

Lietuvių kalbos teksto sintaksinės-semantinės analizės informacinė sistema

LKSSAIS vystymas

Įvardintų esybių atpažinimo lietuvių kalbos tekste komponento
administratoriaus instrukcija

TURINYS

1. Įžanga	3
1.1. <i>Dokumento paskirtis.....</i>	<i>3</i>
1.2. <i>Įvardintų esybių atpažinimo lietuvių kalbos tekste komponento aprašymas.....</i>	<i>3</i>
2. Įvardintų esybių atpažinimo lietuvių kalbos tekste komponento administravimas	4
2.1. <i>Pradiniai reikalavimai</i>	<i>4</i>
2.2. <i>Diegimas.....</i>	<i>4</i>
4.1. <i>Komponento veikimo patikrinimas.....</i>	<i>4</i>
2.2.1. <i>Ner servisas.....</i>	<i>4</i>
2.2.2. <i>Ner asmenų vardų atpažinimas.....</i>	<i>4</i>
2.2.3. <i>Ner geografinių pavadinimų atpažinimas.....</i>	<i>4</i>
2.2.4. <i>Ner datų ir laiko atpažinimas.....</i>	<i>5</i>
2.2.5. <i>Ner organizacijų atpažinimas</i>	<i>5</i>
2.2.6. <i>Ner produktų pavadinimų atpažinimas.....</i>	<i>5</i>
2.2.7. <i>Ner valiutų išraiškų atpažinimas.....</i>	<i>5</i>
2.2.8. <i>Ner studijų krypčių atpažinimas</i>	<i>5</i>
2.2.9. <i>Ner įvardintų esybių atpažinimas teisiniuose dokumentuose.....</i>	<i>5</i>
2.2.10. <i>Ner naujos esybės atpažinimo taisyklės pridėjimas</i>	<i>6</i>
4.2. <i>Komponento atnaujinimas.....</i>	<i>6</i>
4.3. <i>Papildomų esybių atpažinimo taisyklių pridėjimas.....</i>	<i>6</i>
4.4. <i>Ontologijos užkrovimas rankiniu būdu</i>	<i>6</i>
3. Priedai	7
4. Dokumento istorija.....	Error! Bookmark not defined.

1. Įžanga

1.1. *Dokumento paskirtis*

Šiame dokumente aprašoma:

1. Įvardintų esybių atpažinimo lietuvių kalbos tekste komponento diegimas į *kubernetes* klasterį.
2. Įvardintų esybių atpažinimo lietuvių kalbos tekste komponento atnaujinimo procedūra.
3. Papildomų sričių atpažinimui pridėjimo procedūra.
4. Ontologijos užkrovimo procedūra.

1.2. *Įvardintų esybių atpažinimo lietuvių kalbos tekste komponento aprašymas*

Įvardintų esybių atpažinimo lietuvių kalbos tekste komponentas susideda iš vieno konteinerizuoto serviso.

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

2. Įvardintų esybių atpažinimo lietuvių kalbos tekste 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 konfiguracionius kubernetes failus:

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

3. Paleidžiame komponentą:

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

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

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

4.1. Komponento veikimo patikrinimas

2.2.1. Ner servisas

Paleidžiame *proxy* į servisą:

```
kubectl -n tetragrama port-forward service/ner-svc 8080:7002
```

Kitame terminale vykdome komandą:

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

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

2.2.2. Ner asmenų vardų atpažinimas

```
curl --request POST \  
  --url http://localhost:8080/ner \  
  --header 'content-type: application/json' \  
  --data '{"body":"Jonas su Obama ir kitais atstovais nuėjo namo iš Vilniaus į Kauną ir nusipirko IBM kompiuterį rukyt. Prezidentas Obama ir Jonas Jonaitis nuėjo ir susitiko Vilniuje su savo klientais ir nusipirko porą IBM kompiuterių ir Grybauskaitė nusipirko dar kažką.", "annotations":{"lex":{"seg":[[0,5],[6,2],[9,5],[15,2],[18,6],[25,9],[35,5],[41,4],[46,2],[49,8],[58,1],[60,5],[66,2],[69,9],[79,3],[83,10],[94,5],[99,1],[101,11],[113,5],[119,2],[122,5],[128,8],[137,5],[143,2],[146,8],[155,8],[164,2],[167,4],[172,9],[182,2],[185,9],[195,4],[200,3],[204,11],[216,2],[219,12],[232,9],[242,3],[246,5],[251,1]],"s":[[0,100],[101,151]],"p":[[0,252]]}}}'
```

2.2.3. Ner geografinių pavadinimų atpažinimas

```
curl --request POST \  
  --url http://localhost:8080/ner \  
  --header 'content-type: application/json' \  
  --data '{"body":"Kaunas ir Vilnius jau kurį laiką lenktyniauja tarpusavyje. Šiauliai šiuo metu  
džiaugiasi savo dideliu
```

```
uostu.", "annotations": {"lex": {"seg": [[0, 6], [7, 2], [10, 7], [18, 3], [22, 4], [27, 5], [33, 12], [46, 11], [57, 1], [59, 8], [68, 4], [73, 4], [78, 10], [89, 4], [94, 7], [102, 5], [107, 1]], "s": [[0, 58], [59, 49]], "p": [[0, 108]]}}}'
```

2.2.4. Ner datų ir laiko atpažinimas

```
curl --request POST \
  --url http://localhost:8080/ner \
  --header 'content-type: application/json' \
  --data '{"body": "Panoramą per TV rodo 23:00 valandą Gegužės 12 d. ir sausi . visomis dienomis ir kiaura para.", "annotations": {"lex": {"seg": [[0, 8], [9, 3], [13, 2], [16, 4], [21, 5], [27, 7], [34, 1]], "s": [[0, 35]], "p": [[0, 35]]}}}'
```

2.2.5. Ner organizacijų atpažinimas

```
curl --request POST \
  --url http://localhost:8080/ner \
  --header 'content-type: application/json' \
  --data '{"body": "Daugiau kaip 57 proc. apklausoje dalyvavusių darbdavių nurodė, kad KTU absolventų žinios ir gebėjimai labai atitinka jų įmonių poreikius. VDU susijungė su Aleksandro Stulginskio Universitetu.", "annotations": {"lex": {"seg": [[0, 7], [8, 4], [13, 2], [16, 5], [22, 10], [33, 11], [45, 9], [55, 6], [61, 1], [63, 3], [67, 3], [71, 10], [82, 6], [89, 2], [92, 9], [102, 5], [108, 8], [117, 2], [120, 6], [127, 9], [136, 1], [138, 3], [142, 9], [152, 2], [155, 10], [166, 11], [178, 12], [190, 1]], "s": [[0, 137], [138, 53]], "p": [[0, 191]]}}}'
```

2.2.6. Ner produktų pavadinimų atpažinimas

```
curl --request POST \
  --url http://localhost:8080/ner \
  --header 'content-type: application/json' \
  --data '{"body": "Facebook apsirūpino dell kompiuteriais.", "annotations": {"lex": {"seg": [[0, 6], [7, 11], [19, 8], [28, 8], [37, 6], [44, 7], [52, 4], [57, 8], [65, 1], [67, 4], [72, 8], [81, 15], [96, 1]], "s": [[0, 97]], "p": [[0, 97]]}}}'
```

2.2.7. Ner valiutų išraiškų atpažinimas

```
curl --request POST \
  --url http://localhost:8080/ner \
  --header 'content-type: application/json' \
  --data '{"body": "Mėnesinė alga - $ 4000.00. Sulaikyta lėšų alkoholio kontrabandos, kurios vertė juodojoje rinkoje viršija 500000 eurų.", "annotations": {"lex": {"seg": [[0, 8], [9, 3], [14, 1], [16, 4], [21, 4], [25, 1], [27, 9], [37, 8], [46, 12], [58, 1], [60, 6], [67, 5], [73, 9], [83, 7], [91, 7], [99, 6], [106, 4], [110, 1]], "s": [[0, 26], [27, 84]], "p": [[0, 111]]}}}'
```

2.2.8. Ner studijų kryptių atpažinimas

```
curl --request POST \
  --url http://localhost:8080/ner \
  --header 'content-type: application/json' \
  --data '{"body": "Informacinių technologijų studijų kryptis užima 100-150 vietą tarp pasaulio universitetų - jokia kita kito universiteto kryptis nepasiekė tokio aukšto įvertinimo.", "annotations": {"lex": {"seg": [[0, 12], [13, 8], [22, 6], [29, 9], [39, 4], [44, 3], [48, 5], [54, 8], [63, 1], [65, 10], [76, 9], [86, 10], [96, 1]], "s": [[0, 97]], "p": [[0, 97]]}}}'
```

2.2.9. Ner įvardintų esybių atpažinimas teisiniuose dokumentuose

```
curl --request POST \
  --url http://localhost:8080/ner \
  --header 'content-type: application/json' \
  --data '{"body": "Lietuvos respublikos vyriausybės dėl gamtos išteklių nutarimas nr 2. Mano asmens kodas yra 38601011488. Tamsės g. 7, Kaunas. Gegužės 15 d. bus naujas produktas išleistas. +370 654 05103 yra telefono numeris. Du šimtai trys litai. AB SEB bankas. Nacionalinės žemės
```

tarnybos dėl sklypų išvada nr 1XX-15 yra geras. Mano sklypas yra 15 arų. Buto plotas 50 m2. UAB 'ATEA'. Rašykite man el. paštu į simonas@tetragrama.com. Mano sugalvotas prietaisas. MB eridas LT100011119616. Įmonės kodas: 304624480. Lietuvos respublikos civilinio kodekso 3.135 straipsnio. Huawei P10 jis yra tasy topping mano sugalvotas prietaisas yra gearas. Pizza is gerai ypac trumas su zuviene CheeseTopping. naujausia laida. Or deffinitely or deffinately is it?", "annotations": {"lex": {"seg": [[0, 6], [7, 11], [19, 8], [28, 8], [37, 6], [44, 7], [52, 4], [57, 8], [65, 1], [67, 4], [72, 8], [81, 15], [96, 1]], "s": [[0, 97]], "p": [[0, 97]]}}}'

2.2.10. Ner naujos esybės atpažinimo taisyklės pridėjimas

```
curl --request POST \
  --url http://localhost:8080/ner/extend \
  --header 'content-type: application/json' \
  --data '{
    "label": "invention",
    "patterns": [{"LOWER": "mano"}, {"LOWER": "sugalvotas"}, {"LOWER": "prietaisas"}]
  }'
```

4.2. Komponento atnaujinimas

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

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

4.3. Papildomų esybių atpažinimo taisyklių pridėjimas

1. Padarome porto nukreipimą iki komponento kubernetes klasteryje:
`kctl port-forward ner-deployment-85696646b9-bgdkz 5000:8080`
2. Atsidarome naršyklę adresu <http://localhost:8080>
3. Taisyklių pridėjimo skirtuke (pav. 1) įrašome esybės tipą
4. Įrašome aprašymo taisykles (turi būti JSON formatas)
5. Išsaugojame

4.4 Ontologijos užkrovimas rankiniu būdu

1. Padarome porto nukreipimą iki komponento kubernetes klasteryje:
`kctl port-forward ner-deployment-85696646b9-bgdkz 5000:8080`
2. Atsidarome naršyklę ir nueiname adresu <http://localhost:8080>
3. Ontologijos įkėlimo skirtuke (pav. 2) įrašome esybės tipą, kuris bus kuriamas iš ontologijoje esančių esybių
4. Pasirenkame ontologijos failą
5. Išsaugojame

3. Priedai

Įvardintų esybių atpažinimo komponento valdymas

[Taisyklių pridėjimas](#)[Ontologijos įkėlimas](#)

Esybės tipas

Esybės tipas

Esybės aprašymo taisyklės (JSON)

Esybės aprašymo taisyklės (JSON)

Saugoti

Atšaukti

Pav. 1 Taisyklių pridėjimo skirtukas komponento grafinėje sąsajoje

Įvardintų esybių atpažinimo komponento valdymas

[Taisyklių pridėjimas](#)[Ontologijos įkėlimas](#)

Esybės tipas

Esybės tipas

Pasirinkite failą...

Įkelti

Pav. 2 Ontologijos įkėlimo skirtukas komponento grafinėje sąsajoje