

# **Lietuvių kalbos teksto sintaksinės-semantinės analizės informacinė sistema**

## **LKSSAIS vystymas**

Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponento administratoriaus instrukcija

## TURINYS

<b>1. Įžanga .....</b>	<b>3</b>
1.1. Dokumento paskirtis.....	3
1.2. Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponento aprašymas .....	3
<b>2. Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponento administravimas .....</b>	<b>4</b>
2.1. Pradiniai reikalavimai .....	4
2.2. Diegimas.....	4
7.1. Komponento veikimo patikrinimas.....	4
7.1.1. MySQL servisas .....	4
7.1.2. Crawler komponentas.....	4
7.2. Komponento serviso atnaujinimas .....	5
7.3. Komponento darbų sąrašo ir statistikos peržiūra .....	5
7.4. Darbo pridėjimas į sąrašą.....	5
7.5. Darbo šablono, laukų pridėjimas, prisijungimo informacijos įvedimas .....	5
7.6. Darbo valdymas .....	5
7.7. Surinktų objektų statistika .....	6
<b>8. Priedai .....</b>	<b>7</b>
<b>9. Dokumento istorija .....</b>	<b>Error! Bookmark not defined.</b>

# 1. Įžanga

## 1.1. Dokumento paskirtis

Šiame dokumente aprašoma:

1. Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponento diegimas į *kubernetes* klasterį.
2. Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponento atnaujinimo procedūra.

## 1.2. Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponento aprašymas

Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponentas susideda iš kelių konteinerizuotų servisų. MySQL duomenų bazės, Selenium karkaso ir paties tekstų gavybos/surinkimo komponento.

Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponentas diegiamas į *kubernetes* klasterį.

## 2. Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponento administravimas

### 2.1. Pradiniai reikalavimai

Administravimui gali būti naudojamas Linux, MacOS arba Windows operacinę sistemą turintis kompiuteris. Kompiuteryje turi būti įdiegtas *kubectl* įrankis.

*Kubectl* įrankis, kompiuteryje iš kurio diegiame, yra sukonfigūruotas, kad pasiektų aplinką į kurią bus diegiamas komponentas.

Komponentas diegiamas į Kubernetes klasterio vardų sritį: tetragrama.

### 2.2. Diegimas

#### 1. Susikuriame laikiną direktoriją

```
mkdir tmp
```

#### 2. Parsisiunčiame kubernetes konfiguracionius failus:

```
git clone https://github.com/tetragrama/lkssais2-k8s.git
```

#### 3. Inicijuojame talpyklas, kuriose bus saugoma informacija:

```
kubectl apply -f crawler-mysql-pv.yaml
```

```
kubectl apply -f crawler-mysql-pvc.yaml
```

#### 4. Paruošiame komponentų konfigūracijai reikalingus duomenis:

```
kubectl apply -f configmaps.yaml
```

#### 5. Paleidžiame MySQL komponentą:

```
kubectl apply -f repository-mysql-deployment.yaml
```

#### 6. Paleidžiame MySQL servisą:

```
kubectl apply -f repository-mysql-svc.yaml
```

#### 7. Paleidžiame crawler komponentą:

```
kubectl apply -f crawler-deployment.yaml
```

### 7.1. Komponento veikimo patikrinimas

#### 7.1.1. MySQL servisas

Paleidžiame *proxy* į servisą:

```
kubectl -n tetragrama port-forward svc/crawler-mysql-svc 8000:3306
```

MySQL klientinėje programoje prisijungiame prie duomenų bazės „kubernetes“ klasteryje:

```
localhost:8000
```

Turime matyti sukurtą duomenų bazę pavadinimu „crawler“.

#### 7.1.2. Crawler komponentas

Paleidžiame *proxy* į komponentą:

```
kubectl -n tetragrama port-forward deployment/crawler-deployment 8000:2001
```

Kitame terminaliniame lange vykdome komandą:

```
curl localhost:8000/status
```

Turime gauti atsakymą su komponento informacija (versija, veikimo laikas).

## **7.2. Komponento serviso atnaujinimas**

1. Padarome pakeitimus faile *crawler-deployment.yaml*.
2. Atnaujiname komponentą paleidę komandą:

```
kubectl apply -f crawler-deployment.yaml
```

## **7.3. Komponento darbų sąrašo ir statistikos peržiūra**

1. Padarome portų nukreipimą iki komponento kubernetes klasteryje:  

```
kubectl port-forward crawler-deployment-c4d464f67-w54sz 2001:2001
```
2. Naršyklėje įvedame adresą localhost:2001
3. Prisijungiame prie sąsajos naudodami „admin“ naudotojo vardą ir „admin“ slaptažodį.
4. Vaizde „Darbai“ galime peržiūrėti įvestus darbus (pav. 1)
5. „Statistika“ vaizdas (pav. 2) rodo kokią dieną kiek įrašų ir kiek nuorodų buvo apdorota

## **7.4. Darbo pridėjimas į sąrašą**

1. Padarome portų nukreipimą iki komponento kubernetes klasteryje:  

```
kubectl port-forward crawler-deployment-c4d464f67-w54sz 2001:2001
```
2. Naršyklėje įvedame adresą localhost:2001
3. Prisijungiame prie sąsajos naudodami „admin“ naudotojo vardą ir „admin“ slaptažodį.
4. Vaizde „Darbai“ paspaudžiame „Pridėti“ mygtuką
5. Suvedame informaciją į darbo aprašymo langą (pav. 3)
6. Išsaugojame darbo aprašymą

## **7.5. Darbo šablono, laukų pridėjimas, prisijungimo informacijos įvedimas**

1. Padarome portų nukreipimą iki komponento kubernetes klasteryje:  

```
kubectl port-forward crawler-deployment-c4d464f67-w54sz 2001:2001
```
2. Naršyklėje įvedame adresą localhost:2001
3. Prisijungiame prie sąsajos naudodami „admin“ naudotojo vardą ir „admin“ slaptažodį.
4. Vaizde „Darbai“ paspaudžiame pelės kairį mygtuką ant darbo esančio sąrašo
5. Atsidariusiame lange (pav 5.) galima pridėti/redaguoti šabloną
6. Šablonas nurodo koks darbo tipas (straipsnis, forumo įrašas, komentaras) (pav. 6)
7. Laukai naudojami dokumentui suformuoti (pav. 7)
8. Norint pridėti naują lauką spausti mygtuką „Pridėti lauką“, lange (pav. 8) suvesti reikiamą informaciją

## **7.6. Darbo valdymas**

1. Padarome portų nukreipimą iki komponento kubernetes klasteryje:  























































```
kubectl port-forward crawler-deployment-c4d464f67-w54sz 2001:2001
```
2. Naršyklėje įvedame adresą localhost:2001

3. Prisijungiame prie sąsajos naudodami „admin“ naudotojo vardą ir „admin“ slaptažodį.
4. Vaizde „Darbai“ prie kiekvieno darbo yra jo valdymo mygtukai (pav. 4)
5. Pirmas mygtukas leidžia pakoreguoti darbo aprašymą
6. Antras mygtukas paleidžia darbo vykdymą
7. Trečias mygtukas sustabdo darbo vykdymą
8. Ketvirtas mygtukas padaro darbą neaktyviu
9. Paskutinis mygtukas pašalina darbą iš darbų sąrašo

### ***7.7. Surinktų objektų statistika***

1. Padarome portų nukreipimą iki komponento kubernetes klasteryje:  
`kubect1 port-forward crawler-deployment-c4d464f67-w54sz 2001:2001`
2. Naršyklėje įvedame adresą localhost:2001
3. Prisijungiame prie sąsajos naudodami „admin“ naudotojo vardą ir „admin“ slaptažodį.
4. Vaizde „Objektai“ (pav. 9) galima peržiūrėti kokie objektai buvo nuskaityti, pamatyti jų statusą

## 8. Priedai

Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponentas			Darbų sąrašas		
			+ Pridėti		
			Pavadinimas	Būsena	Veiksmai
			sa	Laukia	    
			kurstidijuoti	Laukia	    
			I24	Laukia	    
			technologijos	Laukia	    
			elektronika	Laukia	    
			studijos.tv3.lt	Laukia	    
			kurstoti forumas	Laukia	    
			vinted forumas	Laukia	    
			uzdarbis forumas	Laukia	    
			ieskok forumas	Laukia	    
			draugas forumas	Laukia	    

Pav. 1 komponento grafinės sąsajos vaizdas „Darbai“

Automatinės interneto lietuviškų socialinių tekstų gavybos/surinkimo komponentas			Statistika		
			Straipsnių Socialiniai		
			Diena	Irašai	Nuorodos
			2020-08-26	347	465
			2020-08-25	134	2
			2020-08-24	134	3
			2020-08-23	132	0
			2020-08-22	135	3
			2020-08-21	139	7
			2020-08-20	133	1
			2020-08-19	133	2

Pav. 2 komponento grafinės sąsajos vaizdas „Statistika“





+ Pridėti šabloną

Prisijungimo duomenys

☐ Naudoti prisijungimo duomenis

Pavadinimas

technologijos\_document

Veiksmai

Laukai

Redaguoti

Šalinti

Pav. 5 darbo laukų, šablono ir prisijungimo duomenų koregavimo langas

Naujas šablonas

Pavadinimas

Tipas

Straipsnis

▼

Tėvo CSS klasė (jei puslapyje yra daugiau nei vienas elementas)

✓ Saugoti

✗ Atšaukti

Pav. 6 Šablono pridėjimo langas

+ Pridėti lauką

Pavadinimas	CSS parinktuvas	Tipas	Veiksmai
title	.tarpelis_krastuose ...	STRING	
body	.first_psl_table p	STRING	

X Uždaryti

Pav. 7 Laukų sąrašas

Naujas laukas

Pavadinimas

CSS parinktuvas

Tipas

STRING

▼

✓ Saugoti

X Atšaukti

Pav. 8 Lauko pridėjimo langas

Objektai				
Visi ▼				
Data ▼	Šablonas	Nuoroda	Būsena	Tipas
2020-08-31 06:53:28	15-min_article_template	<a href="https://www.15min.lt/naujiena/aktua...">https://www.15min.lt/naujiena/aktua...</a>	Išsiųstas	Straipnis
2020-08-31 06:53:22	15-min_article_template	<a href="https://www.15min.lt/naujiena/aktua...">https://www.15min.lt/naujiena/aktua...</a>	Išsiųstas	Straipnis
2020-08-31 06:53:12	15-min_article_template	<a href="https://www.15min.lt/naujiena/aktua...">https://www.15min.lt/naujiena/aktua...</a>	Išsiųstas	Straipnis
2020-08-31 06:53:06	15-min_article_template	<a href="https://www.15min.lt/naujiena/aktua...">https://www.15min.lt/naujiena/aktua...</a>	Išsiųstas	Straipnis
2020-08-31 06:52:59	15-min_article_template	<a href="https://www.15min.lt/gyvenimas/na...">https://www.15min.lt/gyvenimas/na...</a>	Išsiųstas	Straipnis
2020-08-31 06:52:53	15-min_article_template	<a href="https://www.15min.lt/naujiena/aktua...">https://www.15min.lt/naujiena/aktua...</a>	Išsiųstas	Straipnis
2020-08-31 06:52:47	15-min_article_template	<a href="https://www.15min.lt/naujiena/aktua...">https://www.15min.lt/naujiena/aktua...</a>	Išsiųstas	Straipnis
2020-08-31 06:52:13	15-min_article_template	<a href="https://www.15min.lt/naujiena/aktua...">https://www.15min.lt/naujiena/aktua...</a>	Išsiųstas	Straipnis
2020-08-31 06:51:47	15-min_article_template	<a href="https://www.15min.lt/kultura/naujien...">https://www.15min.lt/kultura/naujien...</a>	Išsiųstas	Straipnis
2020-08-31 06:51:18	15-min_article_template	<a href="https://www.15min.lt/naujiena/aktua...">https://www.15min.lt/naujiena/aktua...</a>	Išsiųstas	Straipnis
2020-08-31 06:50:59	15-min_article_template	<a href="https://www.15min.lt/naujiena/aktua...">https://www.15min.lt/naujiena/aktua...</a>	Išsiųstas	Straipnis
2020-08-31 06:50:37	15-min_article_template	<a href="https://www.15min.lt/naujiena/aktua...">https://www.15min.lt/naujiena/aktua...</a>	Išsiųstas	Straipnis

Pav. 9 Objektų statistikos vaizdas