcat | 8 |
| 7 | Inštalácia Servisu pre preklad textu..... | 8 |
| 8 | Konfigurácia prekladača..... | 8 |
| 9 | Inštalácia rozhrania spolupracujúceho s webovou službou..... | 9 |

Nasledujúci manuál obsahuje informácie o tom, ako nainštalovať potrebné súčasti s prvotným naplnením údajov.

1 Inštalácia Java SE Runtime Enviroment

Základným predpokladom pre spustenie nášho riešenia je inštalácia rozhrania Java na svojom PC

1. Stiahnite inštalačný balíček Java SE vo verzii 7 a vyššie z nasledujúcej adresy:
<http://www.oracle.com/technetwork/java/javase/downloads/index.html>
2. Nainštalujte Java SE podľa návodu v balíčku

2 Inštalácia ElasticSearch

1. Stiahnite balíček ElasticSearch z nasledujúcej adresy:
<http://www.elasticsearch.org/download/2012/04/04/0.19.2.html>
2. Postupujte podľa návodu v balíčku

3 Inštalácia MySQL na serveri

Pre správny chod nášho prekladača je potrebné mať na serveri databázu MySQL. Dosiahneme to nasledovne:

1. Na serveri v konzole spustíme príkaz na Obrázok 1, ktorý nám stiahne a nakonfiguruje MySQL verziu 5.1.51.

```
wget http://dev.mysql.com/get/Downloads/MySQL-5.1/mysql5.1.51.tar.gz/from/http://mysql.he.net/  
  
tar xzf mysql-5.1.51.tar.gz  
  
cd mysql-5.1.51  
  
./configure --enable-profiling --prefix=/opt/mysql-5.1.51/ --enable-local-infile \  
--with-mysqld-user=mysql --with-big-tables \  
--with-plugins=partition,blackhole,federated,heap,innodb_plugin --without-docs \  
--with-named-curses-libs=/opt/ncurses/lib/libncurses.a #replace the ncurses location a/c to your system
```

Obrázok 1: Stiahnutie a nakonfigurovanie MySQL v5.1.

2. Následne je potrebné nakonfigurovanú verziu skompilovať a nainštalovať. Tieto kroky vykonáme zadaním príkazu **make**.
3. Ak chceme overiť, či inštalácia prebehla správne, môžeme zadať príkaz **ls -lah /opt**. Ten v tabuľke zobrazí obsah adresára **opt**.
4. Gratulujeme, poradilo sa Vám nainštalovať databázu MySQL.

4 Pridanie kópie databáz na server

Po nainštalovaní MySQL je potrebné naplniť databázy dátami, ktoré vyžaduje prekladač na správne vyhodnocovanie prekladu. Súbor, ktoré je nutné importovať, sa nachádzajú na priloženom disku DVD v zložke **data**. Databázy naplníme nižšie opísaným postupom:

1. najskôr sa prihlásime do MySQL a to tak, že v konzole servera spustíme príkaz

```
mysql -u USER -p PASSWORD
```

kde USER je meno používateľa a PASSWORD je jeho heslo (štandardne USER = root, PASSWORD = žiadny)

2. po prihlásení uvidíme na začiatku riadku **mysql>**, tak ako na Obrázok 2. Zadáme nižšie uvedené SQL príkazy.

```
mysql> CREATE DATABASE sme_clanky  
-> CREATE DATABASE translator_db  
-> CREATE DATABASE translator_db2;
```

Obrázok 2: Príkazy na vytvorenie databáz.

3. následne spustíme príkazy z Obrázok 3

```
mysql> mysqldump sme_clanky < sme_clanky.sql  
-> mysqldump translator_db < translator_db.sql  
-> mysqldump translator_db2 < translator_db2.sql;
```

Obrázok 3: Vytvorenie kópií databáz.

4. Gratulujeme, importovali Ste všetky databázy potrebné pre správny chod prekladača aDictIT.

5 Vytvorenie indexov v ElasticSearch

Použitie CURL API:

```
>curl -XPOST http://localhost:9200/sentences/
```

Pričom “sentences” je názov indexu. Ak index predtým neexistoval, Elastic pošle odpoveď v tvare:

```
{“ok”:true,“acknowledged”:true}
```

V prípade ak index už existuje vráti chybu:

```
{“error”:“IndexAlreadyExistsException[[sentences] Already exists]”,“status”:400}
```

Takýmto spôsobom vytvoríme index so základným nastavením.

Niektoré verzie nášho translatora využívajú zložitejšie nastavenia indexov. Verzia prekladača pracujúca so synonymami potrebuje index s nasledovným nastavením:

```
curl -XPOST http://localhost:9200/synpc/ -d '{
  "index":{
    "analysis":{
      "filter":{
        "synonym":{
          "type":"synonym","synonyms_path":"analysis/synonymspc.txt"}
        }
      }
    "analyzer":{
      "myAnalyzer":{
        "tokenizer":"standard","filter":["synonym"]
      }
    }
  }
}'
```

Pri využití cmd vo windows treba ' zamenit' za " , " zamenit' za ""

Príkaz bude teda vyzerat' nasledovne:

```
>curl -XPOST http://localhost:9200/synpc/ -d '{"index": {"analysis": {"filter": {"synonym": {"type": "synonym", "synonyms_path": "analysis/synonymspc.txt"}}, "analyzer": {"myAnalyzer": {"tokenizer": "standard", "filter": ["synonym"]}}}}' -
```

Pre vytvorenie indexu na prácu so slovníkom thesaurus je nutné nastaviť cestu k slovníku na „analysis/synonymsthesaurus.txt“. Cesta hovorí, že textový súbor je umiestnený v adresári „analysis“, ktorý musí byť umiestnený v adresári „config“.

Zmazanie indexov

Treba použiť príkaz:

```
curl -XDELETE http://localhost:9200/sentences/
```

Ak bola operácia úspešná Elastic odpovie:

```
<"ok":true,"acknowledged":true>
```

Naplnenie indexu sentences dátami

Postup:

1.krok - Z priloženého DVD si stiahnite java projekt MySQLToElasticConverter. V súbore src/Hibernate.cfg.xml zmente nastavenia

```
<property name="hibernate.connection.username">pouzivatelске_meno</property>  
<property name="hibernate.connection.password">heslo</property>
```

podľa nastavení vášho MySQL. Súbor skompiluje do formátu spustiteľného jar súboru s názvom mysqltoelasticconverter.jar .

Nahrajte ho na váš server, kde máte nainštalovanú MySQL databázu s tabuľkou sentences a Elasticsearch s vytvoreným indexom sentences.

2.krok - Súboru priradiť práva na vykonanie príkazom:

```
chmod 777 mysqltoelasticconverter.jar
```

3.krok - Súbor spustíte príkazom(predpokladom je nainštalovaná java na serveri):

```
java -jar mysqltoelasticconverter.jar
```

Naplnenie indexu synPC dátami

Postup:

1.krok -Stiahnite si knižnicu pre prácu s Elasticsearch v jazyku Perl *ElasticSearch - An API for communicating with ElasticSearch*¹ vo verzii 0.52.

Súbory nakopírujte do adresáru :

```
usr/share/perl5
```

2.krok - V ľubovoľnom adresári na serveri vytvorte nový súbor s názvom reindex_data_synpc.pl. Do súboru skopírujte nasledujúci zdrojový kód.

```
#!/usr/bin/perl
```

```
use Elasticsearch();  
my $es = Elasticsearch->new(  
    transport => 'http',  
    servers => '127.0.0.1:9200' -- host:port na ktorom beží Elasticsearch  
);
```

¹ <https://metacpan.org/module/ElasticSearch>

```
my $source = $es->scrolled_search(  
  index => 'sentences',      -- index z ktorého chceme získať dáta  
  search_type => 'scan',  
  scroll => '15m',  
  version => 1  
);  
  
$es->reindex(  
  source => $source,        -- dáta ktoré chceme vložiť do indexu  
  dest_index => 'synpc',    -- názov indexu kam chceme vložiť dáta  
  bulk_size => 5000        -- veľkosť dávky po akej sa budú dáta  
);                          -- vkladať
```

3.krok - Priradíte mu práva na vykonanie príkazom:

```
chmod 777 reindex_data_synpc.pl
```

4.krok - Skript spustíte nasledujúcim príkazom:

```
perl reindex_data_synpc.pl
```

Naplnenie indexu synthesaurus dátami

Postup:

1.krok - V ľubovoľnom adresári na serveri vytvorte nový súbor s názvom reindex_data_synthesaurus.pl. Do súboru skopírujte nasledujúci zdrojový kód.

```
#!/usr/bin/perl  
  
use Elasticsearch();  
my $es = Elasticsearch->new(  
  transport => 'http',  
  servers => '127.0.0.1:9200' -- host:port na ktorom beží Elasticsearch  
);  
  
my $source = $es->scrolled_search(  
  index => 'sentences',      -- index z ktorého chceme získať dáta  
  search_type => 'scan',  
  scroll => '15m',  
  version => 1  
);  
  
$es->reindex(  
  source => $source,        -- dáta ktoré chceme vložiť do indexu  
  dest_index => 'synthesaurus', -- názov indexu kam chceme vložiť dáta  
  bulk_size => 5000        -- veľkosť dávky po akej sa budú dáta  
);                          -- vkladať
```

2.krok - Priradíte mu práva na vykonanie príkazom:

```
chmod 777 reindex_data_synthesaurus.pl
```

3.krok - Skript spustíte nasledujúcim príkazom:

```
perl reindex_data_synthesaurus.pl
```

Naplnenie indexu shingleindex dátami

Pred naplnením vytvorte index ak neexistuje

```
curl -XPOST http://localhost:9200/shingleindex/
```

Postup:

1.krok – Súbor bin/elasticSearchFilling/InsertShingleData.jar, data/ngrams.txt nahrajte z priloženého DVD do ľubovoľného priečinku na serveri. Oba súbory sa musia nachádzať v rovnakom adresári.

2.krok - Súboru priradiť práva na vykonanie príkazom:

```
chmod 777 InsertShingleData.jar
```

3.krok - Súbor spustíte príkazom(predpokladom je nainštalovaná java na serveri a vytvorený index s názvom shingleindex):

```
java -jar InsertShingleData.jar
```

6 Inštalácia Webového Servera Tomcat

1. Stiahnite inštalačný balíček Tomcat 7 a vyššie z nasledujúcej adresy:
<http://tomcat.apache.org/download-70.cgi>
2. Nainštalujte Tomcat 7 podľa návodu v balíčku

7 Inštalácia Servisu pre preklad textu

1. Vypnite webový server Tomcat
2. Prekopírujte WAR súbor aDicTrans.war z adresára bin/service z DVD média do adresára kam sa nainštaloval Tomcat do podadresára webapps
3. Zapnite webový server
4. Do webového prehliadača dajte adresu:
<http://ADRESASERVERA:8080/aDicTranslator/support/versions>
Keď sa vám zobrazí JSON s popisom verzií prekladača inštalácia sa podarila a môžete začať používať prekladač.

8 Konfigurácia prekladača

1. Presuňte sa do adresára:
/webapps/aDicTranslator/translator_VERSION.jar/

2. Otvorte súbor:
config.properties
3. Upravte nastavenia pre Elasticsearch, MySQL a ďalšie potrebné konfiguračné parametre.

9 Inštalácia rozhrania spolupracujúceho s webovou službou

Ak chceme inštalovať rozhranie skopírovať adresár z DVD/bin/webRozhranie do adresára na servery prístupného pre web.

Následne je potrebné upraviť niektoré linky v zdrojovom kóde programu:

1. V súbore index.php je potrebné na riadku 30 zmeniť kód z Obr. 1 na kód na Obr. 2.

```
$source = 'http://127.0.0.1:8080/aDicTranslator/support/versions';
```

Obr. 1 – Pôvodný zdrojový kód

```
$source = 'http://147.175.159.145:8080/aDicTranslator/support/versions';
```

Obr. 2 – Zmenený zdrojový kód

2. Obdobne tak musí byť zmenená aj hodnota premennej v súbore service_connect.php. Kód z Obr. 3 na kód z Obr. 4

```
$url = "http://127.0.0.1:8080/aDicTranslator/".$_POST['version']."/more?  
text=".$_textfield. "&from=".$_POST['from']."&to=".$_POST['to']."";
```

Obr. 3 - Pôvodná hodnota premennej \$url

```
$url = "http:// 147.175.159.145:8080/aDicTranslator/".$_POST['version']."/  
more?text=".$_textfield. "&from=".$_POST['from']."&to=".$_POST['to']."";
```

Obr. 4 - Zmenená hodnota premennej \$url

Ostatné požiadavky:

- Odporúčame používať verziu PHP5 alebo vyššie.
- Pre odosielanie požiadaviek na server je potrebné mať inštalovanú v PHP knižnicu lib-curl.