Skosmos

annif

Open source tools for SKOS vocabulary services: Skosmos and Annif

Osma Suominen, Joeli Pokkinen, Mona Lehtinen, Juho Inkinen

NKOS 2025, Tampere, Finland

23 September 2025

1640
NATIONAL LIBRARY OF FINLAND

Part I:
Skosmos and Finto: publishing SKOS controlled vocabularies on the Web

Osma Suominen

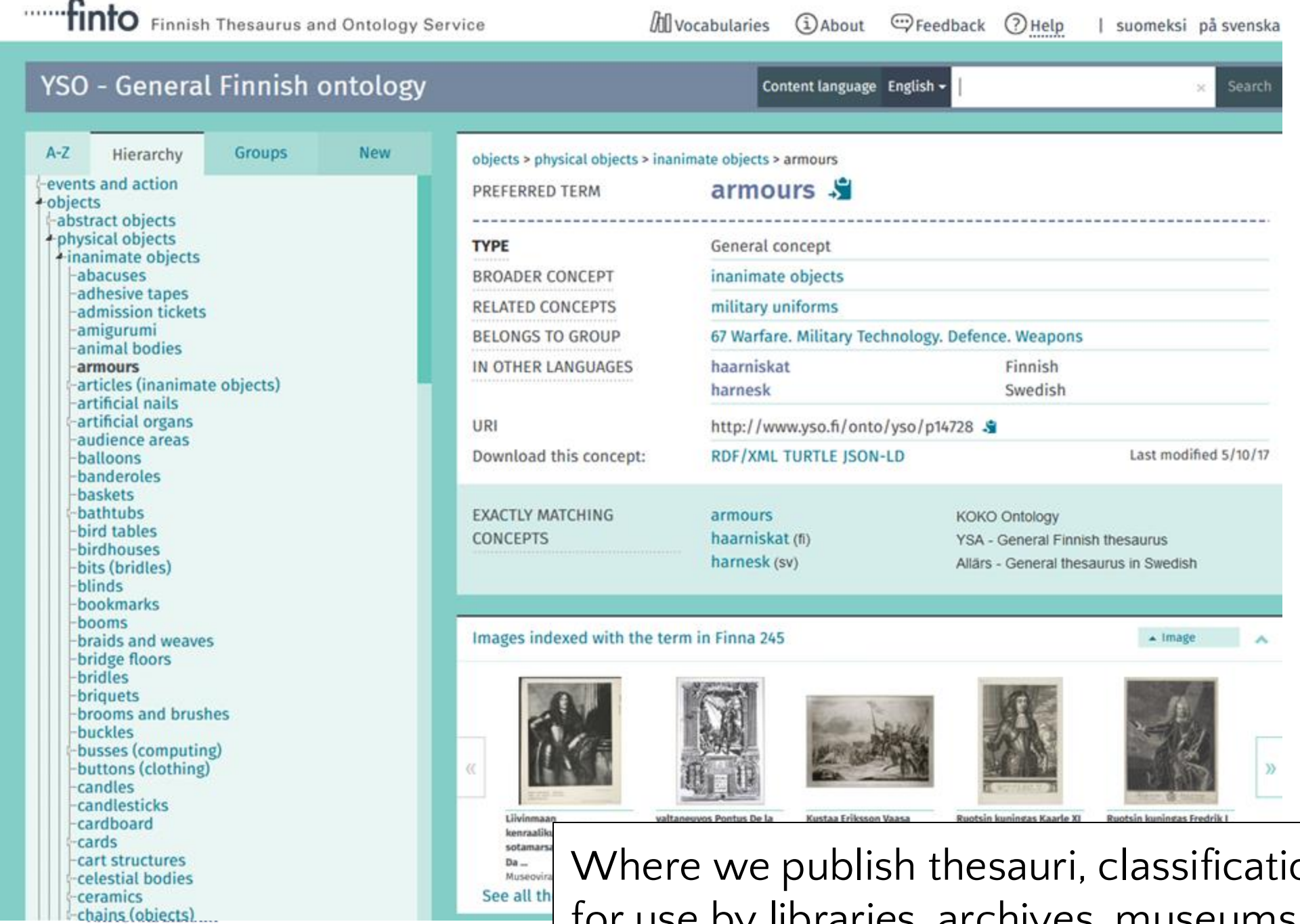
16 40
NATIONAL LIBRARY OF FINLAND

Skosmos

- Web-based SKOS vocabulary browser and publishing tool
  - SKOS is a Linked Data standard for representing thesauri, subject headings, taxonomies, classifications, lightweight ontologies…
- Facilitates subject indexing, information retrieval and vocabulary development
- Developed by the National Library of Finland
  - with contributions from many others
- Open source (MIT license)

Finto service of the National Library of Finland

- Finto started as a project in 2013, financed by Finnish Ministry of Finance and the Ministry of Education and Culture
- Launched to the public in 2014
- Permanent service at the National Library of Finland since 2019
- Service and infrastructure of linked data vocabularies for the *whole Finnish Public Sector* (not just libraries, museums and archives)



finto

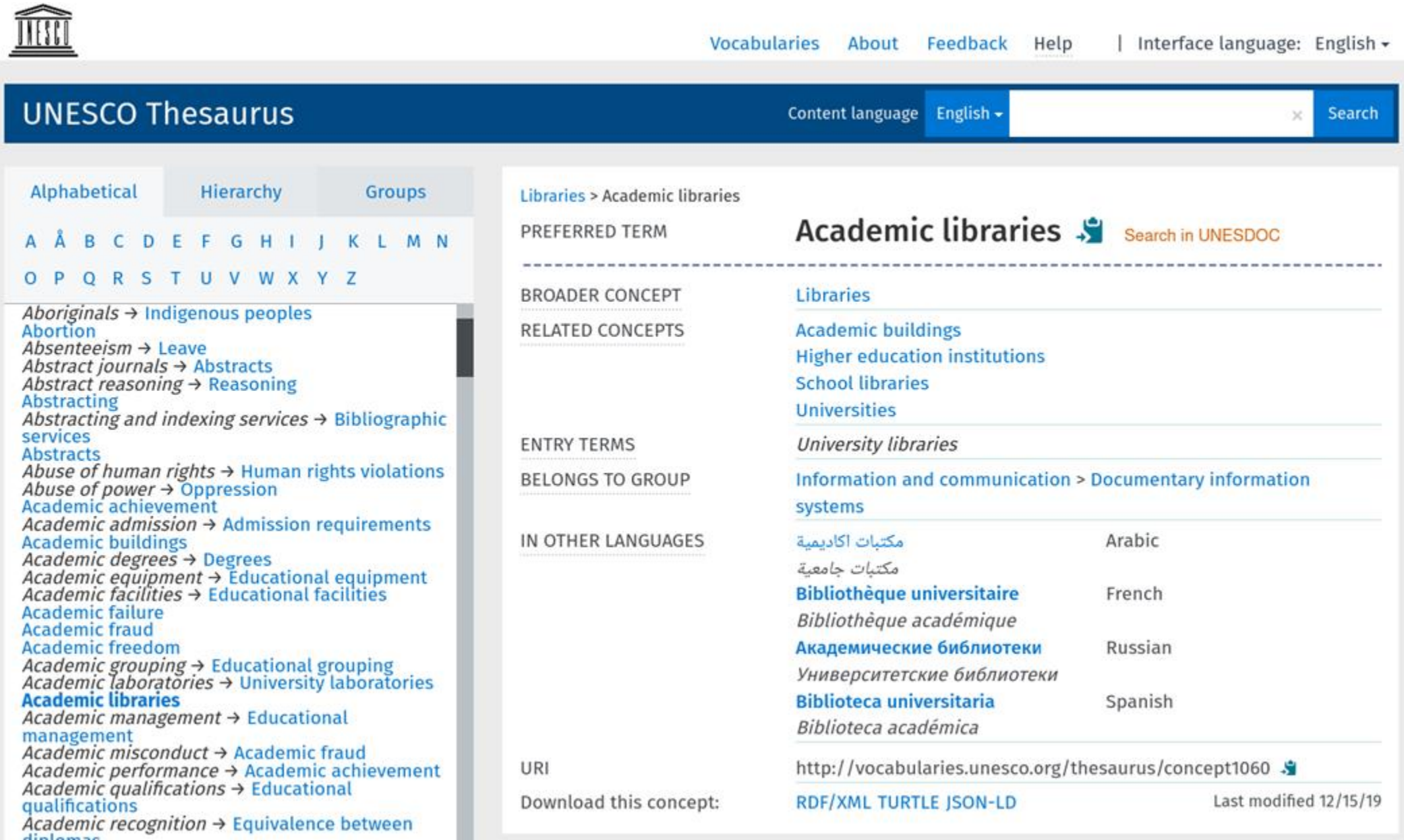
Finto.fi

Where we publish thesauri, classifications, ontologies etc.
for use by libraries, archives, museums, media, students...

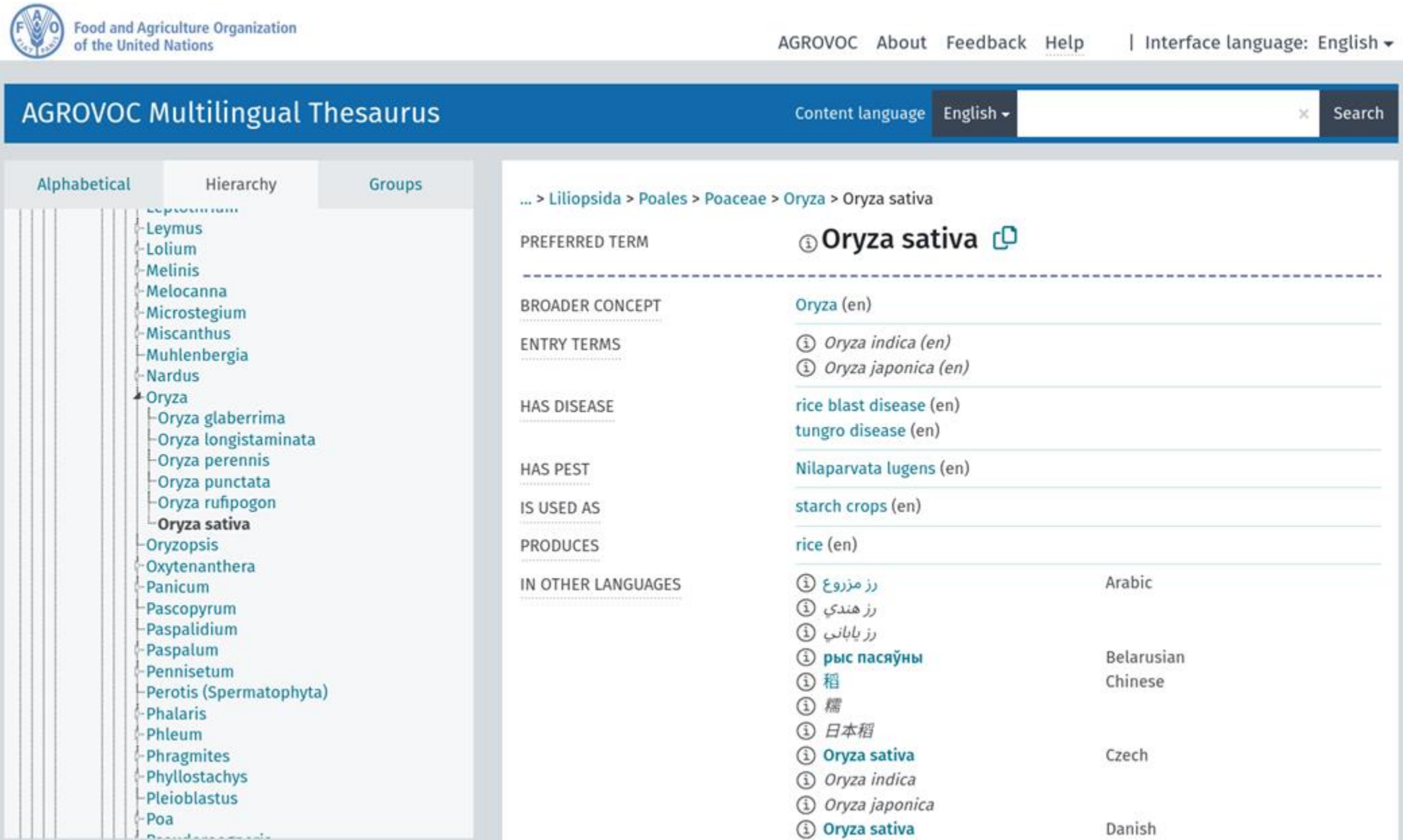
OF FINLAND

Skosmos installations around the world

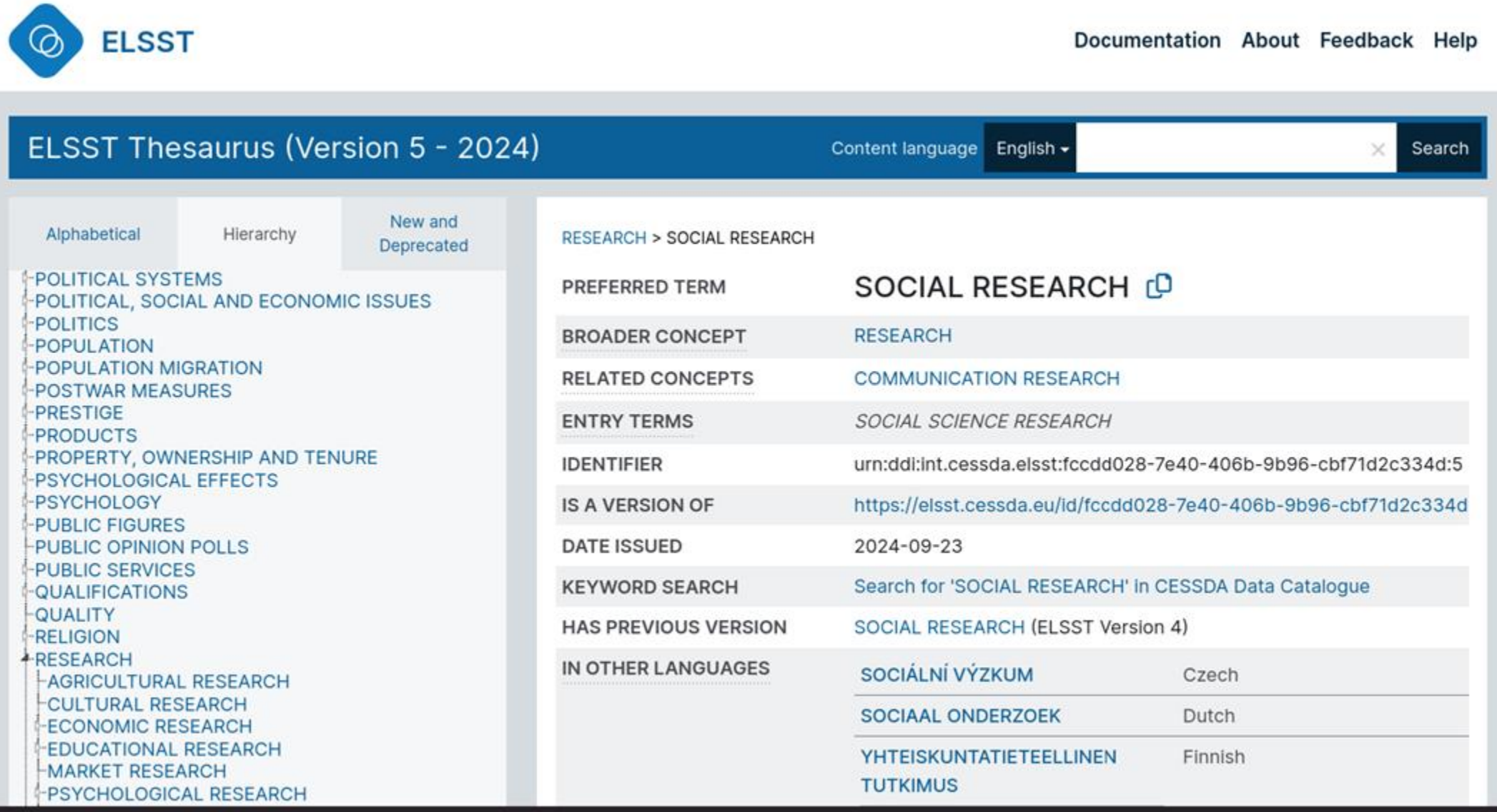
UNESCO Thesaurus (UNESCO, UN)



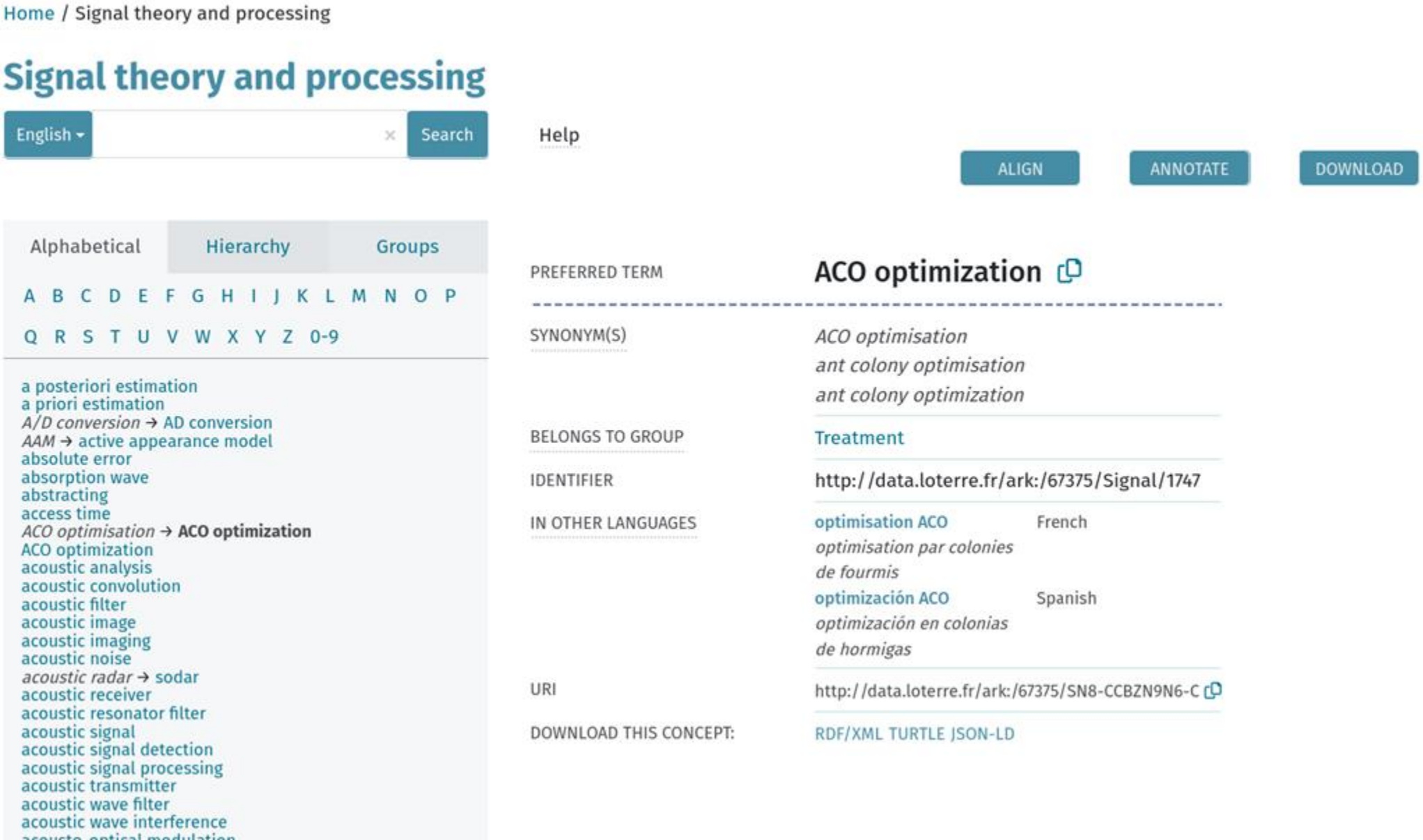
AGROVOC Thesaurus (FAO of the UN)



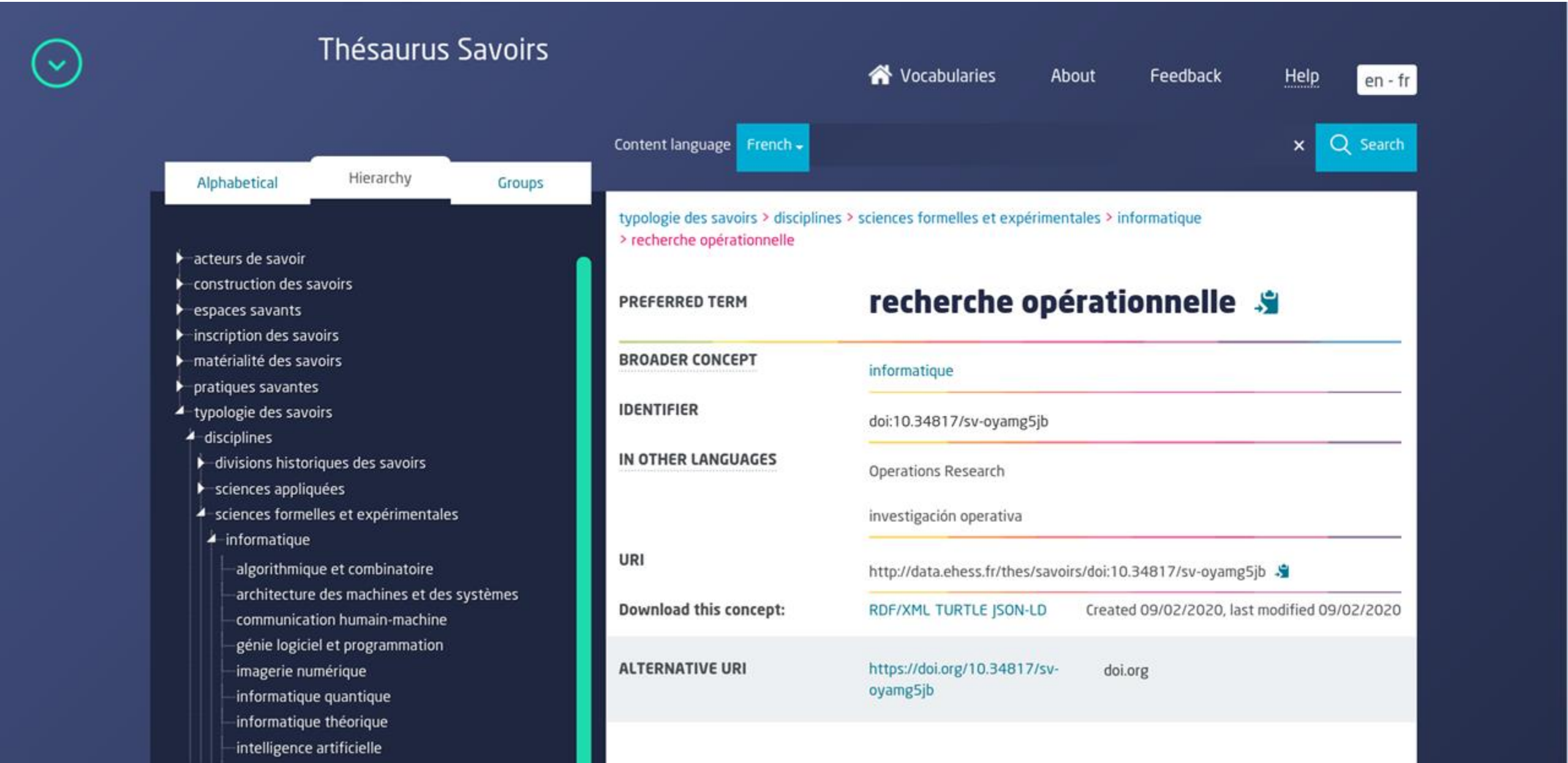
ELSST Thesaurus (CESSDA)



Loterre (ISTEX, France)

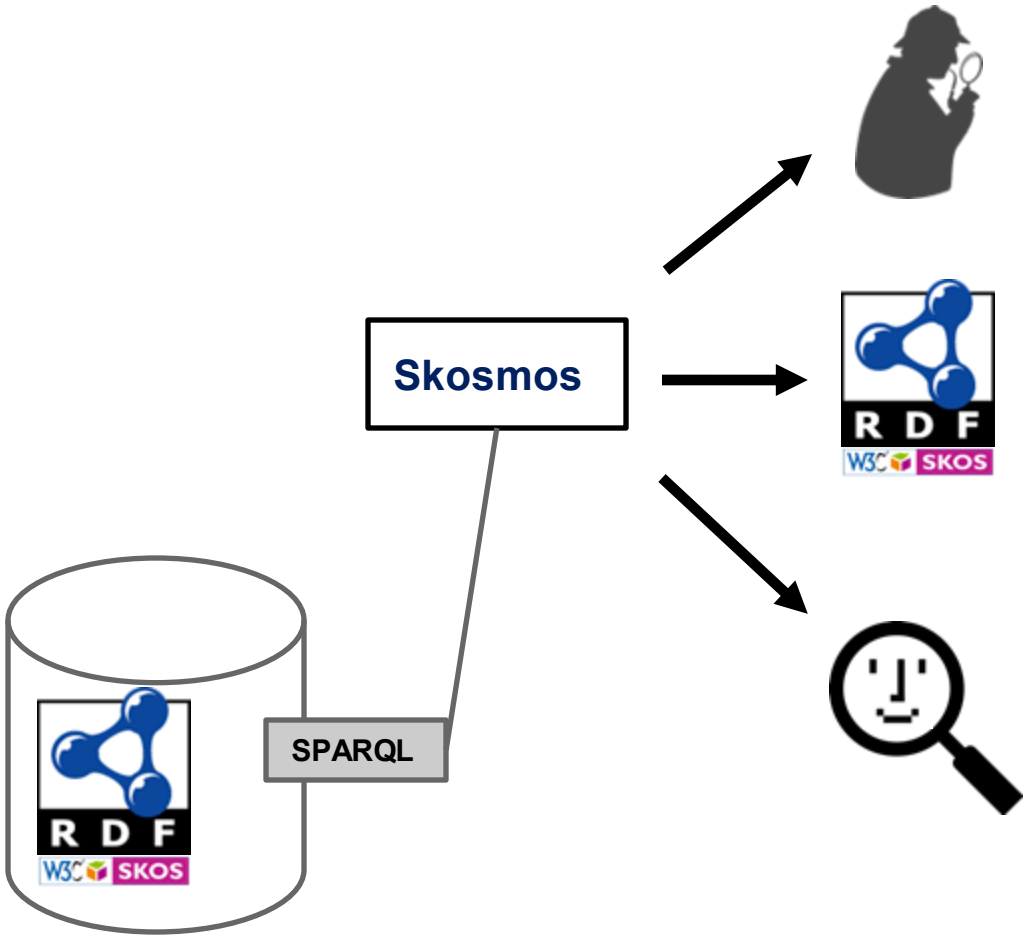


Datu (EHESS, France)



Architecture

- Database: RDF triple store with SPARQL
  - usually Apache Jena Fuseki
  - text index (jena-text) for efficient searching
- PHP backend
  - generates HTML pages for humans/browsers and search engines
  - provides RDF linked data
  - REST API for integrations
- JS/HTML/CSS frontend
  - client side dynamic functionality



Skosmos is a traditional web application with server side generation. Not a Single Page Application (SPA)!

History of Skosmos

2011 Development of ONKI Light started
at the Semantic Computing Research Group (SeCo)
(predecessors ONKI 1, ONKI 2 and ONKI 3 were developed earlier)

2012 First publication "ONKI Light on SPARQL" (EKAW 2012)

2013 Maintenance handed over to the National Library of Finland.

2014 Finto.fi launched. Renamed to Skosmos. Code moved to GitHub.

2015 Skosmos 1.0 released. Skosmos-users group started.

2018 Skosmos 2.0 released: new configuration format and Docker support.

2022 Planning for Skosmos 3 started. New visual design created.

2023 First Skosmos 3 development sprints

Why build Skosmos 3, a new major version?

1. Many UI components used in Skosmos 2 are no longer maintained, so we need to replace them
2. The Skosmos 2 frontend doesn't use any framework except jQuery, which has led to a messy JavaScript codebase
3. Changing the above requires a full front-end rewrite
4. This is an opportunity for improvements in many areas:
    a. Reduce dependencies for easier maintenance
    b. Better JS, HTML, CSS and PHP code quality
    c. Introduce automated frontend testing
    d. Improve UX and accessibility

Overview of changes in Skosmos 3

1. New visual style
2. Reimplemented frontend using Vue.js
3. Small changes to the PHP backend
4. Localization using Symfony Translations and Lokalise

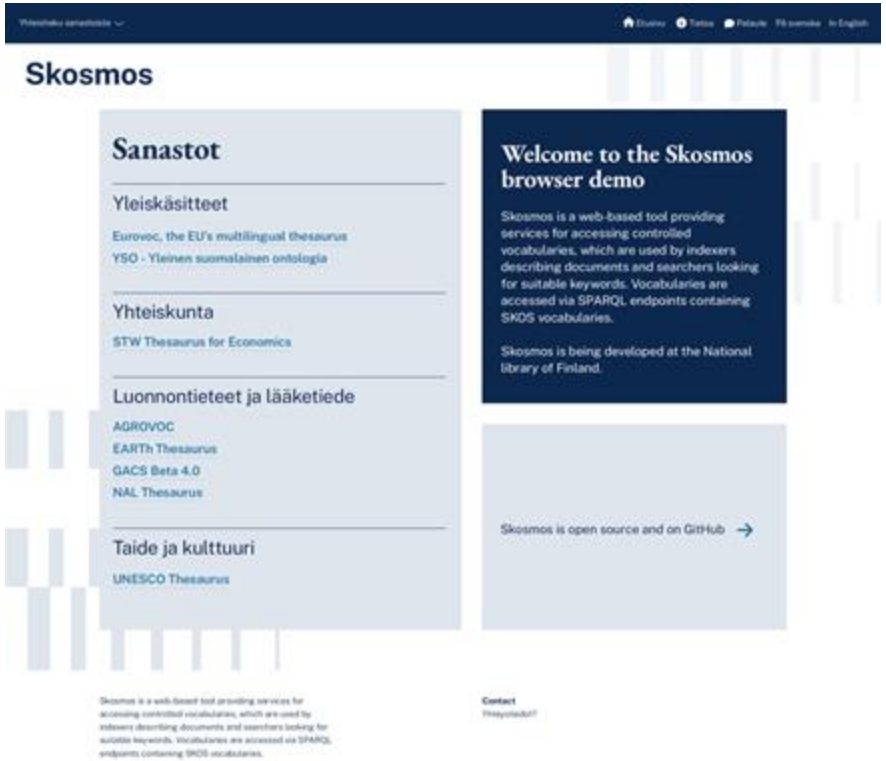
What hasn't changed:

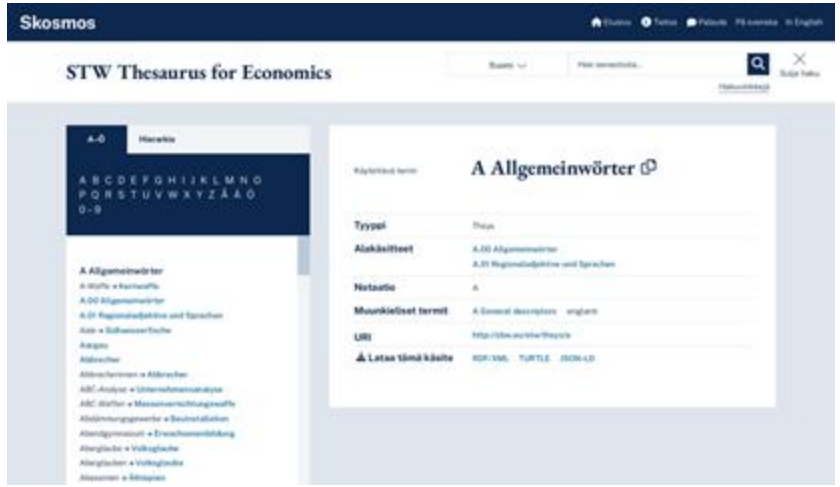
- **core functionality remains the same**
- **architecture remains the same**

New visual style and logo

Skosmos

Skosmos



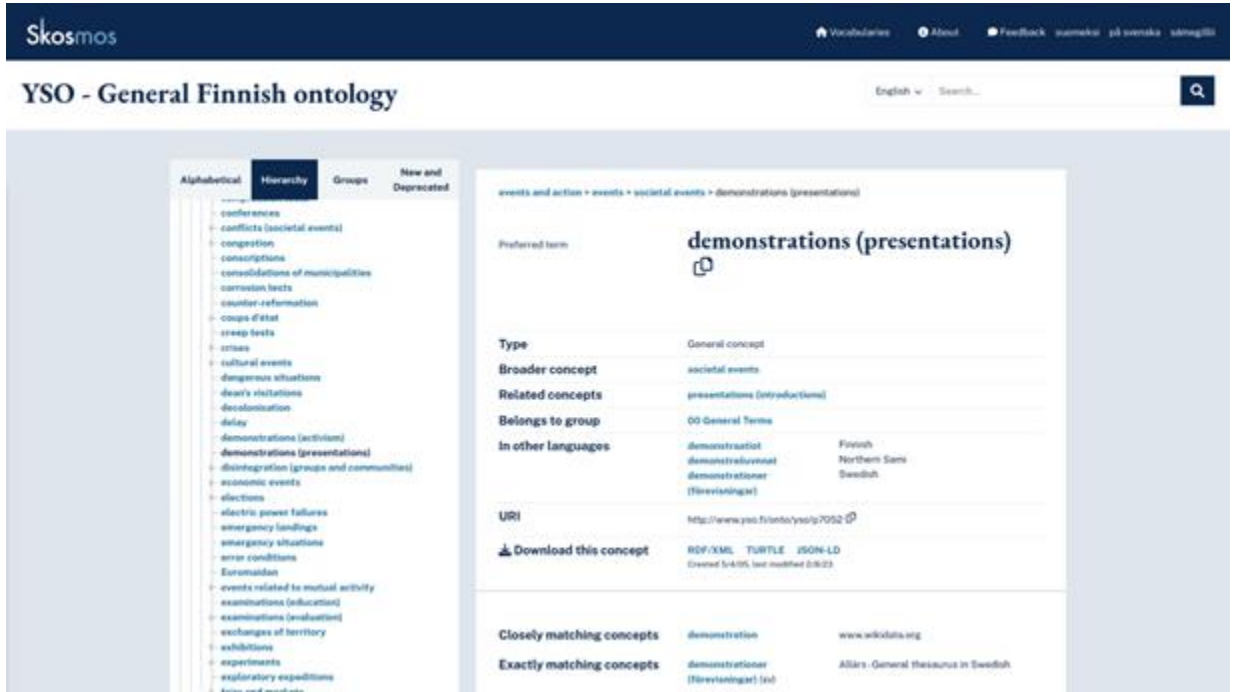


16 40
NATIONAL LIBRARY OF FINLAND

Design by Hahmo Design Oy

Demo time!

Current Skosmos 3 development version (tracking the `main` branch) always available at https://test.dev.finto.fi



16 40
NATIONAL LIBRARY OF FINLAND

Localization: Skosmos UI in 21 different languages

- English is the base/default language
- Finnish, Swedish and Northern Sámi translations maintained by NLF
- translations for other languages crowdsourced using Transifex (so far)

For Skosmos 3, we're setting up **Lokalise** for managing translations.

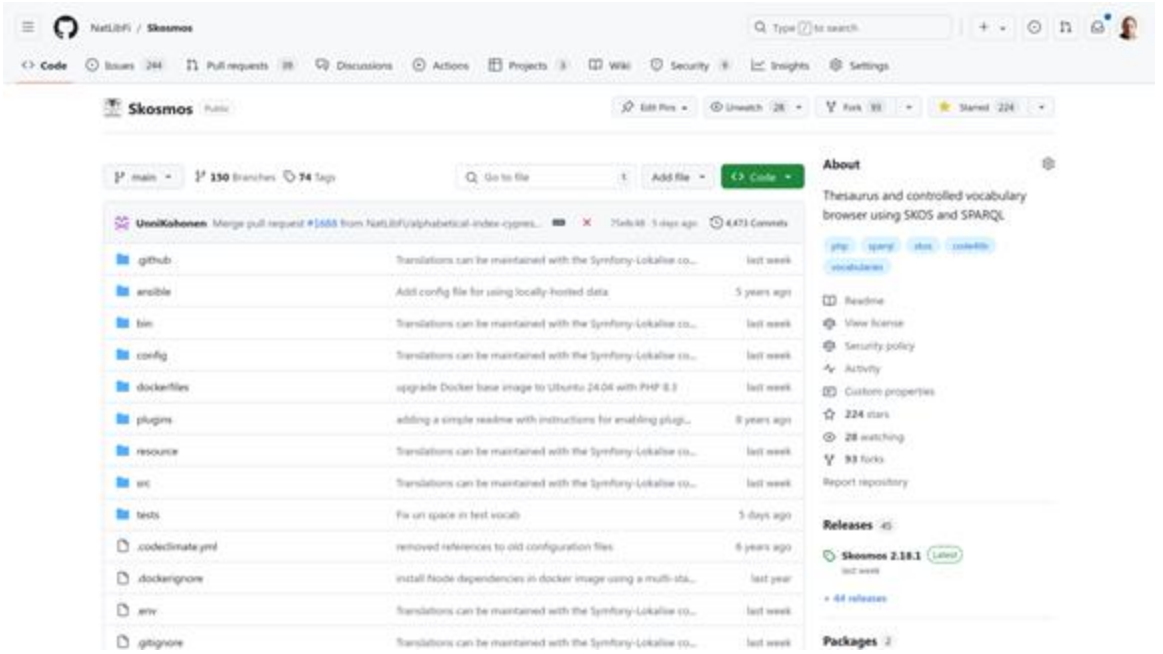
Skosmos on GitHub

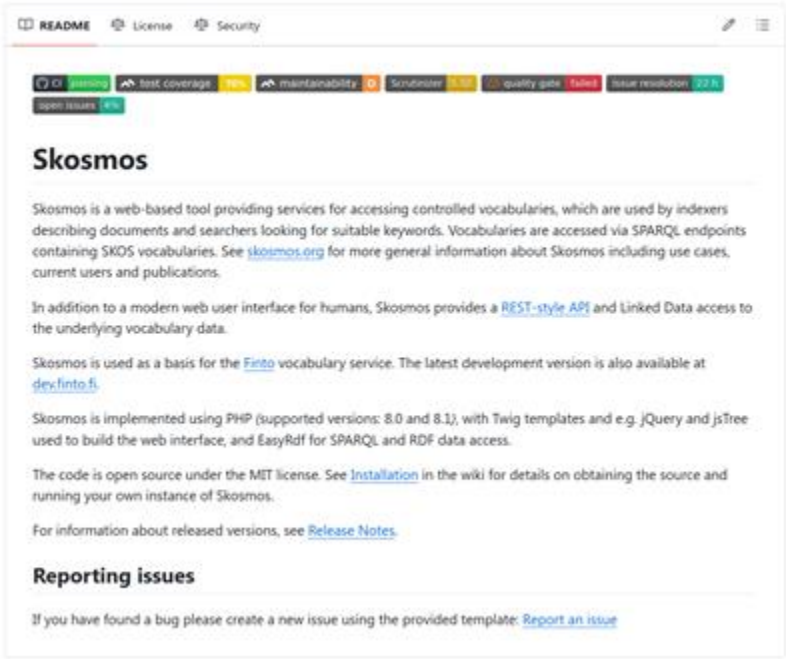
Code repository

- `main` branch is Skosmos 3
- `skosmos-2` branch is Skosmos 2

Issues and **pull requests**

Documentation in the wiki





NATIONAL LIBRARY OF FINLAND

https://github.com/NatLibFi/Skosmos

Development of Skosmos 3

- Most work is done in Scrum-style **sprints**
  - 2 weeks (only Tuesdays to Thursdays) per sprint + planning & review
  - five sprints in 2023, six in 2024, six in 2025 (two to go)
- Development **team**
  - currently 4 developers (Joeli, Mika, Osma, Unni)
  - occasional consultants (Bruno, Henri)
  - product owner (previously Mikko, now Mona)
  - user representatives (Mirja, Tuomas, Maria) in sprint planning & reviews
- Coordinated using GitHub issues and project boards

Roadmap 2025-2026

3.0-alpha.1	Released in March 2025
3.0-alpha.2	Released in May 2025
3.0-beta.1	Released in August 2025
3.0-beta.2	To be released in November 2025
3.0	To be released in December 2025, with core functionality
2026:	...more functionality...
	Customization and plugins needed for Finto.fi
3.1+	Ready to be used in Finto.fi

What does it take to deploy Skosmos?

Hardware & Software:

- a Linux server with modest specifications
- basic web server software: Apache, PHP
- Apache Jena Fuseki triple store (open source)
- git version control utilities

Data:

- one or more controlled vocabularies as SKOS/RDF files

Skills:

- Installing & maintaining the above
- Basic RDF skills, e.g. loading SKOS data into the triple store
- optional: web skills (e.g. CSS) for customizing Skosmos
- being able to read our excellent wiki documentation ;)

Skosmos-users Google Group

Welcome to the Skosmos users' mailing list / web forum! This list can be used for

- general discussion about Skosmos, its features and usage scenarios
- asking for help with installing or running Skosmos
- future directions for Skosmos
- announcements for new versions and other Skosmos-related news

Osma Suominen	**ANN: Skosmos 3.0-beta.1 released** — Skosmos 3.0-beta.2 has been released! https://github.com/NatLibFi/S…	28.8.	☆
Natuurfilmpjes	**Comparing skosmos library against other libraries** — Hi, I am using the Skosmos REST API to get vocabulari…	7.8.	☆
Omid Ghiasvand	**Translations not passed** — Hello, I am using the old interface of Skosmos. Everything works fine, but the prob…	20.7.	☆
Osma Suominen	**ANN: Skosmos 3.0-alpha.2** — Skosmos 3.0-alpha.2 has been released! https://github.com/NatLibFi/Skosmo…	26.5.	☆
Osma Suominen	**ANN: Skosmos 3.0-alpha.1** — Skosmos 3.0-alpha.1 has been released! https://github.com/NatLibFi/Skosmo…	20.3.	☆
Osma Suomin… , … Joeli Taka… 3	**Re: Custom properties - not able to display** — Hello Uldis, No not really. Skosmos won't show the hierarchy if …	5.3.	☆
Osma Suominen	**Apologies for the spam** — Hi all, apologies for the spam messages that were posted through skosmos-users …	24.1.	☆
opo...@gmail.… , Osma Suomi… 2	**Hierarchy disappears when groups are introduced** — Hi Oddrun, can you please show the configuration you u…	23.1.	☆

Skosmos online workshop at SWIB25

"From chaos to Skosmos: publishing controlled vocabularies made easy"

Monday November 17th, 9:00–13:00 CET (probably a bit shorter)

More information and registration:

https://forum.swib.org/t/from-chaos-to-skosmos-publishing-controlled-vocabularies-made-easy/1454

The target audience of this workshop are people working with vocabulary services and knowledge organization systems such as thesauri, classifications, or ontologies. We welcome current Skosmos users as well as those who would be interested in taking Skosmos into use.

Skosmos

Questions about Skosmos?

before we move on to Annif

Part II:
Annif and Finto AI: automated subject indexing and classification

Mona Lehtinen

16 40
NATIONAL LIBRARY OF FINLAND

annif

developed since 2017

- General purpose open source **tool** for automated subject indexing and classification using multiple backends
- Multilingual, supports many vocabularies
- Code on GitHub, website with test form and API (Apache 2.0, except some dependencies)
- Global development and user community; user forum **annif-users** on Google Groups

annif.org

fintoai

launched in 2020

- Automated subject indexing **service** for production use, based on Annif. Web user interface and API service
- Supports indexing with the General Finnish Ontology YSO & PLC - Finnish Public Libraries Classification System (in Fin, Swe & Eng) as well as KAUNO - ontology for fiction (in Finnish)
- Intended to support subject cataloguers in Finland regardless of institution (GLAMs, public administration); sister project to the Finto vocabulary service

ai.finto.fi

16 40 NATIONAL LIBRARY OF FINLAND

Various algorithms

Lexical approaches (MLLM)

Look for matches between terms in text and vocabulary

*"**Renewable resources** are a part of Earth's **natural** environment and the largest components of its ecosphere."*

yso:p14146
"renewable natural resources"

Associative approachest (TF-IDF, fastText, Omikuji)

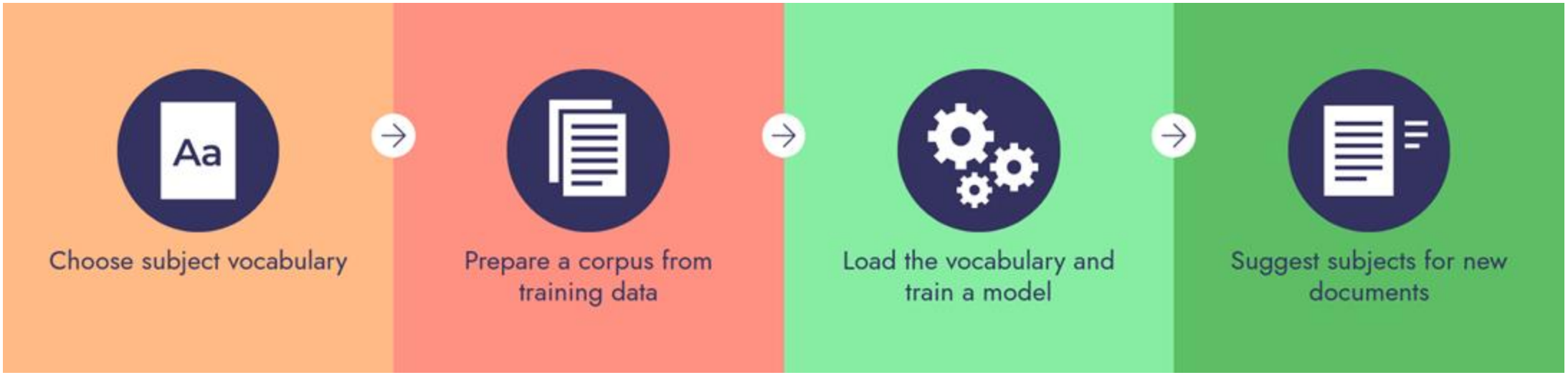
Learn statistical correlations between terms in the text and vocabulary

Less training data needed

More training data needed

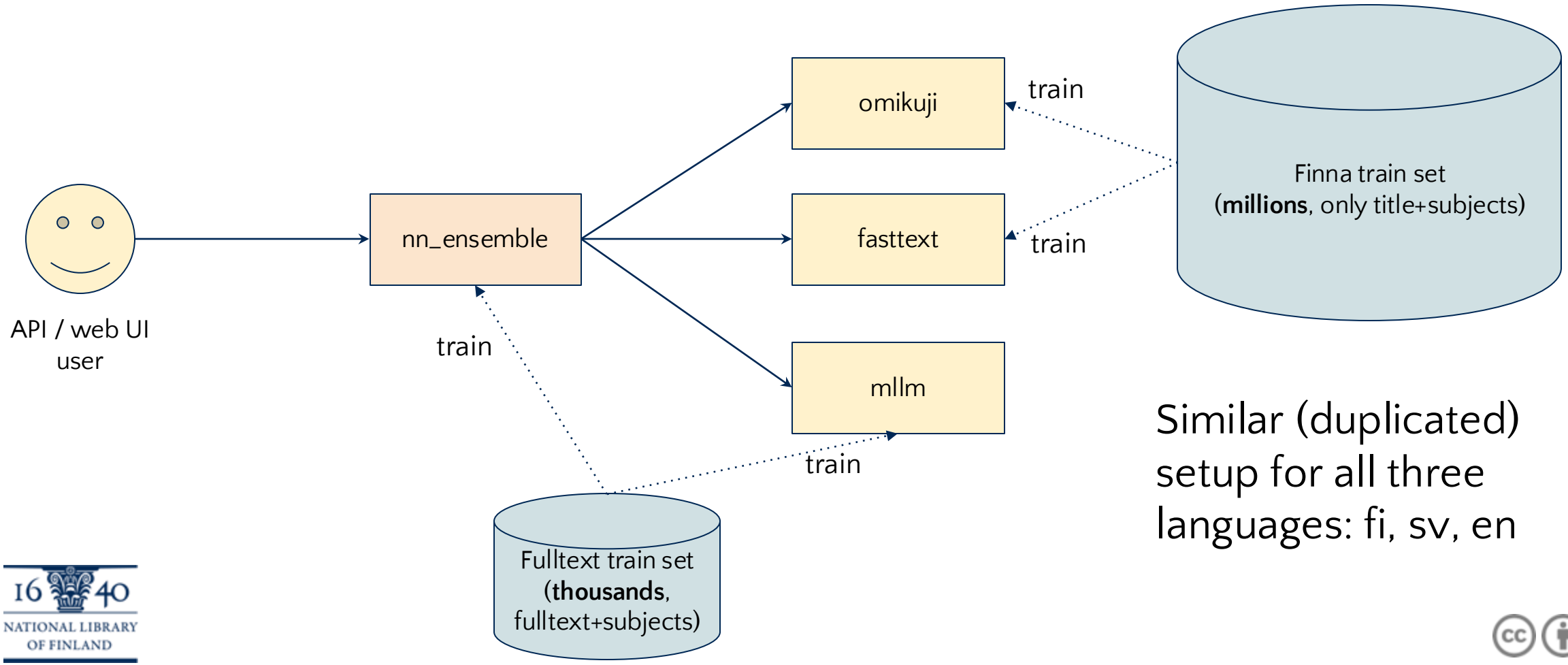
16 40
NATIONAL LIBRARY OF FINLAND

What would it take to use Annif?



Setting up, training and evaluating models

Current Finto AI set of projects/models for YSO subject vocabulary:



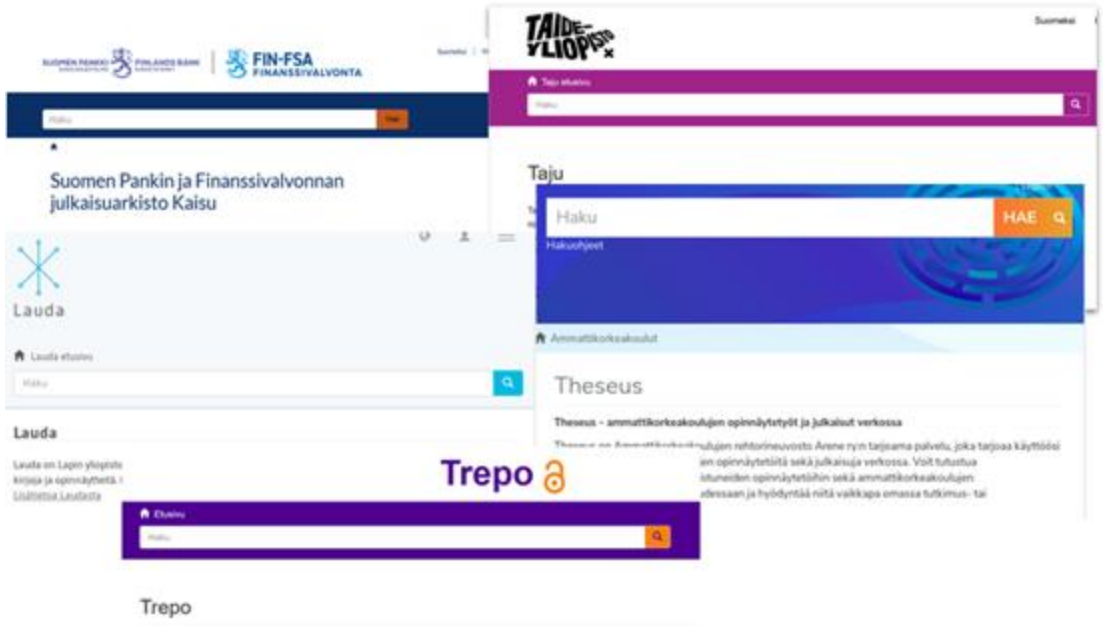
Similar (duplicated) setup for all three languages: fi, sv, en

Annif & Finto AI development in-house

- Within the library, Annif development has ties to
  - Ontologies and vocabularies: Finto
  - Repository services
  - Legal deposits
  - Finna discovery service
  - National Bibliography

finto Finnish Ontology and Vocabulary Service

FINNA.FI



Trepo

16 40 NATIONAL LIBRARY OF FINLAND

Annif & Finto AI development & community

- Work with external partners:
  - EU project High Performance Digitisation (Connecting Europe Facility, 2018–2020) implemented in the supercomputing environment provided by the IT Centre for Science CSC → led to major improvement in Annif
  - ZWB Leibniz Information Centre for Economics: contributions to Annif, Annif-tutorial
  - We also work closely with other cultural heritage and research organisations regarding AI

Annif & Finto AI development & community

- We participated in the LLMs4Subjcets workshop
- Idea: LLMs for efficient subject indexing, competition for different teams
- In the 1st round, Annif was 1st/2nd in quantitative evals, 4th in qualitative eval (of 14 teams)
- In the 2nd round Annif got 1st place (emphasis on resource consumption)
- we used LLMs for **translation** and **synthetic data generation**
- SemEval proceedings; our paper; our poster; our 2nd paper preprint

→ New ideas and lessons for Annif development

Annif & Finto AI community & support

- For hands-on details of working with Annif, see the Annif tutorial
  - The next Annif tutorial / online workshop: SWIB25
    - November 17th 2025, 18:00–22:00 CET
    - more info and registration: https://forum.swib.org/t/introduction-to-the-annif-automated-indexing-tool/1456
  - Materials (videos and exercises) are available for self-study on GitHub and YouTube.

- Annif documentation on GitHub

- Annif-users community: Google forum

Annif-users Google Group

Welcome to the Annif users' mailing list / web forum! This list can be used for

- general discussion about Annif, its features and usage scenarios
- asking for help with installing or running Annif
- future directions for Annif
- announcements for new versions and other Annif-related news

juho.i...@hels... , ... Christoph ... 5	**ANN: Annif 1.4 released** — Dear Annif-team, thank you for releasing the new Annif Version. Including the fully...	4.9.	☆
Annif Users	**Proceedings and Continuation of the LLMs4Subjects Competition** — Dear all, As you may remember, we part...	28.8.	☆
Thomas Guignard	**Annif use cases in Switzerland** — Dear Annif community I'm preparing a short presentation on Annif for the L...	21.8.	☆
Ball, Lakshmi rb 2	**Xtransformer backend** — Hi Lucas, Thanks for testing the Xtransformer backend and reporting this. The issu...	15.8.	☆
Teemu Nuut... , ... Osma Suom... 3	**WSGI service setup fails** — Hi Teemu and Uldis, you are both correct, the wiki page on WSGI is outdated and t...	29.7.	☆
Jan Ja... , ... anna.k...@google... 4	**practice paper 'Automatic Subject Cataloguing at the German National Library'** — Dear João Lima, you might...	29.7.	☆
Robina M, juho.i...@helsinki.fi 2	**XML Support and aksing for best pracitces** — Dear Robina, 1.) Using corpus in XML-format is not possible, y...	16.7.	☆
Sven Sass, juho.i...@helsinki.fi 6	**X-Transformer backend in nn-ensemble** — Hello Juho, just a short follow up - it might be interesting for every...	15.7.	☆

Annif Users

See https://annif.org for more

Use case: YLE

- Suominen, O. & Virtanen, P. Yle meets Annif – an open source tool for automated subject indexing. Presentation at EBU MDN Workshop 2020, 10 June 2020.
- Operates with own installation & vocabulary
  - Weekly updates
- Integrated into workflow in which a journalist submits an article
- Commisioning project took 1-2 years (with a small project group)
- News item: https://yle.fi/aihe/a/20-10001817

Use cases & contributors: ZBW and DNB

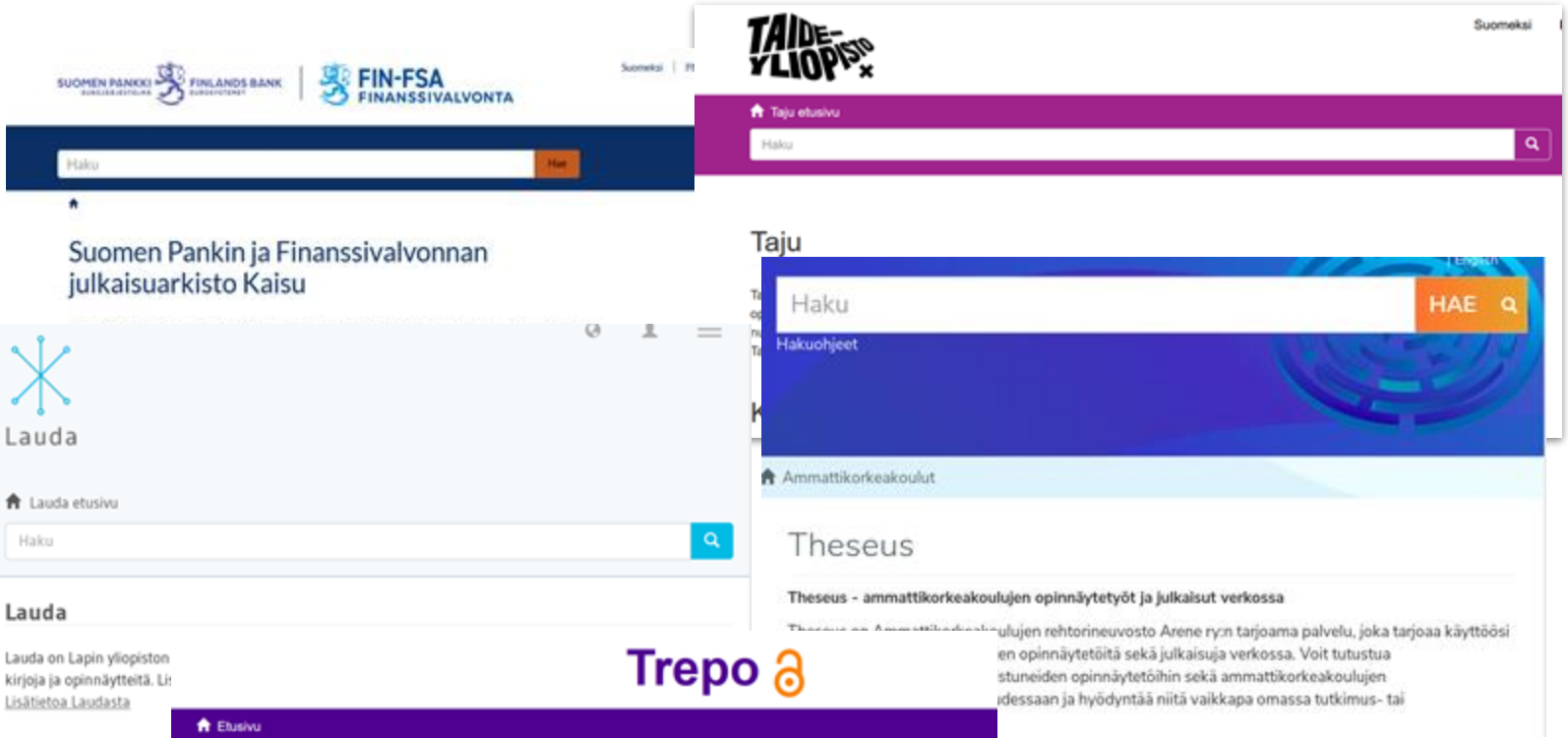
- Kasprzik, A. (2025). Transferring applied machine learning research into subject indexing practice. In E. Balnaves, L. Bultrini, A. Cox, & R. Uzwyshyn (Eds.), *New horizons in artificial intelligence in libraries* (pp. 199–212). De Gruyter Saur. https://doi.org/10.1515/9783111336435-015
- Kasprzik, A. (2023). Automating subject indexing at ZBW – making research results stick in practice. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, 33(1). https://doi.org/10.53377/lq.13579
- Team page
- DNB launches it's cataloguing machine ("Erschließungsmaschine")
- Poley, C., Uhlmann, S., Busse, F., Jacobs, J.-H., Kähler, M., Nagelschmidt, M., & Schumacher, M. (2025). Automatic Subject Cataloguing at the German National Library. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, *35*(1), 1–29. https://doi.org/10.53377/lq.19422
- Kluge, L., & Kähler, M. (2025). DNB-AI-Project at SemEval-2025 Task 5: An LLM-Ensemble Approach for Automated Subject Indexing. arXiv. https://doi.org/10.48550/arXiv.2504.21589

Use case: DESKRYPTOR / National Library of Poland

https://deskryptor.bn.org.pl/tag

16 40 NATIONAL LIBRARY OF FINLAND

Finto AI at NatLibFi

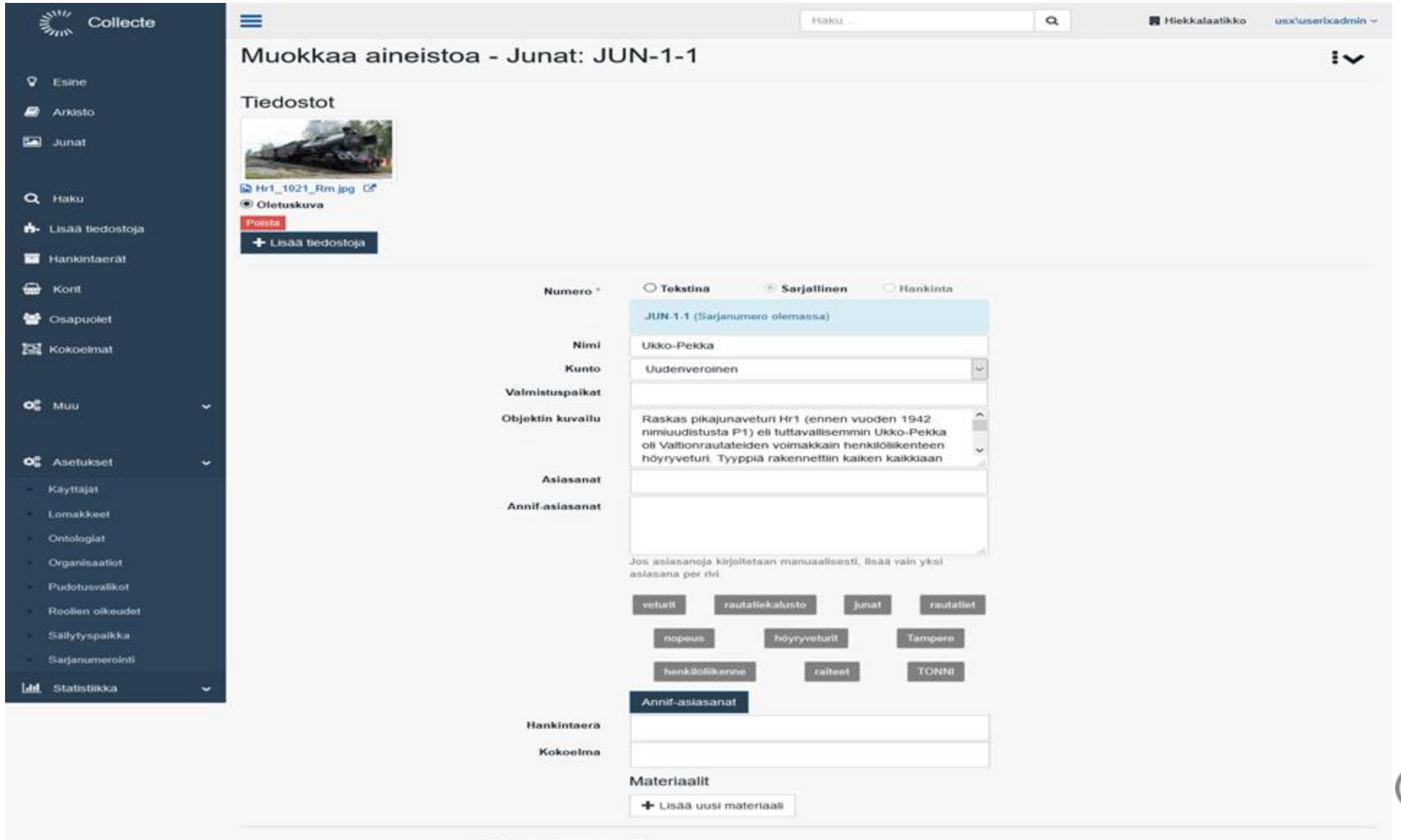


Legal deposits

16 40 NATIONAL LIBRARY OF FINLAND

DSpace-based institutional repositories: Osuva, Trepo, Theseus, Taju, Lauda, Kaisu (Demo of use)

Use case: Finto AI at Finnish museums



16 40
NATIONAL LIBRARY OF FINLAND

Annif users survey

- Started in Nov 2024, 20 responding organizations
- The overall satisfaction with Annif was 4.4 out of 5
- Satisfaction to documentation ganed the highest score of 4.5
- Satisfaction to suggestion quality was 3.6
- Major challenge: gathering and preparing data for training and evaluating
- Inkinen, J., Lehtinen, M., & Suominen, O. (2025). *Annif users survey: Understanding usage and challenges*. Kansalliskirjasto. https://urn.fi/URN:ISBN:978-952-84-1301-1

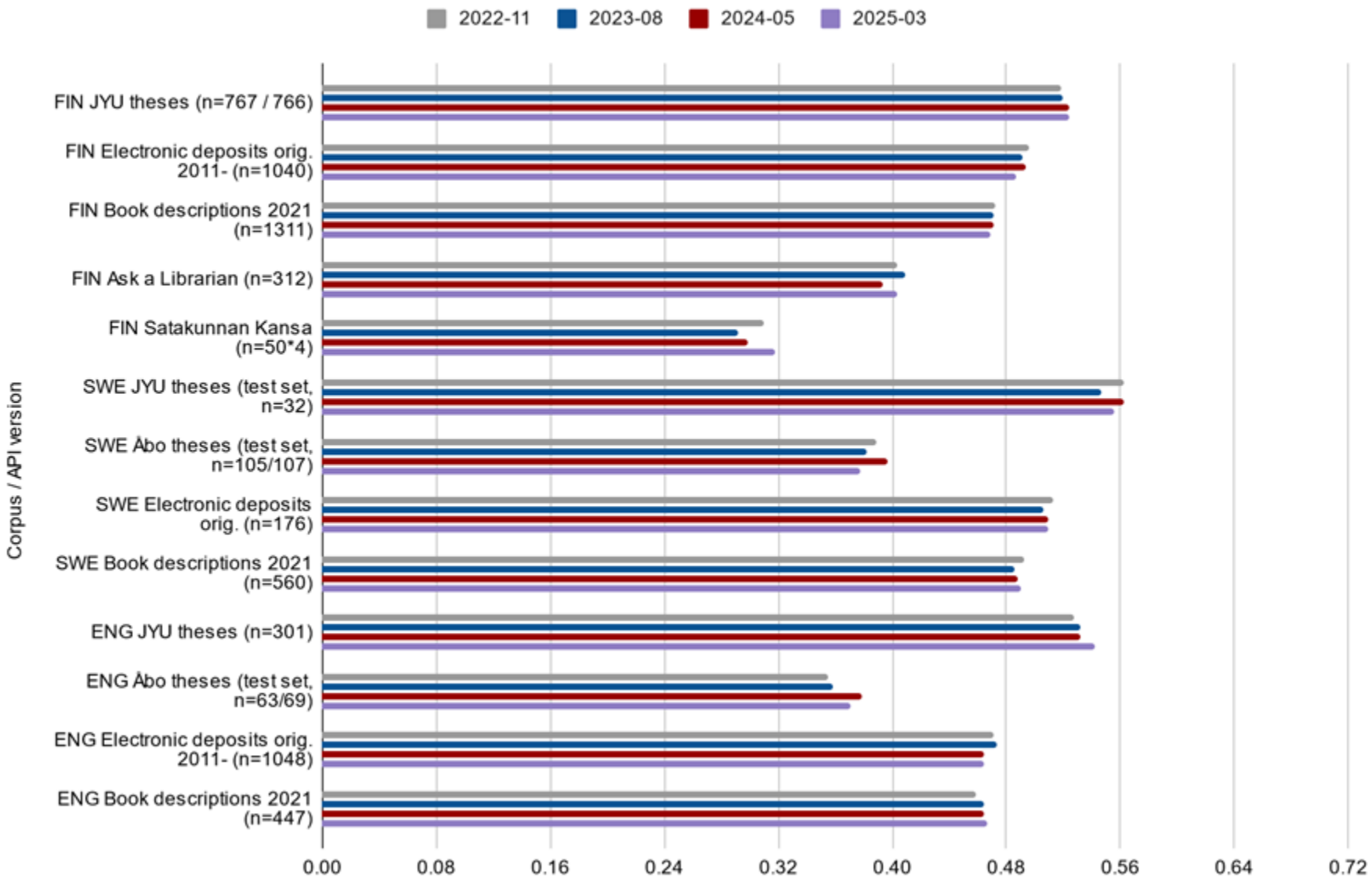
Evaluation approaches (Golub et al. 2016), **emphasis** ours

1. Evaluating indexing quality directly through **assessment by an evaluator** or by **comparison with a gold standard**.
2. Evaluating indexing quality directly **in the context of an indexing workflow**.
3. Evaluating indexing quality indirectly through retrieval performance.

The different evaluation approaches are complementary.
Not a good idea to look at just a single measure.

Golub, K., Soergel, D., Buchanan, G., Tudhope, D., Hiom, D., and Lykke, M. 2016. A framework for evaluating automatic indexing or classification in the context of retrieval. Journal of the Association for Information Science and Technology, 67(1): 3–16.

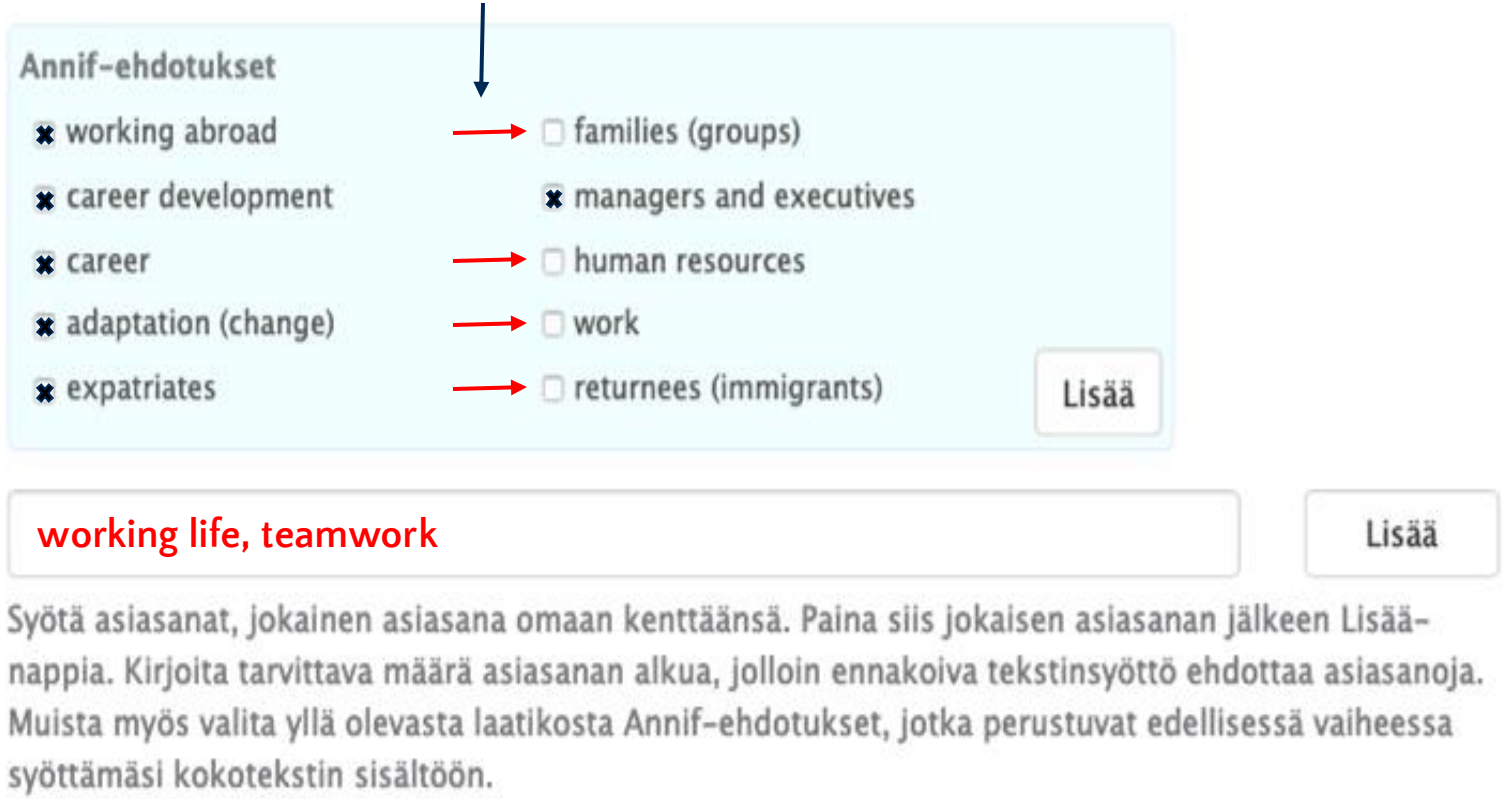
F1@5 scores



Evaluation in the workflow context: Theseus

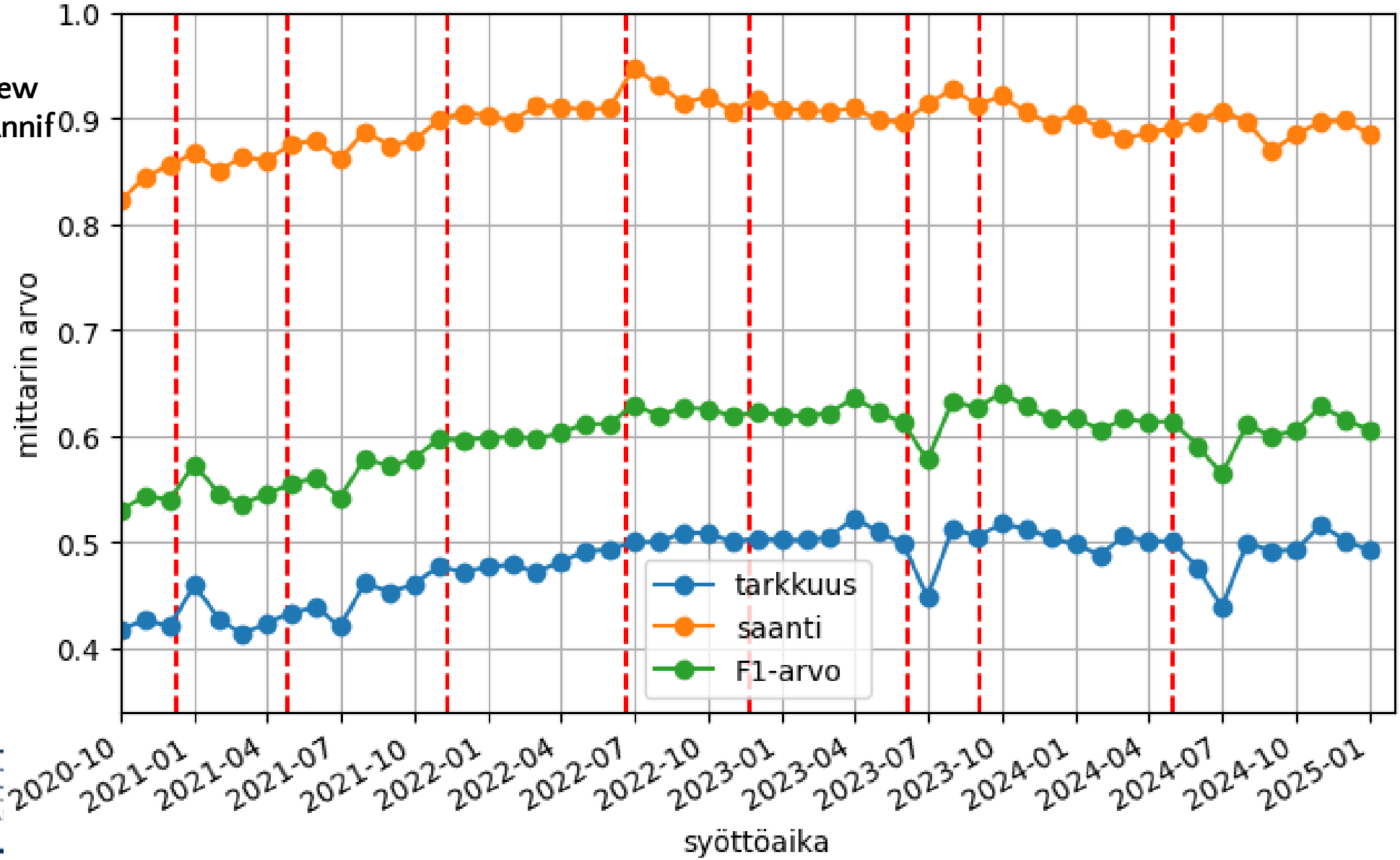
Wrong / irrelevant suggestions can be left out

Missing relevant terms can be added



Evaluation in the workflow context: Theseus

Red lines: new versions of Annif



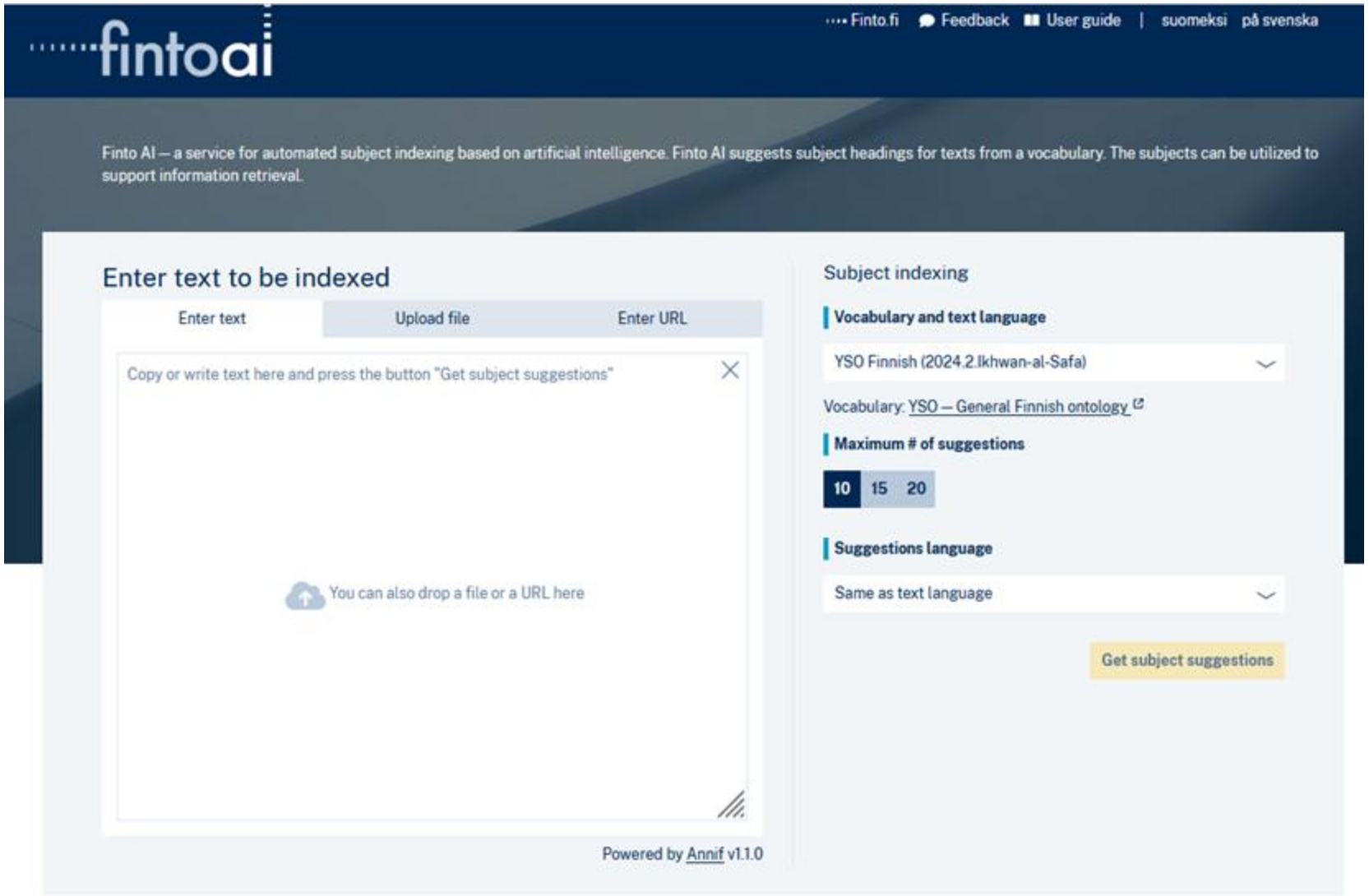
Blue: precision

Orange: recall

Green: F1 score

Try Finto AI!

https://ai.finto.fi



Lessons learned

- Implementing semi-automated indexing with an API service is easy; explaining it to users can be more challenging
- Community & collaboration are key (engage with end-users and other developers)
- Traditional ML approaches are still valid

Thank you!

Osma Suominen, Joeli Pokkinen, Mona Lehtinen, Juho Inkinen

[firstname.lastname]@helsinki.fi