



VYTAUTO  
DIDŽIOJO  
UNIVERSITETAS  
M C M X X I I

## Lietuvių kalbos teksto segmentavimo komponento administratoriaus instrukcija

2020

## TURINYS

<b>1. Įžanga .....</b>	<b>3</b>
1.1. Dokumento paskirtis.....	3
1.2. Lietuvių kalbos teksto segmentavimo komponento aprašymas .....	3
<b>2. Lietuvių kalbos teksto segmentavimo komponento administravimas .....</b>	<b>4</b>
2.1. Pradiniai reikalavimai .....	4
2.2. Diegimas.....	4
2.3. Komponento veikimo patikrinimas.....	4
2.3.1. Lex servisas .....	4
2.3.2. Lex segmentų atpažinimas.....	4
2.3.3. Lex segmentų atpažinimas su stabiliaisiais junginiais .....	4
2.3.4. Lex sakinių atpažinimas .....	5
2.3.5. Lex paragrafų atpažinimas.....	5
2.3.6. Lex segmentų atpažinimas socialiniame tekste .....	5
2.3.7. Lex stabiliųjų junginių pridėjimas.....	5
2.3.8. Lex režimo nurodymas .....	5
2.3.9. Lex naudojamo modelio nurodymas .....	5
2.4. Komponento atnaujinimas.....	6
2.5. Komponento statistikos peržiūra, taisyklių įkėlimas .....	6
<b>3. Dokumento istorija .....</b>	<b>Error! Bookmark not defined.</b>

# 1. Įžanga

## 1.1. Dokumento paskirtis

Šiame dokumente aprašoma:

1. Lietuvių kalbos teksto segmentavimo komponento diegimas į *kubernetes* klasterį.
2. Lietuvių kalbos teksto segmentavimo komponento atnaujinimo procedūra.

## 1.2. Lietuvių kalbos teksto segmentavimo komponento aprašymas

Lietuvių kalbos teksto segmentavimo komponentas susideda iš vieno konteinerizuoto serviso.

Lietuvių kalbos teksto segmentavimo komponentas diegiamas į *kubernetes* klasterį.

Komponentas diegiamas į asmeninį kompiuterį, parsisiuntus iš [www.semantika.lt](http://www.semantika.lt) (šiuo atveju *kubernetes* nereikia, komponentas veikia per vidinį web servisą, reikia vadovautis žemiau aprašyta užklausų formavimo sintakse)

## 2. Lietuvių kalbos teksto segmentavimo 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.

Jei diegiama asmeniniame kompiuteryje, turi būti įdiegtas Docker.

### 2.2. Diegimas

1. Susikuriame laikiną direktoriją

```
mkdir tmp
```

2. Parsisiunčiame konfigūracinius kubernetes failus:

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

3. Paleidžiame komponentą:

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

4. Paleidžiame komponento servisą, kad klasteryje jis būtų prieinamas kitiems komponentams:

```
kubectl apply -f lex-service.yaml
```

5. Jei diegiama asmeniniame kompiuteryje, vadovautis [www.semantika.lt](http://www.semantika.lt) aprašyta parsisiuntimo instrukcija.

### 2.3. Komponento veikimo patikrinimas

#### 2.3.1. Lex servisas

Paleidžiame *proxy* į servisą:

```
kubectl -n tetragrama port-forward service/lex-svc 8080:7000
```

Kitame terminale vykdome komandą:

```
curl http://localhost:8080/status
```

Turime gauti atsakymą iš serviso su pranešimu apie veikiančią versiją, veikimo laiką.

#### 2.3.2. Lex segmentų atpažinimas

```
curl --request POST \  
  --url http://localhost:8080/ \  
  --header 'content-type: application/json' \  
  --data 'Meras norėtu, kad "žmonės galėtu vaikščioti iš vienos vietos" į kita, judėtu... Dėl pirmosios šio teiginio dalies ginčytis netenka.'
```

#### 2.3.3. Lex segmentų atpažinimas su stabiliaisiais junginiais

```
curl --request POST \  
  --url http://localhost:8080/
```

```
--header 'compounds: 1' \
--header 'content-type: application/json' \
--data 'Meras norėtu, kad „žmonės galėtų vaikščioti iš vienos vietos į kita, judėtu... Dėl pirmosios nu ir šio teiginio dalies ginčytis netenka. O bet kas jis toks?'
```

### **2.3.4. Lex sakinių atpažinimas**

```
curl --request POST \
--url http://localhost:8080/ \
--header 'content-type: application/json' \
--data 'Netgi leidžia mums turēt vaikų', - lyg metafora apie homoseksualių asmenų padėti Lietuvoje skambėjo R. Kirilkino daina. Kaip tai baisu. „Soriukas“ - taip jis vadinsis Lietuvoje ir garantuos, kad nieko panašaus iki šiol nesate matę.'
```

### **2.3.5. Lex paragrafų atpažinimas**

```
curl --request POST \
--url http://localhost:8080/ \
--header 'content-type: application/json' \
--data 'Netgi leidžia mums turēt vaikų', - lyg metafora apie homoseksualių asmenų padėti Lietuvoje skambėjo R. Kirilkino daina. Kaip tai baisu. „Soriukas“ - taip jis vadinsis Lietuvoje ir garantuos, kad nieko panašaus iki šiol nesate matę.'
```

Aldas Kikutis, Lietuvos smulkiojo ir vidutinio verslo tarybos pirmininkas, portalui IQ.lt pripažino, kad MMA nebeatspindi dabartinio pragyvenimo lygio, tačiau verslui reikia laiko suplanuoti tokiems pokyčiams.'

### **2.3.6. Lex segmentų atpažinimas socialiniame tekste**

```
curl --request POST \
--url http://localhost:8080/ \
--header 'content-type: application/json' \
--data 'Visu pirma ten nera jokio pesciuju tako, o visu antra kam nepasitaiko. Taip kad autorius geriau eik miegot.'
```

### **2.3.7. Lex stabiliųjų junginių pridėjimas**

```
curl --request POST \
--url http://localhost:8080/compounds \
--header 'content-type: application/json' \
--data '{
  "compounds":["ir šio"]
}'
```

### **2.3.8. Lex režimo nurodymas**

```
curl --request POST \
--url http://localhost:8080/ \
--header 'content-type: application/json' \
--header 'format: line' \
--data 'Meras norėtu, kad „žmonės galėtų vaikščioti iš vienos vietos į kita, judėtu... Dėl pirmosios nu ir šio teiginio dalies ginčytis netenka. O bet kas jis toks?'
```

Pagal nutylėjimą rezultatas grąžinamas JSON formatu. Įvedus header parametrą format: line rezultatas grąžinamas po eilutę.

### **2.3.9. Lex naudojamo modelio nurodymas**

```
curl --request POST \
```

```
--url http://localhost:8080/ \
--header 'content-type: application/json' \
--header 'mode: common' \
--data 'Meras norėtu, kad „žmonės galėtų vaikščioti iš vienos vietos į kitą, judėtu... Dėl pirmosios nu ir šio teiginio dalies ginčytis netenka. O bet kas jis toks?'
```

Modeliui nurodyti naudojamas įkeltų taisyklių identifikatorius, pvz: law, common, medicine.

## ***2.4. Komponento atnaujinimas***

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

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

## ***2.5. Komponento statistikos peržiūra, taisyklių įkėlimas***

1. Padarome portų nukreipimą iki komponento kubernetes klasteryje:  
`kubectl port-forward repository-deployment-b5c56d64b-fpdtl 8080:8080`
2. Atsidarome naršyklę adresu: <http://localhost:8080>
3. Atsidariusiame lange (pav. 1) matome komponento statistiką, taip pat po ja galima įkelti naują modelį, reikia nurodyti identifikatorių, pavadinimą ir įkelti sakinių, bei segmentų modelius.

# LEX v1.0.0

## Statistika

- Darbo pradžia: 2020-08-28T14:37:22.723113
- Užklausų kiekis: 0
- Klaidos: 0
- Aktyvūs modeliai: 4
- Numatytasis sakinių modelis: common
- Numatytasis segmentų modelis: common

## Taisyklių rinkiniai

- Teisiniai dokumentai (law)
- Bendrinė kalba (common)
- Socialiniai tekstai (social)
- Medicininiai tekstai (medicine)

## Papildyti taisykles

Identifikatorius:

Pavadinimas:

Sakinių šablonas:

No file chosen

Segmentų šablonas:

No file chosen

## Klaidos (0)

LKSSAIS 2020