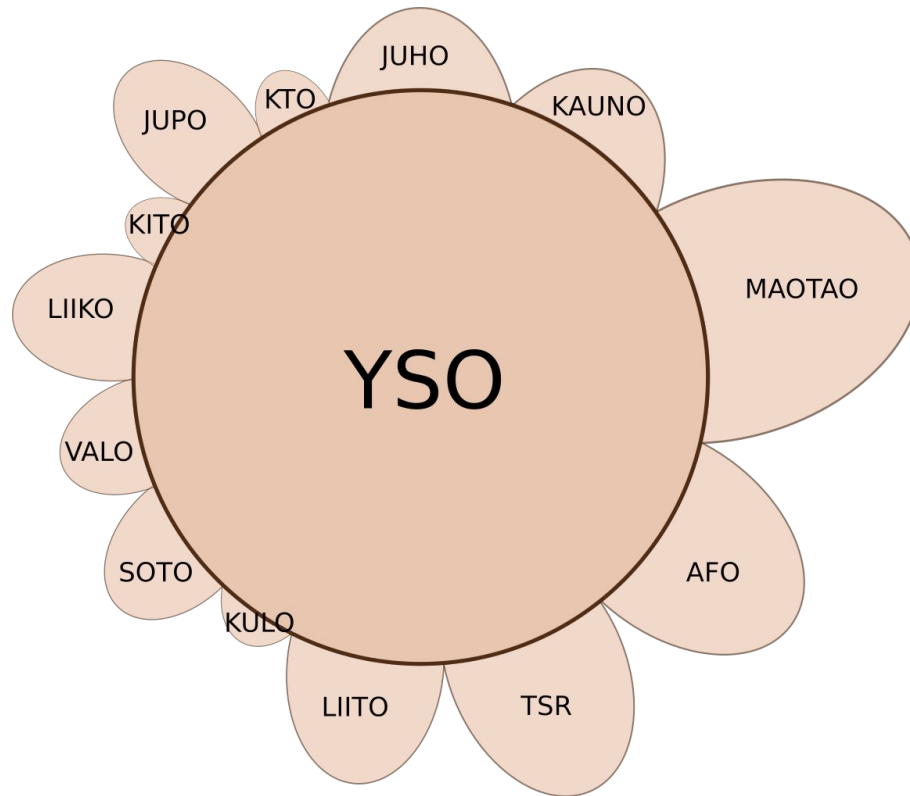![finto — Finnish Thesaurus and Ontology Service]

Automated tools for propagating a common hierarchy from a set of vocabularies

Joeli Takala

Finto service / National Library of Finland

Linked Open Ontology Cloud?

- Vocabularies in Linked Open Data
  - But how interoperable they really are?
- Everybody develops the vocabularies from their own perspective
  - 15 Domain-specific vocabularies
  - One Finnish General Upper Ontology
  - In total 7.6M triples
  - Smash it together into ~1M triple
    Finnish Holistic Collaborative Ontology (KOKO)

Henry George, 1886 - Special Collections & Archives, Georgia State University Library

Tools for the process

- Vocabulary editors (TBC, Vocbench)
- Quality improvement tool SKOSIFY
- Change analysis tool MUTU
➔ Validators
➔ Merge tool KOKOAJA

Data Model, Validator and KOKOAJA

- The detail level of linking in the domain ontologies is determined after the General Finnish Ontology
- Even though the resulting linked cloud is syntactically linked, it does not guarantee a logically sensible outcome
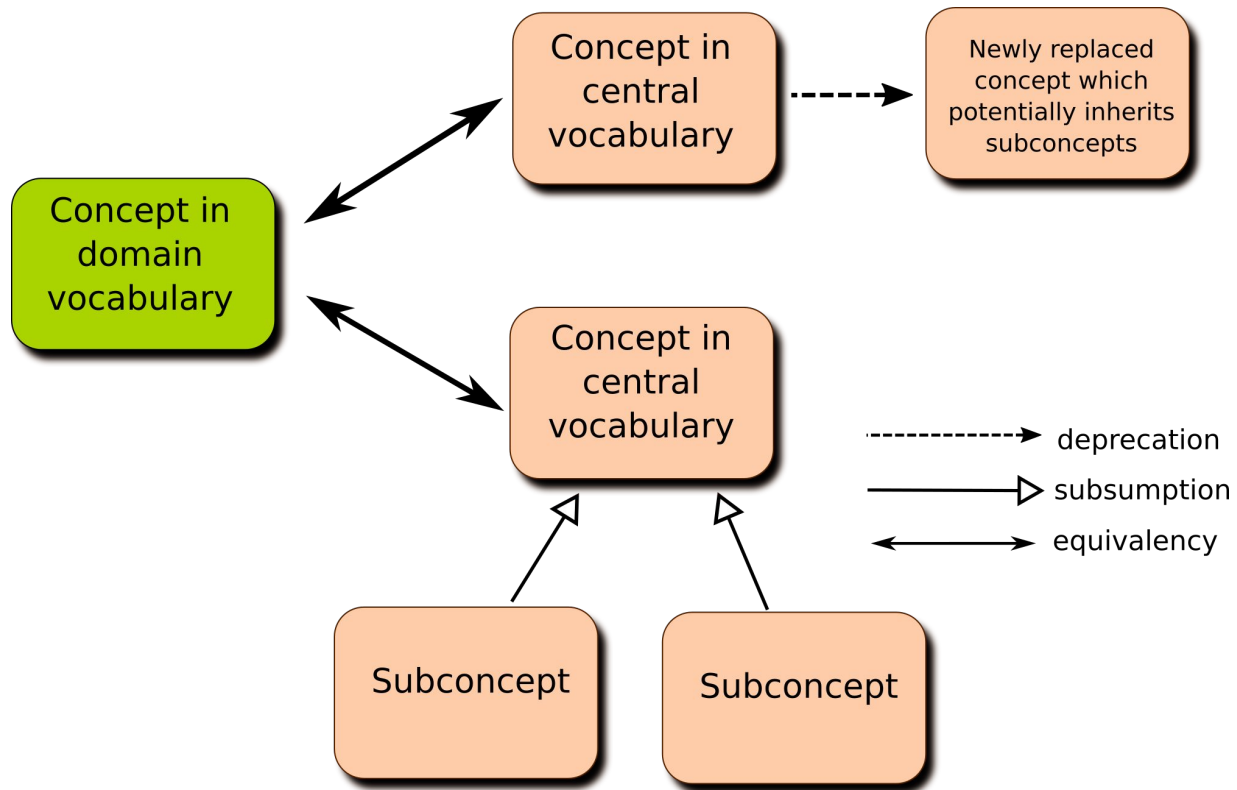- List of principles which aim to ensure a reasonable quality of semantics
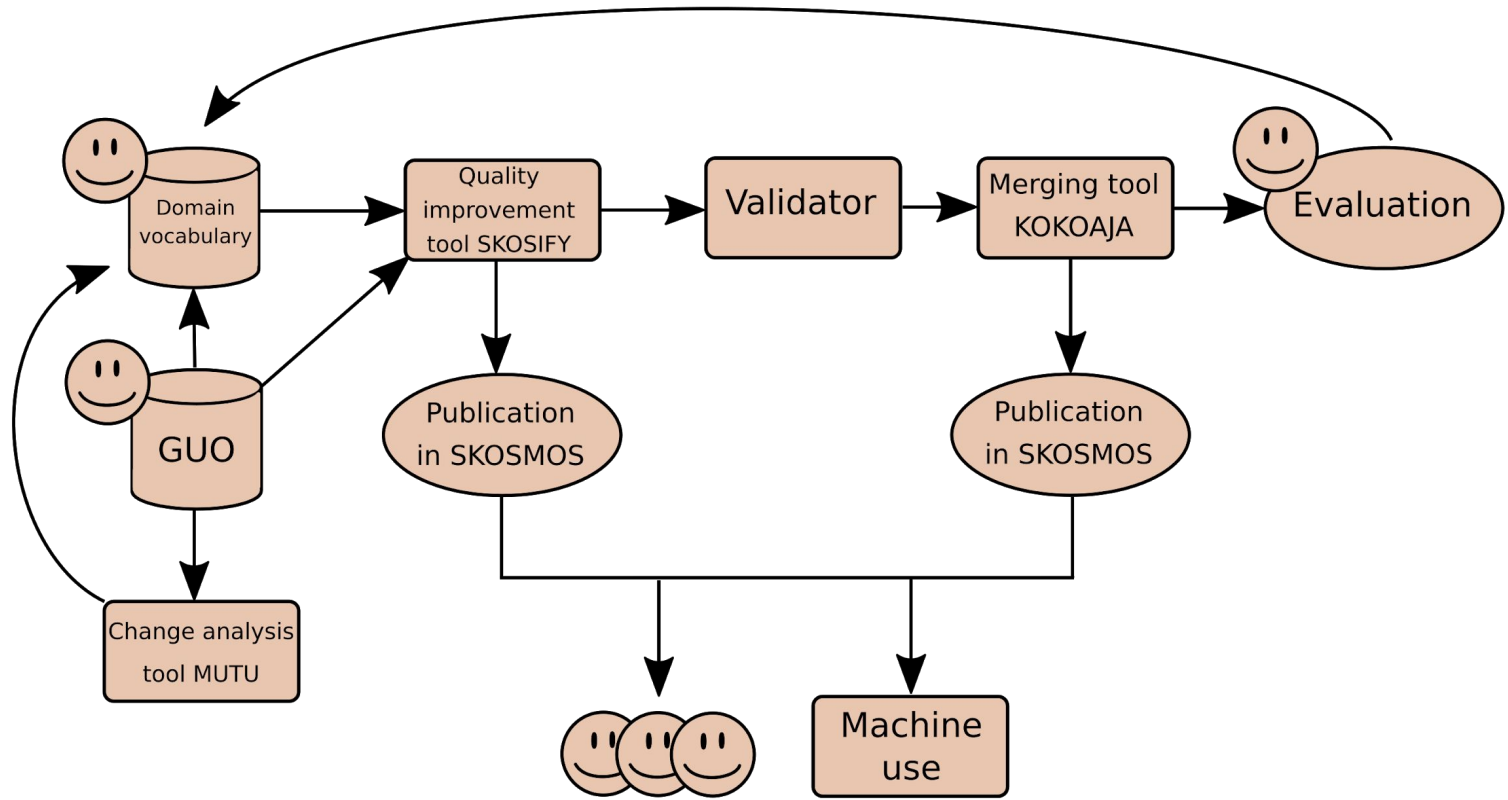
The Update Process

- Automated processes
- Validators
- Proactive design guidelines
- Synchronization
  - Push or pull?
 - Central or distributed?
- Manual labour

Error Types

- Guiding principles:
  - Each concept should have its own URI
  - Each URI should be unique
  - No URIs should be removed when a concept is removed/split
- Domain vocabularies should be linked together only via the central ontology
- Statements in the hierarchy of the domain vocabularies should not change the hierarchy of the central ontology

Concept in central vocabulary

Newly replaced concept which potentially inherits subconcepts

Concept in domain vocabulary

Concept in central vocabulary

- - - - - → deprecation

———▷ subsumption

←———→ equivalency

Subconcept

Subconcept

The Merge Tool Procedure

1. Load vocabularies to a graph store
2. Choose the main terms for each concept
3. Group equivalent concepts together
4. Assign URIs for the new concepts
5. Replace old links to point to the current URIs
6. Add references to old KOKO URIs
7. Update concept substitutions
8. Fix outlier concepts in the hierarchy
9. Report remaining error situations
10. List changes in relation to the previous output
11. Save the list of newly created URI equivalencies

Acceptance testing

Skosify

Kokoaja

Validation

Skosify

Concept analysis
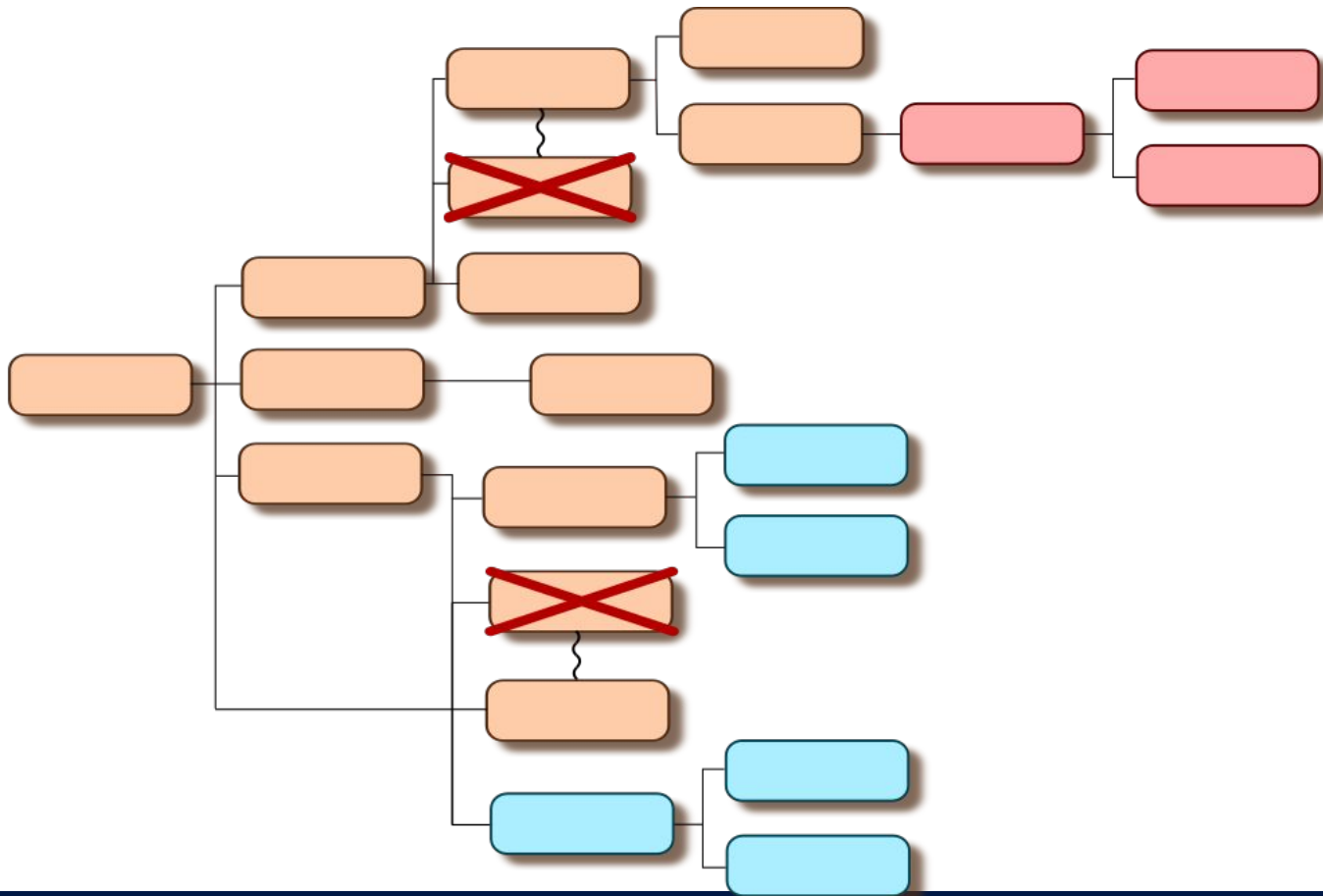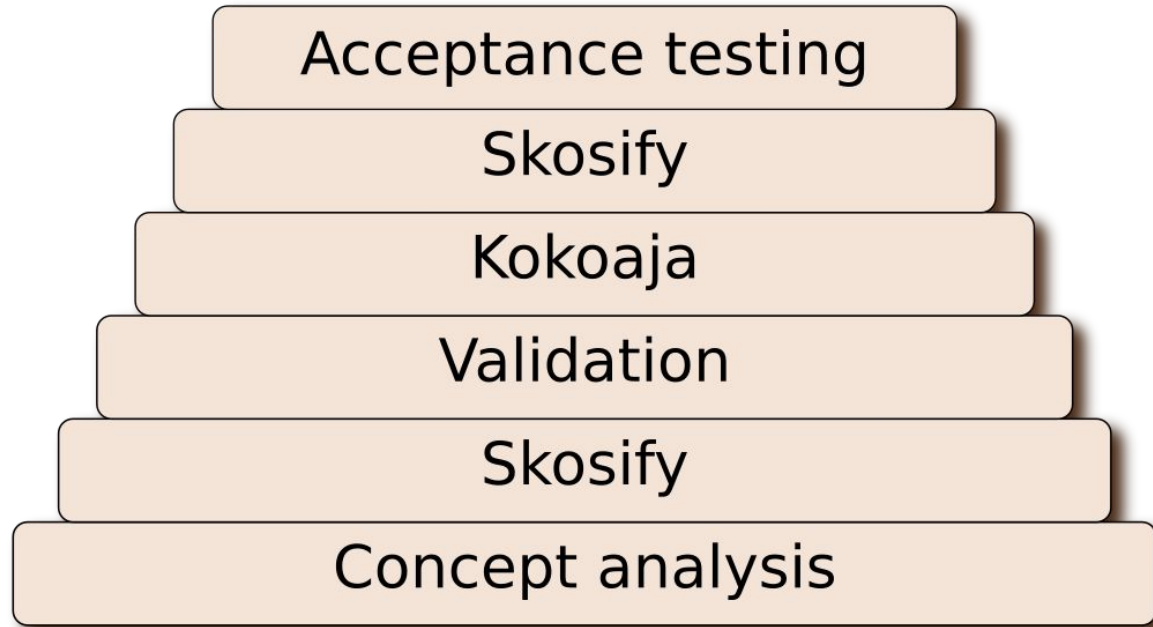
How much the process can be automated

- Algorithms

- Vocabulary development guidelines

- Synchronized updating process

- Validator reports

Major users of Finto

- All libraries
- Many museums
- Finnish Broadcasting Company Yle
- Archives
- National Institute for Health and Welfare
- Prime Minister's Office
- Government web portals
- Ministry of Employment and Economy
- EnterpriseFinland portal
- National Land Survey
- City of Helsinki
- Parliament of Finland

KOKO

- In development since 2003:
  - E. Hyvönen, M. Frosterus, S. Pessala, O. Suominen
- Based on a modular set of tools to solve each subproblem
- Ongoing iterative development process
- Balance between manual labour, automated processes and proactive guidelines

finto-posti@helsinki.fi

https://finto.fi/