

KOHA-integraatioita (sormiharjoituksia)

- miksi ja miten
- case JYU

Ari Häyrinen
Matti Lassila



Miksi?

- Kohassa on tietoa, jota tarvitaan muissa järjestelmissä
- Kohan tietoja halutaan analysoida

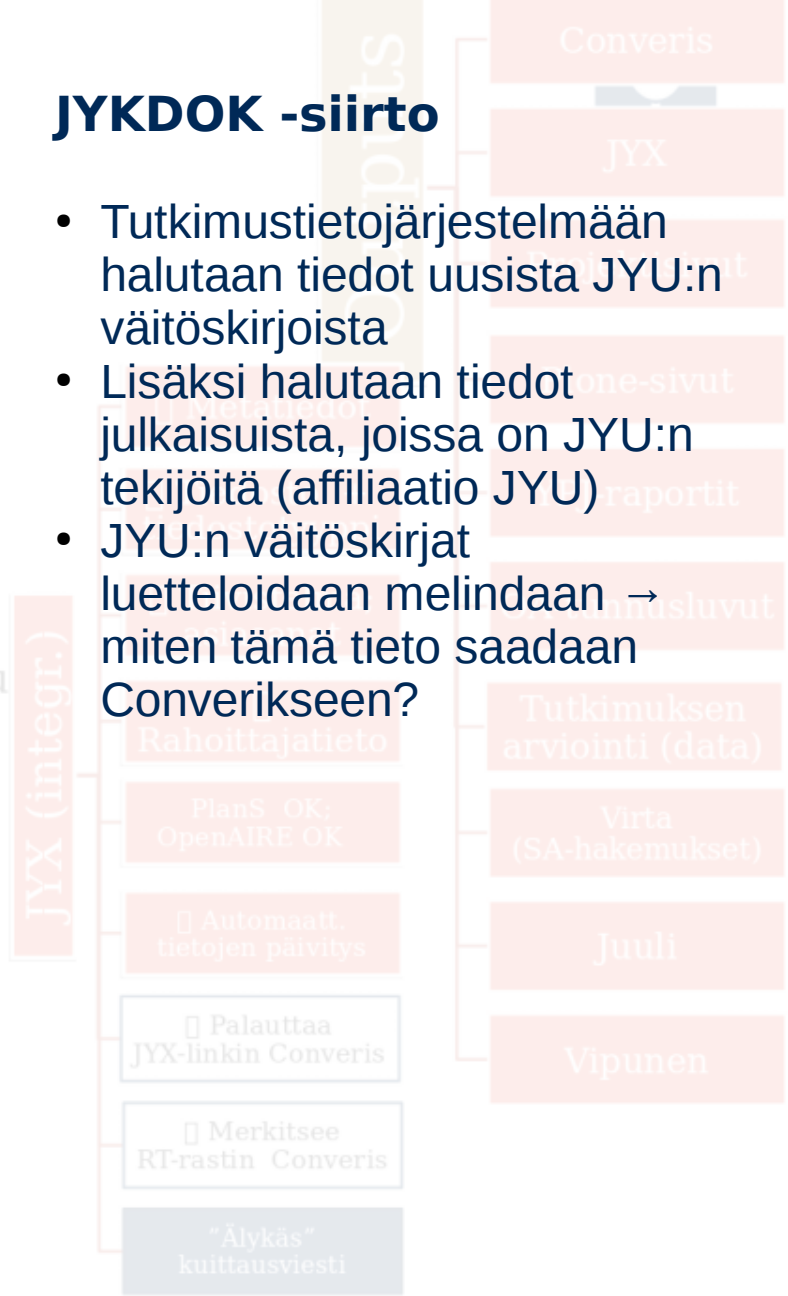
(M-L Harjuniemi
JYU, 3.10.2019)

Inputs



JYKDOK -siirto

- Tutkimustietojärjestelmään halutaan tiedot uusista JYU:n väitöskirjoista
- Lisäksi halutaan tiedot julkaisuista, joissa on JYU:n tekijöitä (affiliaatio JYU)
- JYU:n väitöskirjat luetteloidaan melindaan → miten tämä tieto saadaan Converikseen?





Mitä minun piti saada?

Selkokielineen JSON -tiedosto, jonka voi viedä työnkulkuun*

- pelkkiä JYU-isbn:niä tai JYU:n väitöskirjat liputettuina

*)työnkulku = mongo (GLAMpipe) + selainpohjainen sovellus



Mitä KOHA tarjoaa?

- Haravointirajapinta OAI-PMH
- Raportit
- REST-rajapinta
- SRU/SRW

perustavanlaatuinen ongelma:

- osa tarvittavista tiedosta MARC-tietueen sisällä (hyvin hitaita kyselyjä)
- kyselyt tehdään tuotantopalvelimella



Vaihtoehto 1: CSV

- SQL-clientin kautta otettu CSV -dumppi
 - kaikki tieto
 - n. 1,5 miljoonaa tietuetta
 - MARCXML yhtenä kenttänä
 - iso tiedosto (n. 5Gt)



Vaihtoehto 1: CSV → väitöskirjat

- 1) CSV -dumppi (SQL-client)
- 2) CSV luetaan dokumenttietokantaan
- 3) MARCXML muutetaan JSON-rakenteeksi
- 4) poimitaan JYU:n isbn
- 5) merkataan lipulla väitöskirjat (502 a)
- 6) ”puretaan” MARC selkokieleiseksi
- 7) väitöskirjan tiedot saadaan JSON-rakenteisina työnkulkuun



Vaihtoehto 1: CSV → väitöskirjat

- tiedot saatiin
- ison CSV:n käsittely raskasta
 - CSV ei ollut validi → vaati merkkijonomuunnoksia
 - 8 CPU:n kone käsitteli n. tunnin
- validin CSV:n kanssa tuonti CSV-dumpista on ihan ok vaihtoehto



Vaihtoehto 2: Finna

- Haetaan tiedot Finnasta!
 - [https://finna.fi/Search/Results?limit=0&filter\[\]=~format_ext_str_mv:"1/Thesis/Thesis/"&filter\[\]=~building:"0/JYU/"&lookfor=jyväskylä&type=AllFields&filter\[\]=search_daterange_mv:"\[2019 TO 2019\]"](https://finna.fi/Search/Results?limit=0&filter[]=~format_ext_str_mv:)



Vaihtoehto 2: Finna → väitöskirjat

- Nopea ja kevyt
- Toimii väitöskirjojen tapauksessa
- MUTTA kyselyllä ei saada JYU:n affiliaatiolla olevia julkaisuja Finnasta (AFAIK)



Vaihtoehto 3: MARC

- MARC -dumppi Kohasta (esisuodatettu, M. Lassila)
- `yaz-marcdump biblio-metadata-dump.mrc`
`>marcxml.xml`
- Prosessoidaan Pythonilla
 - ”puretaan” MARCXML
- Tuloksena ”selkokielineen” JSON, joka voidaan viedä työnkulkuun



Sormiharjoitukset

- CSV -dumpppi
 - Mongo + JS → JSON
- Finna
 - Mongo + JS → JSON
- MARC -dumpppi
 - python → JSON

Tuloksena kaikista ”selkokielineen” JSON,
joka voidaan viedä työnkulkuun



Vaihtoehto 4:

- Matti Lassila jatkaa

Rajatun käyttötarkoituksen hakuindeksointi

Tavoitteena yhtenäinen opinnäytehaku

- JYKDOK
 - Koha
 - MARC
- JyX
 - DSpace 6.x
 - Dublin Core





kyberturvallisuus Hae

Menetelmä Valitse ▾
Tieteenala Valitse ▾
Taso 5 valittuna ▾
Vuodesta 1920 ▾
Vuoteen 2019 ▾
PDF saatavissa i

➤ Hakuvalinnat

CLOSE

Tuloksia yhteensä: 33

Taso	Oppiaine	Tekijä	Otsikko	Vuosi	PDF
Väitöskirja	Tietojenkäsittely ja informaatiotieteet Valtio-oppi, hallintotiede Muut yhteiskuntatieteet	Kari, Martti J.	Russian strategic culture in cyberspace : theory of strategic culture - a tool to explain Russia's cyber threat perception and response to cyber threats	2019	
Pro gradu - tutkielma	Tietojärjestelmätiede	Palonen, Olli-Pekka	Kyberturvallisuuden johtaminen Virossa, Israelissa ja Alankomaissa : mitä voimme oppia?	2019	
Pro gradu - tutkielma	Tietotekniikka	Niemelä, Jukka	Kyberturvallisuusalan työvoiman kysyntä, saatavuus ja kehittäminen vastaamaan työvoiman tarvetta Suomessa	2019	
Pro gradu - tutkielma	Tietotekniikka	Immonen, Riku	Teollisuusautomaatiojärjestelmien ja teollisen internetin kyberturvallisuus	2019	

Catmandu



Extract
Transform
Load

Koha ⇨ MARC-osadumpppi ⇨ JSON
⇨ Elasticsearch

```
select marc_any_match('502', '*.Jyv.*')
```

```
select marc_any_match( '020a' , '978-951-39.*' )
```

```
select marc_any_match( '100u', '.*Jyv.*' )
```



```

156 1308806 3362 rdacontent
157 1308806 337a käytettävissä ilman laitetta
158 1308806 337b n
159 1308806 3372 rdamedia
160 1308806 567a Tapaustutkimus.
161 1308806 579a XLUETTELOITU
162 1308806 650a koulutus
163 1308806 6502 ysa
164 1308806 650a opetus
165 1308806 6502 ysa
166 1308806 650a opettajat
167 1308806 6502 ysa
168 1308806 650a opinto-ohjelmat
169 1308806 6502 ysa
170 1308806 650a koulut
171 1308806 6512 ysa
172 1308806 651a maseutu
173 1308806 6502 ysa
174 1308806 651a Suomi
175 1308806 6512 ysa
176 1308806 700a Kimonen, Eija.
177 1308806 700u Jyväskylän yliopisto
178 1308806 7737 nnam
179 1308806 773w (FI-J)1307266
180 1308806 773t Transforming teachers' work globally : in search of a better way for schools and their comm
unities. -

```

<https://asciinema.org/a/276009>

```

osboxes@osboxes:~/Documents/2019/catmandu-demo/data$ catmandu breaker --fields 100u jyu-author-affil.breaker
| name | count | zeros | zeros% | min | max | mean | variance | stdev | uniq~ | uniq% | entropy |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| # | 1453 | | | | | | | | | | |
| 100u | 1474 | 0 | 0.0 | 1 | 2 | 1.01445285615967 | 0.0 | 0.1 | 30 | 2.0 | 0.9/10.5 |
osboxes@osboxes:~/Documents/2019/catmandu-demo/data$ catmandu breaker --fields 700u jyu-author-affil.breaker

```