

TEIE ETTE ASTUB...



Urmas (andmehalduse peaspetsialist)

22
01
20

Back-up &
Storage



Surf andmejärve lainetel

ALSO[▼]
Akadeemia



EESTI
RAHVUS-
RAAMATUKOGU

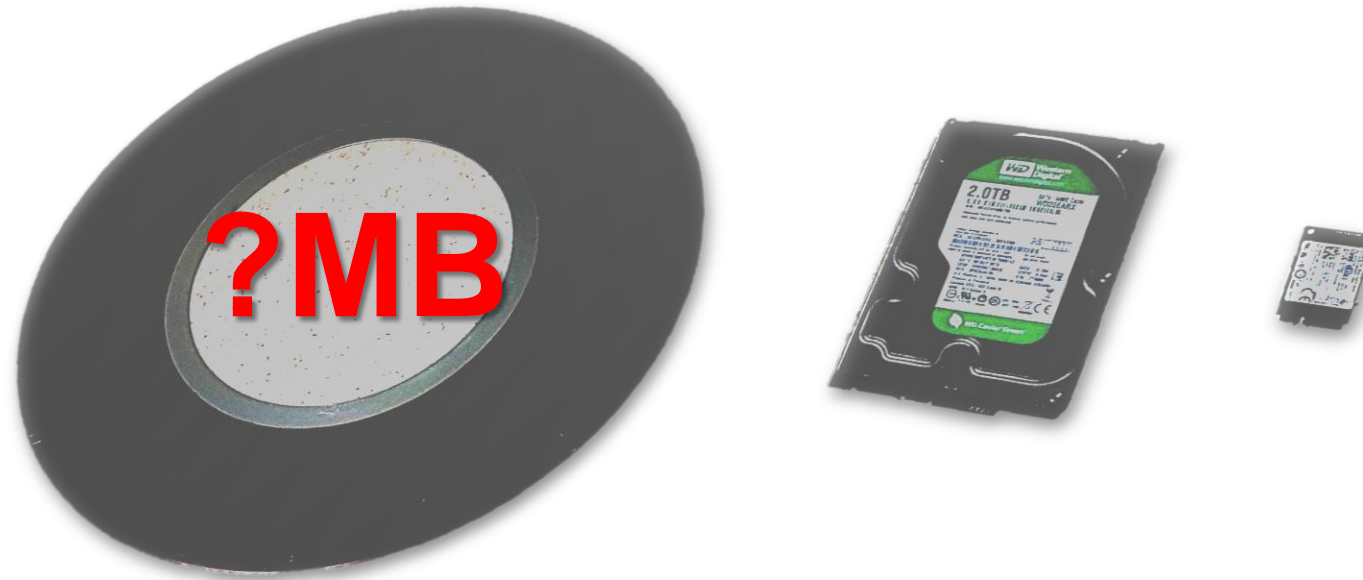
Surf andmejärve lainetel. (Surfing on DataLake Waves)

Urmas Sinisalu

Data steward

22.01.2020

.... it was long long time ago



and today You all are selling to me?

.... and data was born in library

- Book descriptions
- Library systems
- Digital content
- Digital archiving
- Long-term preservation
(longer than 10 years)
- Open Data
- Artificial Intelligence
(automated subject indexing)

.... Where is the biggest data?

- Book scanning
- Newspaper files
- Legal Deposit Books
- Web archiving
- Long-term preservation
 - Backup

Yes - Backup!

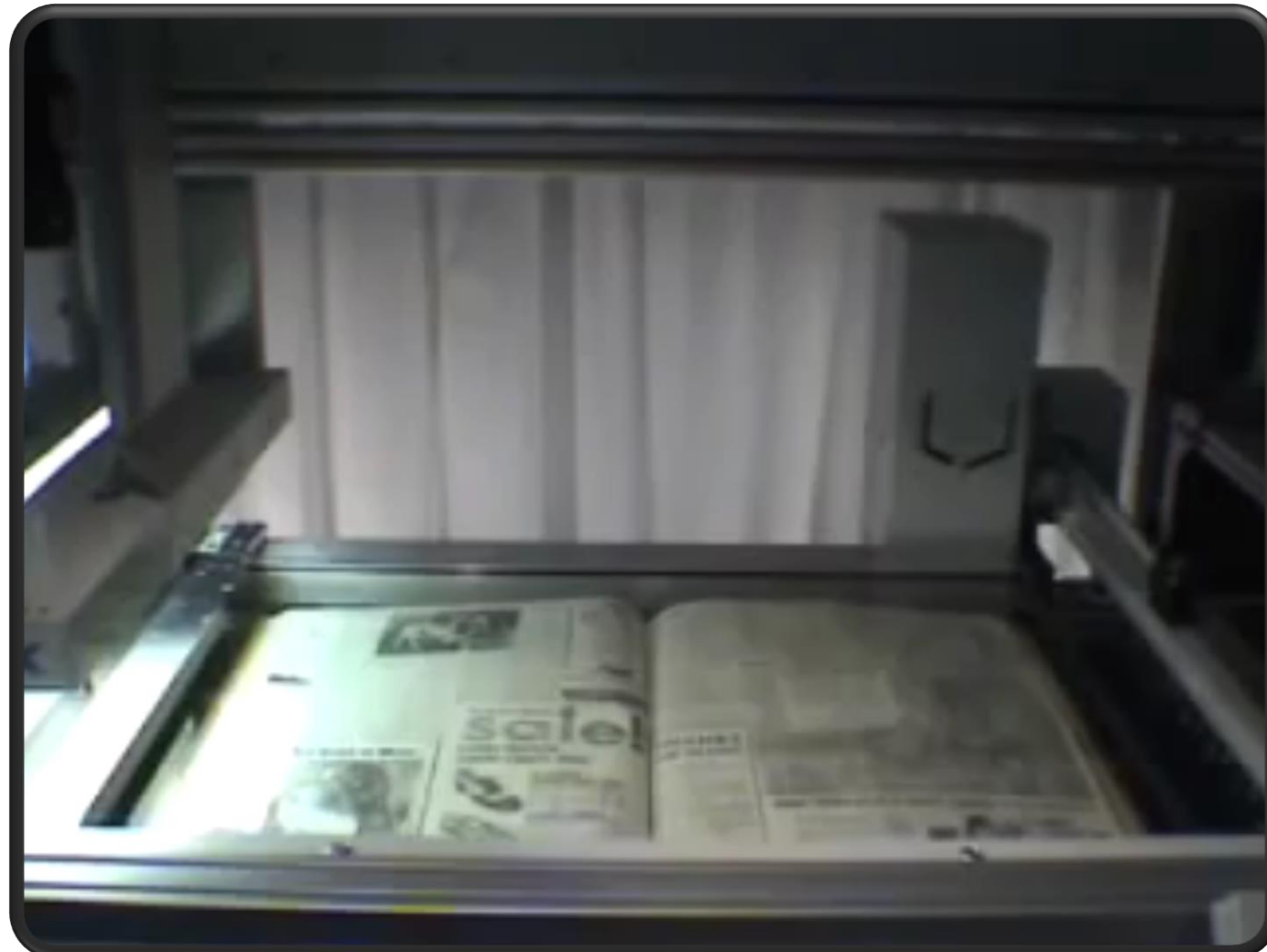
... Scanning books in „old way“!



.... Scanning book „new way“



.... Scanning newspapers!



.... Mass Digitization



primagest+

blw9d6e+

Otsi ainult kättesaadavaid

Detailne otsing

OTSI

Täistekst



Otsingute ajalugu

VALI KUUPÄEV

2020 (319)

jaanuar (319)

E	T	K	N	R	L	P
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

 Valitud kuupäev Kättesaadav Juurdepääs piiratud

TERE TULEMAST!

Riigi Teataja 100. juubeliaasta puhul digiteeris Rahvusraamatukogu koostöös Justiitsministeeriumiga 1918.-1944. aastal avaldatud Riigi Teataja väljaanded. Saad nendega lähemalt tutvuda [SIIT](#).

Ajalehed (1468)

Ajakirjad (425)

Jätkväljaanded (356)

NBI Ajalehed on portaalis alates 1821.a. (nõukogude perioodi väljaandeid ei ole veel lisatud). Ajakirjad ja jätkväljaanded on portaalis aastast 2017 (varasemaid väljaandeid vaata digitaalarhiivis [DIGAR](#)).

Artiklite portaali täiendatakse iga päev ja see sisaldab hetkel: 388394 väljaannet, 3674168 lehekülge ja 6361691 artiklit.

HUVIPAKKUVAT



PARIMAD KORREKTORID

1. Marike V. 73,814
2. Kalle G. 54,291

Preprint newspaper files - DEA.digarar.ee

[AVALEHT](#)
[OTSING](#)
[VÄLJAANDED](#)
[ILMUMISAEG](#)
[TEEMAD](#)
[ABI](#)
[LOGI SISSE](#)
[ENG](#)
[EST](#)
[PYC](#)

[Pärnu Postimees](#) > 8 jaanuar 2020

[f](#)
[t](#)
[e](#)
[+](#)

Pärnu Postimees, nr. 4, 8 jaanuar 2020

Väljaanne PDF (33.37 MB)

Väljaanne | Artikkel


Sisu

- Päismik
- Kohalikud muretsevad, mis segadus raiest metsa jääb
- ARVAMUS
 - Kas kodanikku peab tõesti kogu aeg jälgima?
 - Suur silm ja kõrv
- Leht 3 Kuulutused
 - Leht 3 Kuulutused Veerg 1
- UUDISED
 - Ühistranspordikeskus jätkab õpilaste koolitamist
 - Purjus mees ründas eakat ema, kes sai viga
 - Bussiliin 46 pikeneb kahel väljumisel Männiku peatuseni
 - Hoonete andmed tuleb kanda ehitisregistrisse
 - TABRIA TEEÄÄRSED PEAVAD OOTAMA VEEL AASTA
 - Audit viitab Rail Balticu projekti vigadele
 - Metskits veetis öö Liblika lasteaia hoovis


Väljaanne ← →



Leht 1



Leht 2



Leht 3

...Old newspaper files - DEA.digar.ee

AVALEHT
OTSING
VÄLJAANDED
ILMUMISAEG
TEEMAD
ABI
LOGI SISSE
ENG
EST
PYC

🏠 > Perno Postimees ehk Näddalileht > 5 juuni 1857
f t e +

📄 Perno Postimees ehk Näddalileht, nr. 1, 5 juuni 1857
📄 Väljaanne PDF (12.54 MB)

Väljaanne
Artikkel

Postimehhe essimenne terretamine.
<https://dea.digar.ee/article/pernopostimees/1857/06/05/2>

Tekst

NBI Tekst võib sisaldada vigu. Loe lähemalt...
 Paranda seda teksti. Logi sisse raamatukogu kasutajatunnuse, ID-kaardi või Mobiil-ID-ga

Postimehhe essimenne terretamine.
 Terre, armas Eesti rahwas! Minna, Perno Postimees. „Kulvtama wüWM " " Keik, mis sünnib ilma sees, Annan teada sulle heaks, Et ka sinno rahwas teaks, Kuida keik ma-ilma maad Ellawad ja teggewad. Awwalikud kulutussed, Rigi kässud, seädussed, Sündind asjad, mällestussed Ja keik teised teädussed Sündko siin, ehk teises rikis, Saad siin leida nende likis, Mis se kirri teada annab, Jggamehhe kätte kannab.

Kirrikudest, laste kotist, PöNo-tööst ja harrimisest, Mitmest asjast, iggast polist, Ellajatte kaswatussest: Nende terwist, arsti-rohto, Hädda aial abbi-ohto Luggedes sa leiad siit; Agga mitte ebba wiit! Andko nüüd siis Zummal sull' Head meelt ja mõistust ka, leida rõme teidust kül' Waaga alle allada: Zummal

Teemad (0)

Väljaanne



Pernus, leffnädal sei 5^{mal} Juni-ka päiwol.

Postimehhe edsimenne terretamine.

Terre, armas Eesti rahwas!
 Minna, Perno Postimees.
 Kulwtama eilen wotusok
 Keik, mis sünnib ilma sees,
 Annan teada sulle heaks,
 Et ka sinno rahwas teaks,
 Kuida keik ma-ilma maad
 Ellawad ja teggewad.

Kirrikudest, laste kotist,
 PöNo-tööst ja harrimisest,
 Mitmest asjast, iggast polist,
 Ellajatte kaswatussest:
 Nende terwist, arsti-rohto,
 Hädda aial abbi-ohto
 Luggedes sa leiad siit;
 Agga mitte ebba wiit!
 Andko nüüd siis Zummal sull'
 Head meelt ja mõistust ka,
 leida rõme teidust kül',
 Waaga alle allada:
 Zummal

Kirrikudest, laste kotist,
 PöNo-tööst ja harrimisest,
 Mitmest asjast, iggast polist,
 Ellajatte kaswatussest:
 Nende terwist, arsti-rohto,
 Hädda aial abbi-ohto
 Luggedes sa leiad siit;
 Agga mitte ebba wiit!
 Andko nüüd siis Zummal sull'
 Head meelt ja mõistust ka,
 leida rõme teidust kül',
 Waaga alle allada:
 Zummal



Tutvustus
Tagasiside
Abi
Koostööpartnerid

Eesti Rahvusraamatukogu
Tõnismägi 2, 15189 Tallinn
Kontakt: 630 7611
dea@nlib.ee



**EESTI
RAHVUS-
RAAMATUKOGU**

• ~tip!

www.rahvusraamatukogu.ee

.... DIGAR.ee

- ~300 000 objects
- biggest image file ~1GB
- ~30TB ingest buffer
- ~280TB raw data
- + ~50TB raw data only for viewer

Books in archive - DIGAR.ee

- ~300 000 objects
- biggest image file ~1GB
- ~30TB ingest buffer
- ~280TB raw data
- + ~50TB raw data only for viewer

.... Veebiarhiiv.DIGAR.ee

- ~60TB raw data

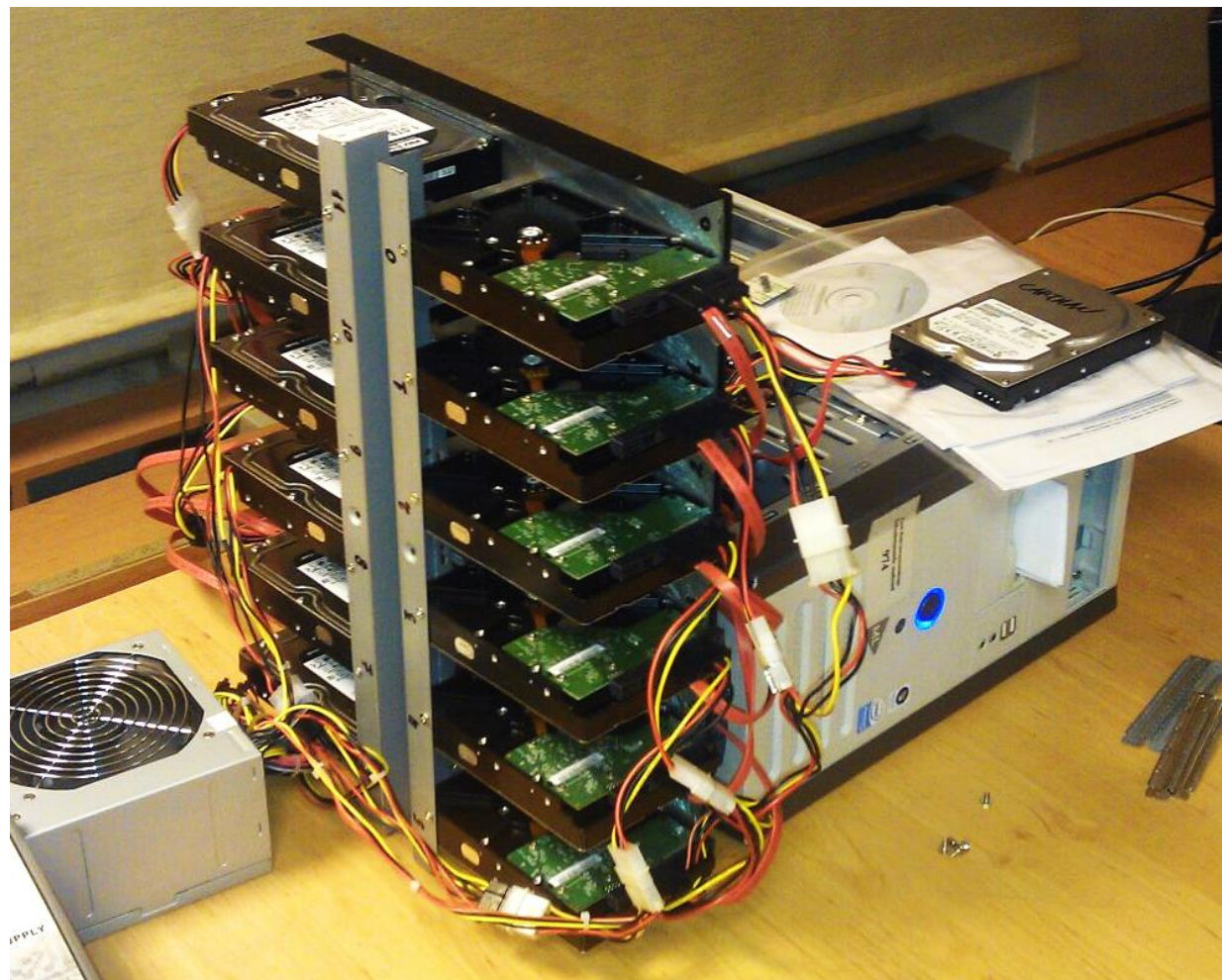
... 24h site history in 1min.

22.01.2020

The screenshot shows the ETV website interface. At the top, there is a navigation bar with 'TV', 'RAADIO', and 'POOD' options, along with a search bar and flags for Estonia, Latvia, and Lithuania. Below the navigation bar, there are several sections:

- UUDISED**: A grid of news articles with images and headlines. Headlines include 'Sakkov lahkus Nordicaest, etab vastuolusid nõukoguga', 'Galerii: jaanipäevaks peatinsa hülgavad inimesed ajasid Tallinna tänavad umbe', 'Juncker ja Tsipras: Brexiti-hirmud näitavad, et Brexit oli looli idee', 'Prokuratuur andis Savisaare advokaatidele süüasja materjalidega tutvumiseks lisaaraga', and 'Konkurentsi puudumine on Kopenhaagenisse lendamise veerandi jagu kallimaks teinud (1)'.
- OTSELEKANDED**: A section for live streaming, featuring 'OTSE kl 17: heategevuskontsert "Laulud sõdurile"'. Below this is a 'VAATA JA KUULA OTSE' section with logos for ETV, ETV2, ETV+, Vikar, Raadio 2, Klassika, Raadio 4, and R Tallinn.
- TÄNA ETV-S**: A section for today's TV programs, featuring 'Heategevuskontsert "Laulud sõdurile" Täna 17:00'.
- TÄNA VIKERRAADIO**: A section for today's radio programs, featuring 'Huvitaja, Puhkumisvõimalused Eestis'.
- Interaktiivne kaart: jaanilaupäeva ilm üle Eesti**: A large interactive map of Estonia showing weather data for January 17th. Below the map, there is a list of news items with timestamps: '12:05 Pikad pühad peatavad pankadevahelised maksed esmaspäevani', '10:20 Õged grillimisvõtted aitavad vältida vähki põhjustavate ainete teket', and '20:15 Päästekeskuse juht: jaanitõkke pole koht, kus aastaga kogunenud praht ära põletada'.

.... Long-term preservation ;)



DATA.digar.ee

DIGAR

DEA

ERB

ISIKUD/KOLLEKTIIVID

Andmekaeuurile

Näited

Kontakt

arendab mitmekesist e-kogude (Digitaalarhiiv DIGAR, Eesti rahvusbibliograafia andmebaas ERB jne) kolleksioone ning pakub erinevaid e-teenuseid, arendab personaalseid sisuteenuseid ning sotsiaal- ja humanitaarvaldkonna ja kitsamatele sihtrühmadele suunatud teenuseid.

Avaandmete kasutamisel nõustute Creative Commons Public Domain (CC0) litsentsitingimustega.

Rahvusraamatukogu digiarhiiv DIGAR

DIGAR (<http://www.digar.ee>) on Rahvusraamatukogu digiarhiiv, mis pakub juurdepääsu digitaalarhiivis talletatud väljaannetele. Nende hulgas on e-raamatuid, ajalehti, ajakirju, kaarte, noote, fotosid, postkaarte, plakateid, illustratsioone, audioraamatuid ning muusikafaile. Raamatute ja perioodika vorming on enamasti pdf või epub, pildimaterjalil jpeg ja helisalvestistel wav.

DIGAR'i andmefailid tehakse kättesaadavaks EDM vormingus (Europeana Data Model) läbi OAI-PMH protokoll.



what next - artificial intelligence (AI)

Kratt - tekitab ikke ainult kahju (kellelegi) st varastas. Ja lõpuks viis omaniku hinge minema, sest kratt sai tekkida (elustuda) ainult lepingust (kolm tilka verd) vanatühjaga. Omaniku tahe kratti luues oli asju naabritelt öö varjus „oma aita tõsta“.

Pätt - programmse ärioloogikaga toetatud tehnoloogia

Tehistaip

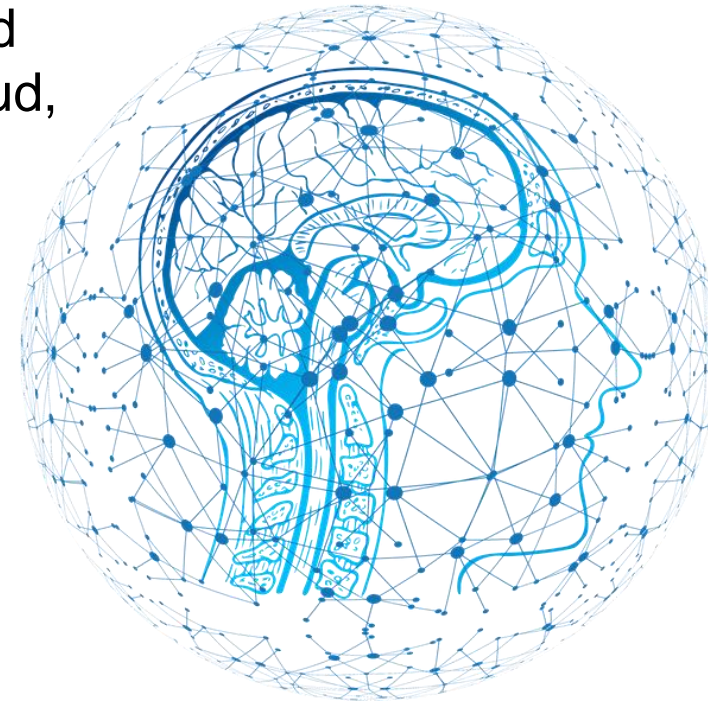


<https://www.flickr.com/photos/46785534@N06/15028865166/in/photostream/>

www.rahusraamatukogu.ee

Automaatne märksõnastamine

- **Tulem:**
 - Uute väljaannete analüüs
 - Robot pakub välja sobivad märksõnad (teemad, kohanimed, ajaperiood, isikud, organisatsioonid)
 - Töötaja valideerib ja teeb valiku
 - Märksõnad saadetakse kataloogi bibliokirjesse
- **Kus kasutada?**
 - Tervikväljaannete analüüs
 - Artiklite märgendamine
 - Veebiarhiivis veebilehtede sisuanalüüs



Automaatne märksõnastamine

- **Based ja õpetamine:**
 - EMS (40 000 subjects, 20 000 äraviidet)
 - ERB (200 000 book record)
 - DIGAR (~10 000 objects, 1940 - 2019)
- **Eeldus:** tekstide edukaks märgendamiseks iga ennustatava märgendi jaoks vaja vähemalt 50 näidisdokumenti
- **Tulem:**
 - Kasutuses kõigest 8000 märksõna
 - ~7600 tekstilist faili



Päringurobot „Küsi meilt“

- **Probleem:**
 - Palju standardse sisuga päringuid
 - Raamatukogu ei ole kättesaadav 24/7
 - Küsimustele vastamine võtab aega
 - Vaegnägijad peavad päringute esitamiseks pöörduma
- **Õppematerjal:**
 - Olemasolev päringute andmebaas
 - RR veebileht
- **Tulem**
 - *chatbot* ja *voicebot*

