



EESTI  
RAHVUS-  
RAAMATUKOGU

29.11.2020

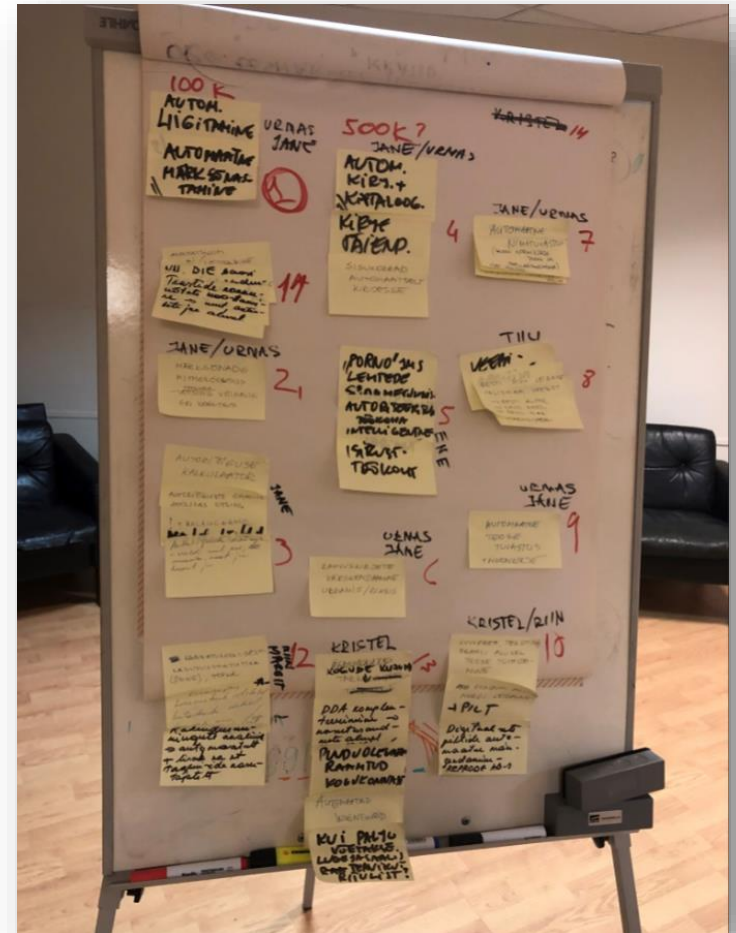
# RAHVUSRAAMATUKOGU KRATIMAANIA

Jane Makke

Raamatukoguteenuste arenduskeskus  
Eesti Rahvusraamatukogu

# Meie tee algas...

- 4. detsembril 2018  
Kutsega osaleda *Deep Dive* projektis
- Ideekorje / ajurünnak
  - ~20 ideed
  - sõelale jäi 13 ideed
- Fookus:
  - mida teeme täna käsitsi
  - mida ei tee üldse
- Hindamine ja ümarlaud



# Potentsiaalsed ideed

- Eestiaineliste veebisaitide leidmine välisveebist
- Automaatne märksõnastamine
- Automaatne kataloogimine
- Autorikaitse tähtaja kalkulaator
- Luulerea, viisijupi, fraasi alusel teose tuvastamine
- DIGARis olevate piltide automaatne märgendamine
- Automaatne nimetuvastus/teose tuvastus täistekstidest ja sidumine autoriteetfailiga
- Päringukratt – vestlusrobot
- Rahvuskirjete automaatne uuendamine URRAMis ja RIKSis
- Nutikas kogude kujundamine – targad ostuotsused
- Kogude haldamine, inventuurid jm analüütika



# Kaks kratti

- **Automaatne märksõnastamine**
- **Päringukratt – vestlusrobot**
  
- **Rahataotlus**
- **Hange**
  - detailanalüüs
  - prototüüp (*Proof of Concept*)



# Automaatne märksõnastamine

- **Probleem:**

- märksõnastamine on subjektiivne
- märksõnastamine ei ole alati ühetaoline
- märksõnad ei ole alati suupärased
- märksõnastamine on aeganõudev
- märksõnastamine on kallis



- **Eesmärk:**

- Kvaliteedi kasv, suurem kasu kasutajale
- Integreerida robot igapäevasesse töövoogu
- Palju rakendusvaldkondi

# Automaatne märksõnastamine

- **Oodatav tulemus:**
  - Väljaannete analüüs
  - Robot ennustab märksõnad:  
teemad, kohanimed, ajaperiood, isikud, organisatsioonid
  - Töötaja valideerib ja teeb valiku
  - Märksõnad saadetakse kataloogi bibliokirjesse
  
- **Potentsiaalne kasutusala**
  - Dokumentide analüüs ja märgendamine
  - Artiklite analüüs ja märgendamine
  - Veebiarhiivis veebilehtede sisuanalüüs



# Automaatne märksõnastamine

## Alusandmed, analüüs ja õpetamine:

- Eesti rahvusbibliograafia ERB: ~200 000 raamatu andmed
- RR digitaalarhiiv DIGAR: ~10 000 täisteksti (1940-2019)
- Eesti märksõnastik EMS: 40 000 märksõna, 20 000 äraviidet

## Tulem:

- EMSist on kasutuses ca 8000 märksõna



## Eeldus:

- iga ennustatava märgendi jaoks on vaja vähemalt 50 näidisdokumenti ehk märksõna kasutuskorda.

# Automaatne märksõnastamine

Andmeanalüüs: teemamärksõnad (8 000)

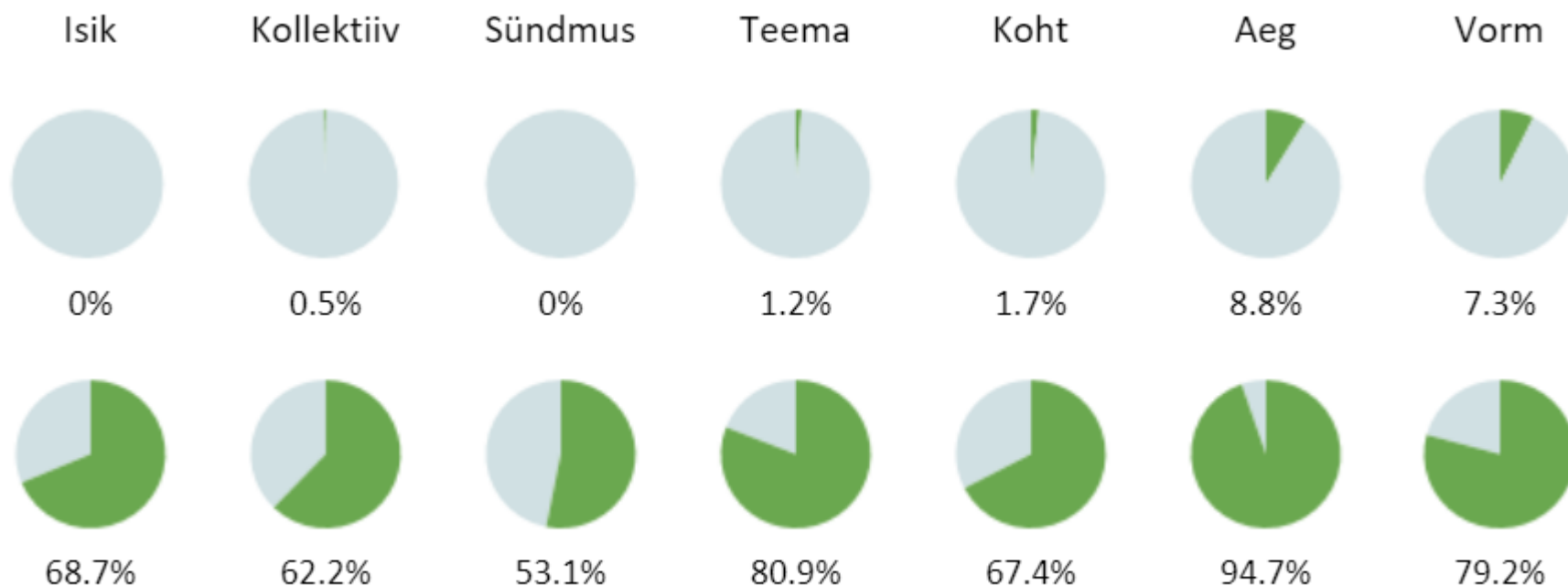
Enne segmenteerimist			Pärast segmenteerimist		
1.	eesti	556	1.	eesti	42954
2.	auditeerimine	400	2.	parlamendid	42360
3.	kontroll	393	3.	istungjärgud	39334
4.	eesti keel	279	4.	ajalugu	28533
5.	ajalugu	198	5.	eesti keel	27447
6.	uuringud	185	6.	uuringud	16002
7.	hindamine	157	7.	auditeerimine	14132
8.	raamatupidamine	136	8.	kontroll	13826
9.	parlamendid	133	9.	hindamine	12008
10.	finantseerimine	128	10.	kodulugu	11400

# Automaatne märksõnastamine

Andmeanalüüs: isikunimed märksõnana (457)

Enne segmenteerimist			Pärast segmenteerimist		
1.	Jeesus Kristus	17	1.	Jeesus Kristus	2123
2.	Konrad Mägi	4	2.	Konrad Mägi	1395
3.	Enn Nõu	4	3.	Enn Nõu	718
4.	A. H Tammsaare	4	4.	Raivo Uukkivi	691
5.	Friedebert Tuglas	4	5.	Konrad Mägi	645
6.	Aino Kallas	3	6.	Helga Nõu	627
7.	Friedrich Reinhold Kreutzwald	3	7.	Marek Vahula	584
8.	Pahlen	3	8.	Istvan Kantor	534
9.	Lev Tolstoi	3	9.	Heino Jõe	528
10.	Raoul Kurvitz	3	10.	Peeter	517

# Automaatne märksõnastamine



*Joonis 10. Eritüüpi märgendite kasutatavus enne segmenteerimist (ülemine rida) ning pärast segmenteerimist (alumine rida)*

# Automaatne märksõnastamine

<https://booktags.nlib.ee/>

## Tegelik tulemus:

- Väljaannete analüüs
- Robot ennustab EMSi märksõnad:  
teemad, kohanimed, vormid
- Töötaja valideerib:
  - kinnitab märksõna või lükkab tagasi
  - märksõnad saab kopeerida MARC21 formaadis kataloogikirjesse



# Automaatne märksõnastamine

<https://booktags.nlib.ee/>

## Kus kasutada?

- Terviklike tekstdokumentide analüüs ja märgendamine
- Eestikeelsed

## Kus mitte kasutada?

- Artiklid
- Ilukirjandus
- Pildid, noodid, muusika, kaardid jmt
- Võõrkeeled tunneb ära, aga ei märksõnasta



# Automaatmärksõnastaja

Teavik URL-i kaudu  Teavik faili kaudu

Teavik kasutaja arvutist \*

Otsi



Lehekülgede arv \*

10

Märksõnasta



Euroopa Liit  
Euroopa  
Regionaalarengu Fond



Eesti  
tuleviku heaks

<https://booktags.nlib.ee/>

# Automaatmärksõnastaja

Teavik URL-i kaudu

Teavik faili kaudu

Teavik kasutaja arvutist \*

a\_Eesti arstide tööga rahulolu\_2007.pdf



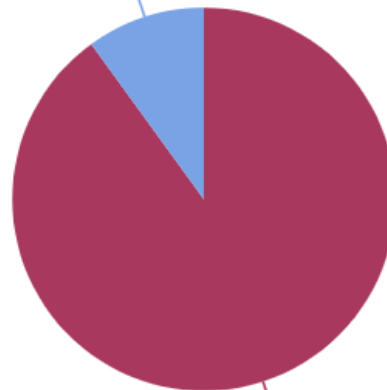
Lehekülgede arv \*

10

Märksõnasta

## Leitud keeled

inglise



eesti



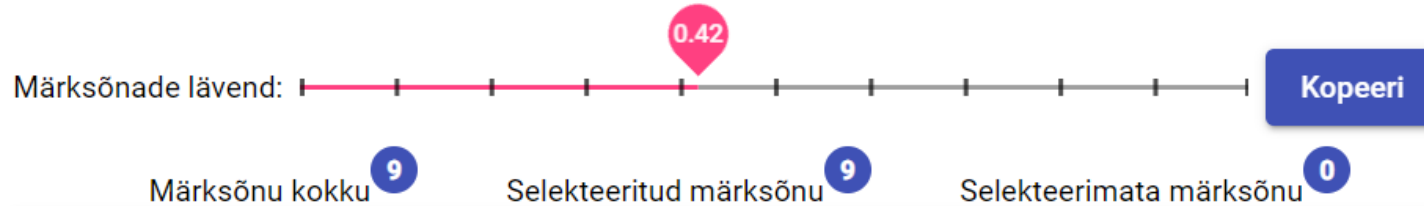
Euroopa Liit  
Euroopa  
Regionaalarengu Fond



Eesti  
tuleviku heaks



## Leitud märksõnad



### Teemamärksõna ^

- arstid
- küsitlusuuringud
- väärtusorientatsioon
- töökeskkond
- töörahulolu
- sotsiaalne käitumine

### Kohamärksõna ^

- Eesti



# Automaatne märksõnastamine

## Märksõnad MARC21 vormingus

- d 650 9 arstid
- d 650 9 küsitlusuuringud
- d 650 9 väärtusorientatsioon
- d 650 9 töökeskkond
- d 650 9 töörahulolu
- d 650 9 sotsiaalne käitumine
- d 651 9 Eesti
- d 651 9 eesti
- d 655 9 võrguväljaanded



<https://booktags.nlib.ee/>

# Päringukratt – vestlusrobot

- **Probleem:**

- Palju standardse sisuga päringuid
- Raamatukogu ei ole kättesaadav 24/7
- Küsimustele vastamine võtab aega
- Vaegnägijatele päringute esitamine keeruline

- **Õppematerjal:**

- RR päringute andmebaasid
- RR veebileht

- **Tulem**

- *chatbot ja voicebot*



# Päringukratt – vestlusrobot

## Alusandmed:

- Kokku on 839 vestlust chati logidest
- Soovituslikud kasutajalood koos fraaside ja vastustega
- Need viidi omavahel kokku

## Tulemid:

- **Algne:** tehisintellekt viis kokku 13 vestlust ja teemat → 1.6%
- **Pärast treenimist:** RR teemade osakaal tõusis 37%-ni → 312 vestlust viidi kokku etteantud teemadega
- **Lõpptestimine:** ~30% juhtudel vastas Vesta õigesti

# Päringukratt – vestlusrobot



EESTI  
RAHVUS-  
RAAMATUKOGU

🕒 E-R 10-20, L 12-19, P suletud [Muud lahtiolekuajad >](#)

Info: 630 7100

E-post: [info@nlib.ee](mailto:info@nlib.ee)

ENG

Otsi otsinguportaalist...



REGISTREERU KASUTAJAKS

[Muuda kasutaja andmeid >](#)

[enused](#) ^ [Haridustegevused](#) ^ [Mäluasutusele](#) ^ [Kirjastajale](#) ^ [Konverentsikeskus](#) ^ [Kultuurikeskus](#) ^ [Raamatukogust](#) ^

EESTI RAHVUSRAAMATUKOGU  
NATIONAL LIBRARY OF ESTONIA

TARBE-  
KUNSTMÜÜK



FUAJEE JÕULIST JÕULU,  
FOYER UHKET UUT!

JÕULUKUU  
RAAMATUMÜÜK



Algas 2019. aasta kauneimate raamatute  
konkurss — 3.12.2019

[Kuidas saan aidata?](#)





Vesta



Euroopa Liit  
Euroopa  
Regionaalarengu Fond

Eesti  
tuleviku heaks

Tere! Mina olen  
Rahvusraamatukogu  
virtuaalne assistent Vesta.  
Mind alles testitakse 😊.  
Palun esita üks küsimus  
korraga.



Otsin kataloogist

Probleemid sisselogimisel 🔑

Raamatukapp 📖

Lahtiolekuajad 🕒

Mis kell on?



Vesta



virtuaalne assistent Vesta.  
Mind alles testitakse 😊.  
Palun esita üks küsimus  
korraga.



Otsin kataloogist

Probleemid sisselogimisel 🔑

Raamatukapp 📖

Lahtiolekuajad 🕒

Mis kell on?



Kell on ajanäitaja 🕒

Kirjuta oma küsimus ...





Vesta



Euroopa Liit  
Euroopa  
Regionaalarengu Fond

Eesti  
tuleviku heaks

Tere! Mina olen  
Rahvusraamatukogu  
virtuaalne assistent Vesta.  
Mind alles testitakse 😊.  
Palun esita üks küsimus  
korraga.



Otsin kataloogist

Probleemid sisselogimisel 🔑

Raamatukapp 📖

Lahtiolekuajad 🕒

Kus saab koopiaid teha?



Vesta



Lahtiolekuajad 🕒

Kus saab koopiaid teha?



Põhjaliku info  
Rahvusraamatukogu  
kopeerimis- ja  
digiteerimisteenuste kohta  
leidad siit 🖱️

Rahvusraamatukogust koopiate  
tellimine

Soovin ise koopiaid teha

Vaata hinnakirja


Kirjuta oma küsimus ...



# Autorikaitse tähtaja kalkulaator

New Arrivals | Course Lists | Publishing Soon | Help ▾

My ESTER – Log in

 E-KATALOOG  
ENTIRE COLLECTION

ENTIRE COLLECTION

SEARCH

Keyword | Author | Title | Subject | ISBN/ISSN | Call Number | UDC | Articles | Modify Search



1164 results found. Sorted by Relevance | Year | Title

ENTIRE COLLECTION

Return To Browse Record: « Previous | Next »

Export Save to List

Author Luts, Oskar, 1887-1953  
Title **Kevade. II : pildikesed koolipõlvest / O. Luts**  
Imprint Tartus : Noor-Eesti, 1920 (Tartu : H. Laakmann)  
Description 208 lk. ; 20 cm  
Edition 3. tr.  
Permalink <https://www.ester.ee:444/record=b1212963>

 Request 

**View Rights** Report a Problem

Items Full Record MARC Display Additional Information

# Authorikaitse tähtaja kalkulaator

## COPYRIGHT EXPIRATION DATE\*

***Author:*** Luts, Oskar

***Title:*** Kevade.

***Publishing year:*** 1920

***Estimated copyright expiration  
date:*** 01.01.2024

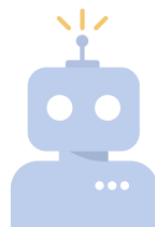
[View calculation](#)

# Mis saab edasi?

- **Kevad 2020 valmisid**
  - märksõnastaja prototüüp
  - vestlusroboti prototüüp
- **Sügis 2020**
  - Märksõnastaja lisaanalüüs
- **2021/2022**
  - Artiklite märksõnastaja (prototüüp)
  - Automaatne kataloogia (prototüüp)
  - Pildituvastaja (prototüüp) koos Muinsuskaitseametiga



+



# Kokkuvõtteks

- Tea, mida masinõpe võimaldab, mida mitte!
- Tea, mis järjekorras masinõppe projektides asjad käivad
- Ettevalmistustööd
  - Veendu, et sul oleks piisavalt alusandmeid
  - Veendu, et need oleks kasutatavad
  - Vajadusel korrasta ja puhasta
- Ole valmis, et tulemusi ei saavutata!