

Topeltkirjete liitmise automatiseerimisest

Martin Pajuste, ELNET Konsortsium

Sissejuhatus

Ülesandepüstitus

Tööl on kaks eesmärki:

1. leida topeltkirjed,
2. mestida leitud kirjed nõuetekohaselt.

Tööplaan

1. Leia võimalikud kandidaadid
2. Analüüsi tulemusi
3. Täienda reegleid
4. Korda samme 1.-3.
5. Automaatne liitmine

LIHTNE

Ülesandepüstitusest

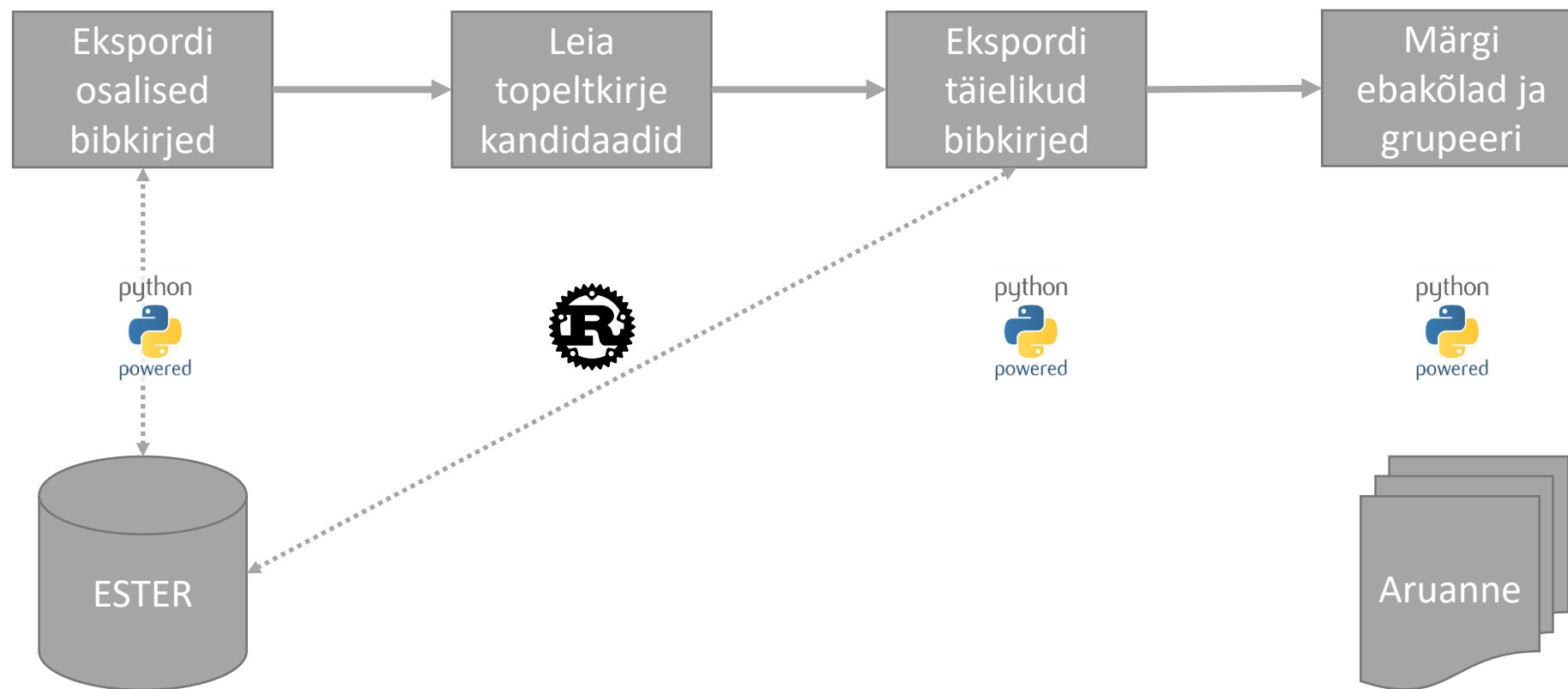
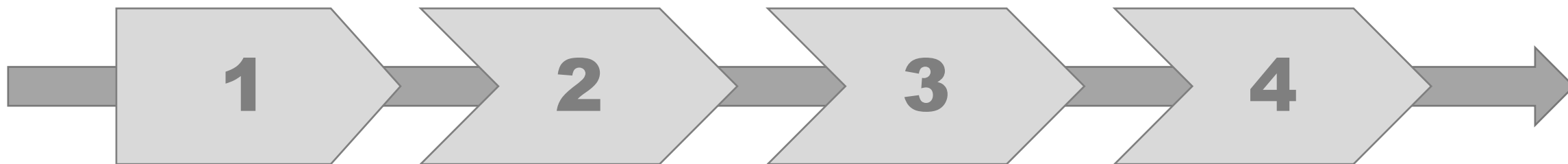
1. Topeltkirjete leidmine

- Vali MARC väljad
- Jõu kasutamine ei vii kuhugi
- Kitsenda võrreldavate paaride hulka
- Leia viis kirjeldamiseks paaridevahelist sarnasust
- Genereeri interaktiivne aruanne

2. Topeltkirjete mestimine

- Topelt tähendab 2 või rohkem
- Eksemplarikirjete ümbertõstmine
- Sierra REST API
- Mestitavate kirjete vastavus kataloogimisreeglitele
- Arenduse (osade) hilisem kasutamine

Sierrast aruandeni



1

Lähtematerjal

Isiklikust kogemusest:

- MARC21,
- kataloogimisjuhendid.

Ekspordi
osalised
bibkirjed

ESTER



1

Lähtematerjal

Kirjest on eraldatud ühe pika sõnena loendis olevad väljad koos kõigi alamväljadega:

- 100 - ISIKUNIMI PEALDISENA (MK)
- 110 - KOLLEKTIIVAUTOR PEALDISENA (MK)
- 240 - ÜHTLUSPEALKIRI (MK)
- 245 - PEALKIRJA- JA VASTUTUSANDMED (MK)
- 250 - EDITSIOONIANDMED (K)
- 260 - ILMUMISANDMED (K)

Ekspordi
osalised
bibkirjed



ESTER

1

Lähtematerjal

- 300 - FÜÜSILINE KIRJELDUS (K)
- 440 - SEERIAANDMED (K)
- 490 - SEERIAANDMED (K)
- 505 - SISU MÄRKUS (K)

Ekspordi
osalised
bibkirjed

ESTER

1

Lähtematerjal

Ekspordi
osalised
bibkirjed

ESTER

```
id search
420909489491 |aBeljajev, Riina |aõppimine ja õppimisoskuste arendamine täiskasvanuna|h[V...
420909489492 |aMapping colonial conquest :|bAustralia and Southern Africa /|cedited by M...
420909489493 |aPõdra, Andres |aVanglatöö taktika|h[Võrguteavik] :|bteenistus tugevdatud ...
420909489494 |aLindgren, Astrid,|d1907-2002 |aPipi ajab kõik joonde ja teised lood /|cAs...
420909489495 |aStrata :|bdeserts past, present and future : an environmental art project...
420909489496 |aВойны русского народа :|b1558-1878 гг. : библиографический указатель оспо...
420909489497 |aTables of fractional powers /|cprepared by the Mathematical Tables Projec...
420909489498 |aГо, Мо-жо,|d1892-1978 |aИзбранные сочинения :|b[стихи, рассказы, пьесы, с...
420909489499 |aRaipala-Cormier, Virpi,|d1952- |aLoomulikult ilus :|btee ilu ja tervise j...
420909489500 |aГо, Мо-жо,|d1892-1978 |aПеснь о бушующей волне :|b[воспоминания об антияп...
420909489501 |aVuorela, Einari,|d1889-1972 |aRunoja /|cEinari Vuorela ; esipuheen kirjoi...
420909489502 |aМенгер, Антон, 1841-1906 |aОбщественные задачи правоведения :|b(речь, чит...
420909489503 |aГо, Мо-жо,|d1892-1978 |aЦюй Юань :|bисторическая трагедия /|cГо Мо-жо ; п...
420909489504 |aDenizer, Cevdet |aStabilization, adjustment and growth prospects in trans...
420909489505 |aВознесенский, А. А. |aЛенин и оборона социалистического отечества :|bk XV...
420909489506 |aSherratt, Tim |aInigo Jones :|bthe weather prophet /|cTim Sherratt ; Aust...
420909489507 |aMichalopoulos, Constantine |aPayments and finance problems in the commonw...
420909489508 |aМенгер, Антон,|d1841-1906 |aПраво на полный продукт труда :|bперевод с не...
```



2

Eeltöötlus ja paaripanek

Leia
topeltkirje
kandidaadid

Sõnele rakendatakse järgmisi teisendusi:

- UNICODE → ASCII,



2

Unicode

Leia
topelkirje
kandidaadid

- ü - 'LATIN SMALL LETTER U WITH DIAERESIS',
ü - 'LATIN SMALL LETTER U', 'COMBINING DIAERESIS'
- Suur ja väiketähtede võrdlemisel erilised reeglid -
ss·ß
- Sarnased märgid erinevates tähestikes



2

Eeltöötlus ja paaripanek

Leia
topeltkirje
kandidaadid

Sõnele rakendatakse järgmisi teisendusi:

- UNICODE → ASCII,
- ABC → abc,
- eemaldatakse „|[a-z0-9]“,
- teisendatakse bigrammideks.



3

Rohkem andmeid

Ekspordime:

- Püsipikkusväljad
- Muutpikkusväljad

Sierra DNA

Ekspordi
täielikud
bibkirjed



ESTER



4

Reeglid ja eeskirjad



Märgi
ebakõlad ja
grupeeri

- Vastavalt kata töörühma juhilt saadud reeglitele kontrollitakse kahe kirje vastavust.
- Vastuolud paarides märgistatakse.
- Kirjete grupeerimine kindla tunnuse alusel.

Demo

Aitäh