



Aalto University
School of Science



Oopperasammon datapalvelu ja Sampo-UI-työkalu sammon taontaan

Heikki Rantala

Datapalvelu

- “Sampo-mallissa” data ja käyttöliittymä pyritään erottamaan
- Tutkija ei ole riippuvainen vain web käyttöliittymästä
- Data on ladattavissa eri muodoissa ja kyseltävissä SPARQL-rajapinnan kautta
- <https://www.idf.fi/dataset/operasampo>



SPARQL kyselykieli

- Tietokantaa voi kysellä suoraan SPARQL kyselykielellä
- Lähinnä sovelluskehittäjille, mutta myös teknisesti osaaville tutkijoille
- Esimerkki: https://api.triplydb.com/s/eyY8_ZKa-



SPARQL esimerkki

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX opera-s: <http://ldf.fi/schema/operasampo/>

```
SELECT ?theaterLabel (COUNT(DISTINCT ?performance) AS ?performances)
WHERE {
  ?performance a opera-s:Performance . # Valitaan kaikki esitykset
  ?performance opera-s:performedIn ?theater . # Teatterit joissa esitys on esitetty
  ?theater skos:prefLabel ?theaterLabel . # Teatterin nimike
}
GROUP BY ?theaterLabel
ORDER BY DESC(?performances)
```



Tulevaisuus?

- **Linkitetty data on joustavaa ja sitä on helppo yhdistää erilaisiin lähteisiin**
- **Yhdistävät datamallit?**
- **Yhteiset sanastot?**



Sampo-UI sovelluskehys

Sampo-UI

Search all content

PERSPECTIVE 1 PERSPECTIVE 2 PERSPECTIVE 3 CLIENTFS

FEEDBACK INFO ▾ INSTRUCTIONS EN ▾

Sampo-UI

"Here to forge for us the Sampo,
Hammer us the lid in colors"

Select a perspective to search and browse the knowledge graph:

Perspective 1
Perspective 1 description

Perspective 2
Perspective 2 description

Perspective 3
Perspective 3 description

ClientFS
Client-side faceted search

Sampo-UI

- Avoimen lähdekoodin kehitysympäristö linkitetyn datan portaaleille
- <https://github.com/SemanticComputing/sampo-ui>
- <https://seco.cs.aalto.fi/tools/sampo-ui/>
- Otettavissa helposti käyttöön ja ylläpidettävissä



Fasettihaku

- Keskeisenä periaatteena **fasettihaku**
- Käyttäjälle näytetään aina kaikki valittavissa olevat vaihtoehdot ja niiden osumamäärät
- Tulokset voi esittää eri tavoin: lista, karttoja, kaavioita...
- Haun voi tehdä “exploratiivisesti”
- Fasettien osamaluvut kertovat itsessään paljon
- Fasettihaku ja visualisaatiot voivat paljastaa virheitä datassa brutaalilla tavalla



FASETIT / SUODATTIMET

Hae entiteettien kotisivu

Tulokset valitussa muodossa

Sisältö ⓘ

Puheenvuoron tyyppi ⓘ

Puhuja ⓘ

Puhujan puolue ⓘ

Hae...

- Kansallinen Kokoomus [20987]
- Vihreä liitto [7523]
- Vasemmistoliitto [6880]
- Suomen Kristillisdemokraatit [44]
- Suomen ruotsalainen kansanpuo

TAULUKKO

AIKAJANA

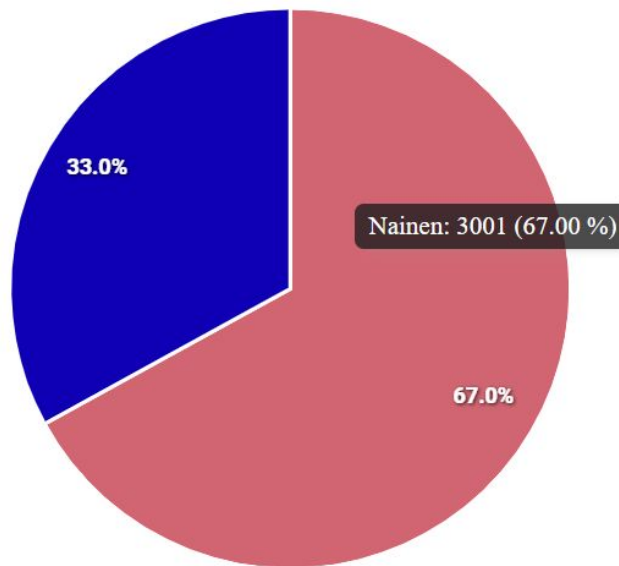
PIIRAKKA/PYLVÄSK

KARTTA

LÄMPÖKARTTA

Lajitteluperuste: puhujan sukupuoli ▾

Kaavion tyyppi: piirakkakaavio ▾



Nainen: 3001 (67.00 %)
Mies: 1478 (33.00 %)

Puheenvuorot 1907- (i)

Tulokset: 5491 puheenvuoroa

Valitut suodattimet POISTA KAIKKI

Sisältö: nato* X

Rajaa:

Sisältö (i)
Hae...
nato* 🔍

Puheenvuoron tyyppi (i) ▼

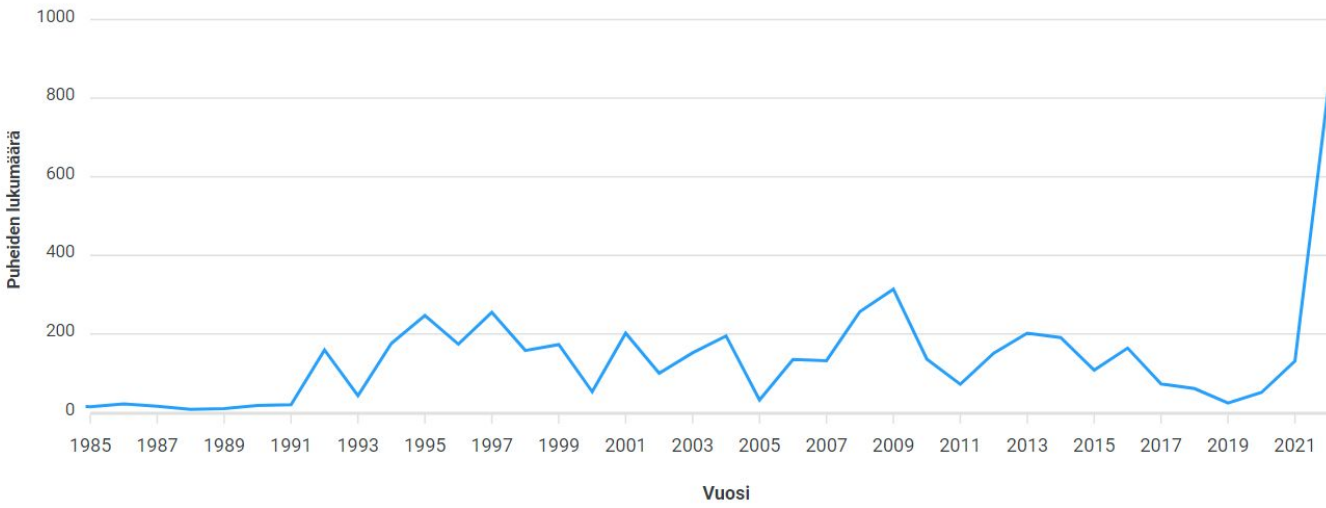
Puhuja (i) ▼

Puhujan puolue (i) ▼

Puhujan sukupuoli (i) ▼

TAULUKKO **AIKAJANA** PIIRAKKA/PYLVÄSKAAVIO CSV SPARQL-KYSELY

Puheiden lukumäärä vuosittain



Sampo-UI:n käyttö

- Kaikki SeCo:n uudet “sampo” sovellukset
- KAVI:n televisioarkisto (tulossa)
- Maailmalla esim. norjalaiset nimistöntutkijat
- Käyttöliittymä on lopulta helppo juttu
- Hyvä data on vaikeinta tuottaa!

