

Raplamaa Rakenduslik Kolledž

**NewsParser**

**Kasutusjuhend**

Marten Filippov, Mykyta Minaiev

IT25

Juhendaja: Marko Livalent

Kehtna 2025

# SISUKORD

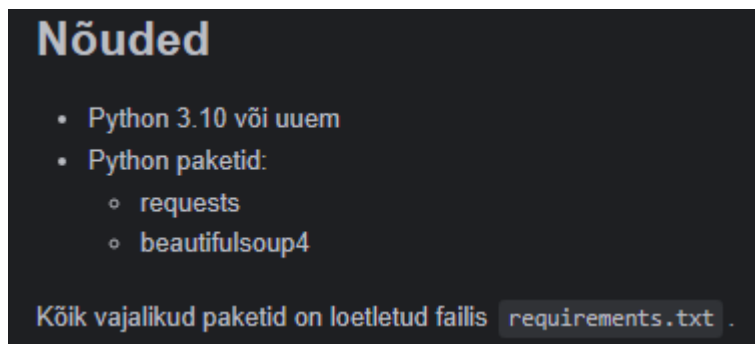
<b>Tutvustus.....</b>	<b>3</b>
Versioon.....	3
<b>Laadimine.....</b>	<b>4</b>
Käivitamine.....	5
<b>Näited.....</b>	<b>6</b>
„Ühendus ei õnnestu“ .....	7
„Tulemus tühi“ .....	7

# Tutvustus

See projekt on Pythonis kirjutatud rakendus, mille eesmärk on laadida Eesti uudisteportaalide avalehed ja tuvastada pealkirjadest inimeste nimed. Programm töötab portaalidega postimees.ee, delfi.ee ja ohtuleht.ee. Leitud nimed kogutakse kokku ning salvestatakse iga portaali kohta eraldi JSON-faili. Rakendust saab kasutada kas käsurea kaudu või graafilise kasutajaliidese abil, sõltuvalt projekti teostusest.

## Versioon

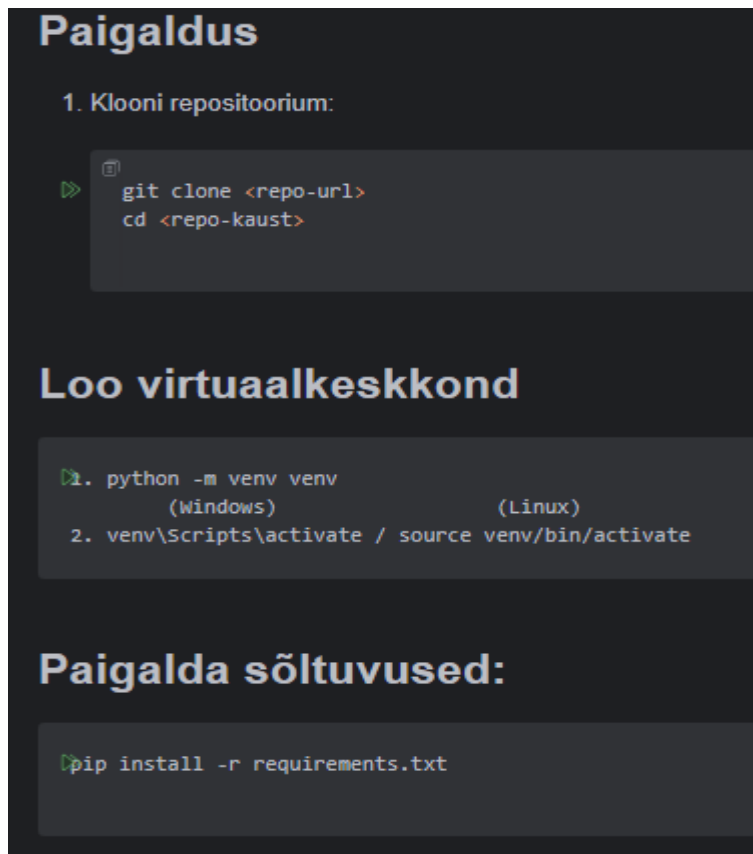
Programmi kasutamiseks on vajalik Python versioon 3.10 või uuem (v.t pilt 1). Rakendus vajab toimivat internetiühendust, kuna uudisteportaalide sisu laaditakse veebist. Programmi tööks kasutatakse Python pakette requests ja beautifulsoup4, mis võimaldavad HTML-lehtede allalaadimist ja parsimist. Kõik vajalikud paketid on kirjas failis requirements.txt.



Pilt 1. Versiooni nõuded.

# Laadimine

Projekt tuleb kõigepealt kloonida (v.t pilt 2) GitHubist või alla laadida. Seejärel tuleb luua virtuaalkeskkond (v.t pilt 2) ning see aktiveerida. Pärast virtuaalkeskkonna aktiveerimist paigaldatakse vajalikud paketid (v.t pilt 2) requirements.txt faili alusel. Pärast edukat paigaldust on projekt valmis käivitamiseks.



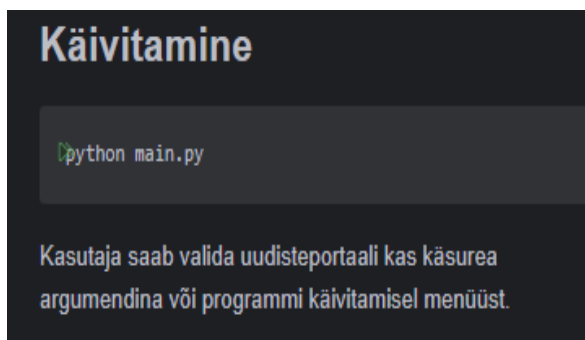
Pilt 2. Laadimise sammud ehk etapid.

## Käivitamine

Programmi käivitamine (v.t pilt 4) toimub käsurea kaudu käsuga `python main.py`. Pärast käivitamist palutakse kasutajal valida uudisteportaal (v.t pilt 3).

```
-----  
      --- Nimede Otsija ---  
-----  
1. Delfi  
2. Postimees  
3. Õhtuleht  
4. Kõik  
0. Välju  
-----  
Vali:
```

Pilt 3. GUI'st ja kus mis on.



Pilt 4. Kuidas käivitada programmi.

## Näited

Kasutaja sisendiks on valitud uudisteportaal, näiteks Postimees. Programmi töö tulemusena luuakse JSON-fail, mis sisaldab kõiki pealkirjadest tuvastatud inimeste nimesid (v.t pilt 5, 6). Iga uudisteportaali kohta luuakse eraldi fail, näiteks postimees.json või delfi.json. Failis on kirjas domeeni nimi ning nimekiri nimedest, mis vastavad etteantud tingimustele.

```
1. Delfi
2. Postimees
3. Õhtuleht
4. Kõik
0. Välju

-----

Vali: 4
Delfi skaneeritud. Leiti 21 nime.
Postimees skaneeritud. Leiti 26 nime.
Õhtuleht skaneeritud. Leiti 139 nime.
```

Pilt. 5 Näide kus näitab mis juhtub peale uudiseportaali valimist.

```
{
  "name": "Rainer Saks",
  "headline": "SÕJARAPORT | Rainer Saks: Venezuela kurdab oma liitlase Venemaa täielikult puudunud toetuse üle USA armee rünnaku ajal",
  "url": "https://arvamus.delfi.ee/artikkel/120429596/sojaraport-rainer-saks-venezuela-kurdab-oma-liitlase-venemaa-taielikult-puudunu",
},
{
  "name": "Kaarel Kasper Kõrge",
  "headline": "Kokaiini tarvitanud suusataja Kaarel Kasper Kõrge positiivne dopinguproov tuli ilmsiks aasta ja neli kuud hiljem 35",
  "url": "https://www.delfi.ee"
},
```

Pilt. 6. Näitab mis läheb json faili. (delfi.json)

# Levinud vead ja lahendused

## „Ühendus ei õnnestu“

See viga tekib siis, kui rakendus ei suuda luua ühendust serveri (v.t pilt 7) või välise teenusega. Põhjuseks võib olla puuduv või ebastabiilne internetiühendus, vale serveri aadress või võrgu piirangud. Probleemi lahendamiseks kontrolli oma internetiühendust, veendu serveri aadressi õigsuses ning proovi rakendus uuesti käivitada.

```
-----  
1. Delfi  
2. Postimees  
3. Õhtuleht  
4. Kõik  
0. Välju  
-----  
Vali: 3  
Viga ohtuleht skaneerimisel: HTTPSConnectionPool  
Ohtuleht skaneeritud. Leiti 0 nime.
```

Pilt. 7 Viga kus skript ei saa ühendust ning ei suuda skaneerita uudiseportaali.

## „Tulemus tühi“

See teade tähendab, et rakendus käivitub korrektselt, kuid ei tagasta ühtegi tulemust. Tavaliselt on põhjuseks vigased, puudulikud või sobimatud sisendandmed. Kontrolli, kas sisendfail sisaldab andmeid ja on õiges formaadis. Vajadusel testi rakendust näidisandmetega.