

# Finto AI, a service for automated subject indexing

Finto AI is a service for automated subject indexing. It can be used to suggest subjects for text in Finnish, Swedish and English. It currently gives suggestions based e.g. on concepts of the General Finnish Ontology YSO.

## Web Interface

The screenshot shows the Finto AI web interface. At the top, there's a navigation bar with links to Finto.fi, Feedback, User guide, and language options (suomeksi, på svenska). Below the navigation bar, a banner explains the service: "Finto AI — a service for automated subject indexing based on artificial intelligence. Finto AI suggests subject headings for texts from a vocabulary. The subjects can be utilized to support information retrieval."

The main interface is divided into two columns. The left column is titled "Enter text to be indexed" and contains three tabs: "Enter text", "Upload file", and "Enter URL". The "Enter text" tab is active, showing a large text area with a sample text: "The purpose of the subject cataloging is to express the topic of the material (e.g. book, article, website, object, etc.) to facilitate finding the material. For subject cataloging, you can use e.g. classification systems, thesauri or free keywords. It is traditionally expert work done by professionals in libraries, archives and similar organizations. Authors also often describe the content of their own work in some form." Below the text area, it says "Powered by Annif v1.0.0".

The right column is titled "Subject indexing" and contains several settings: "Vocabulary and text language" set to "YSO English (2023.6.Hypatia)", "Vocabulary" set to "YSO — General Finnish ontology", "Maximum # of suggestions" set to 10, and "Suggestions language" set to "Same as text language". A yellow button labeled "Get subject suggestions" is located below these settings.

Below the button, there is a section titled "Suggestions" with a "Copy" button. It lists four suggestions: "subject cataloging", "cataloguing", "thesauri", and "libraries". Each suggestion has buttons for "TERM", "URI", and a clipboard icon.

Finto AI can be used via the form at [ai.finto.fi](https://ai.finto.fi). You can use the form by copy-pasting text to the large text field and then clicking the button "Get subject suggestions". In the drop-down menu you can choose the language of the text. You can also set the maximum number of suggestions you would like to receive. You can also copy the suggestion (as a word, URI or record) to the clipboard using a button.

## REST-style API

Finto AI also has an API, which makes it easy to integrate with other systems. More information of the API can be found from this wiki on the page [Finto AI open API service](https://ai.finto.fi/v1/ui/), a detailed OpenAPI/Swagger technical documentation is available at <https://ai.finto.fi/v1/ui/>.

An API integration is already in place at the University of Jyväskylä : students submitting their Master's thesis to the JYX repository get suggestions from Annif that they can use or discard, then a librarian/informatician does a final check. A similar workflow is being piloted in the Osuva repository of the University of Vaasa.

## Vocabularies and Languages

Finto AI currently uses the latest version of the General Finnish Ontology including place names (YSO Places) and the Finnish Public Libraries Classification System in three languages (Finnish, Swedish and English) as well as KAUNO (an ontology for fiction in Finnish). We are planning to expand the choice of subject vocabularies and languages in the future.

In the development process we have discovered and tested several algorithms, and selected the currently best combination for Finto AI (MLLM, Omikuji and a TensorFlow based NN ensemble). The algorithms have been trained primarily with metadata from the Finna discovery service, but full text documents have also been used for fine-tuning the models. The development work of is ongoing and we will offer updates and improvements to Finto AI accordingly.

## From Annif API to Finto AI, the Production Version

Finto AI is based on Annif, a tool for automated subject indexing. You can read more about using Annif in it's [GitHub Wiki](#). To work, Annif needs a controlled vocabulary (subject headings, thesaurus or classification) and existing metadata - Annif can then be used to assign subjects for new documents. This tool is built upon a combination of existing natural language processing and machine learning tools including e.g. [Omikuji](#), [fastText](#) and [Gensim](#). It is designed to be multilingual and it can support any subject vocabulary (in SKOS or a simple TSV format). It can be used either via a command-line interface or a microservice-style REST API. In fact, this demo API under [Api.annif.org](#) and the demo form at [annif.org](#) are the basis of Finto AI. As a development tool, Annif offers more methods than Finto AI, successful features will be integrated to Finto AI in time.