



Account ▼

<https://readcoop.eu>[Transkribus \(/transkribus/\) >](/transkribus/)[How-To Guides \(https://readcoop.eu/transkribus/resources/how-to-guides/\) >](https://readcoop.eu/transkribus/resources/how-to-guides/)

How to Search Documents with Smart Search

How to Search Documents with Smart Search

[💡 Transkribus Tools](#) [📄 Transkribus Lite](#)*Last update 8 months ago*[« How-To's overview \(https://readcoop.eu/transkribus/resources/how-to-guides/\)](https://readcoop.eu/transkribus/resources/how-to-guides/)

► About Transkribus ()

Table of Contents



How to Search Documents with Smart Search

Introduction

[Smart Search in Transkribus Lite](#)

Preparation: Text Recognition

Using Smart Search

Difference between Smart Search and Fuzzy Search

Introduction

Smart Search enables to perform a more advanced and powerful type of search on the documents recognised using a PyLaia HTR (<https://readcoop.eu/glossary/pylaia/>) model. This feature enables you to search the automatically generated transcriptions more accurately, without the need to correct them manually. It can prove very useful with, but not limited to, records and registers.

While the standard search goes through the transcription as it appears in the text editor, Smart Search also looks to several alternatives for each recognised word of the automatic transcription. The alternatives do not appear in the text editor but have been stored in addition to the transcription. Using Smart Search, you can find words even if they have been transcribed incorrectly by the HTR model. Thus, Smart Search can produce valuable results also with automatic transcriptions with a high error rate (CER up to 30%).

Smart Search in Transkribus Lite

Preparation: Text Recognition

To use the Smart Search feature, you need first to enable it at the time of Text Recognition, so that all the possible alternatives of the words are stored and can be browsed any time you launch a search.

Select the pages or the document you want to recognise automatically and click on the "Text Recognition" button on the left-side "Actions" menu. Then, select the most suitable PyLaia model for your documents. Only if you have chosen a PyLaia model, the "Smart Search" checkbox will appear above the green "Start Recognition" button (Fig. 1).

Tick the checkbox to enable Smart Search, i.e. to save not only the best match but also the other alternatives on which the HTR model is less confident in their correctness. By default, 100 is the maximum number of variants taken into account and stored.

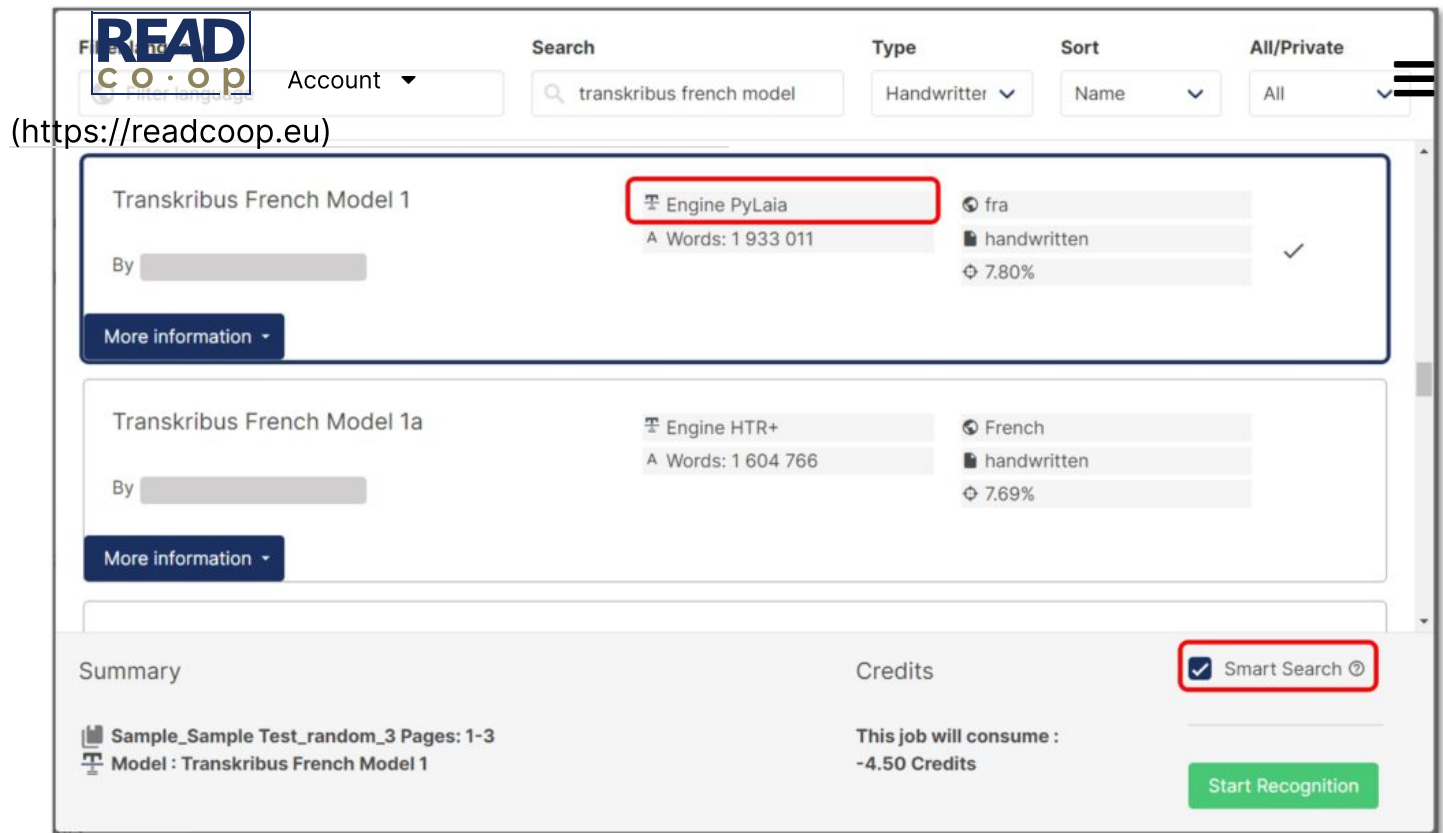


Figure 1. Smart Search checkbox in Transkribus Lite

Since generating the Smart Search data during the text recognition is an intense computing task and requires additional storage space (10x more than normal), a 50% credits surcharge is applied. This means that instead of consuming 1 credit per page, as is usually the case with PyLaia HTR models, you will consume 1.5 credits per page.

Before launching the text recognition, you must therefore evaluate whether the Smart Search feature is helpful for your documents, depending on how you intend to use the HTR transcripts. If you want to make Smart Search available at a later stage, you will have to relaunch the text recognition on all pages, which consumes more credits than implementing Smart Search in the first place.

Using Smart Search

When the Text Recognition is finished, you can search your pages by using the search bar in the top right of Transkribus Lite. At this stage, you do not have to select any options: just type the term and launch the search. Automatically, the search goes through both the words appearing in the text editor and all the saved alternatives.

Clicking on the result will load the page where it has been found. When the term has been found among the variants, it appears correctly in the search result list (Fig. 2). However, when you open the corresponding page, you will notice that the transcription contains a

different word, i.e. the word that the model rated as best during the recognition. Looking at the highlighted word in the image, you will see if the variant found by Smart Search is the correct transcription (very likely) or if it is an incorrect guess (Fig. 3). (<https://readcoop.eu>)

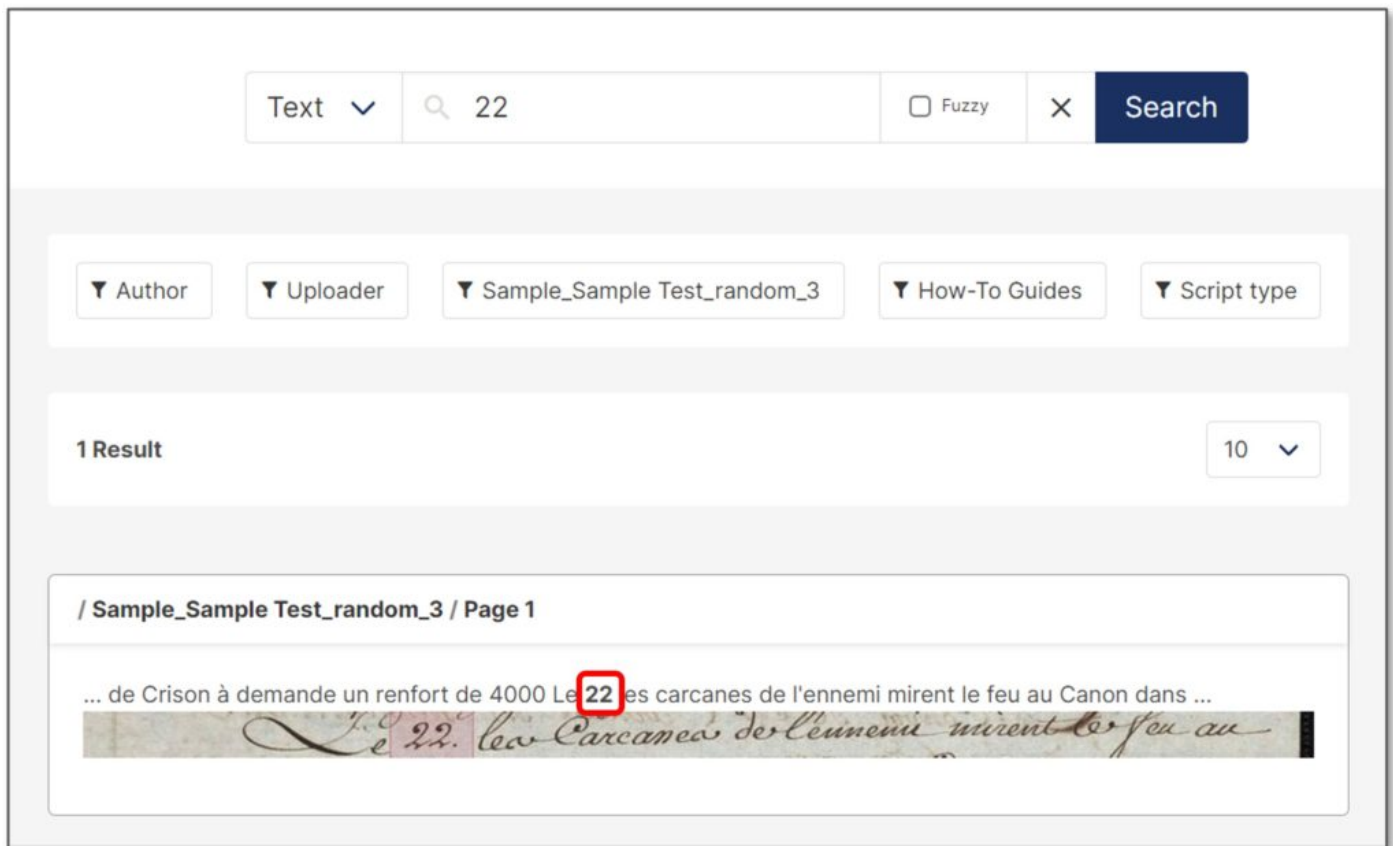


Figure 2. Search results in Transkribus Lite

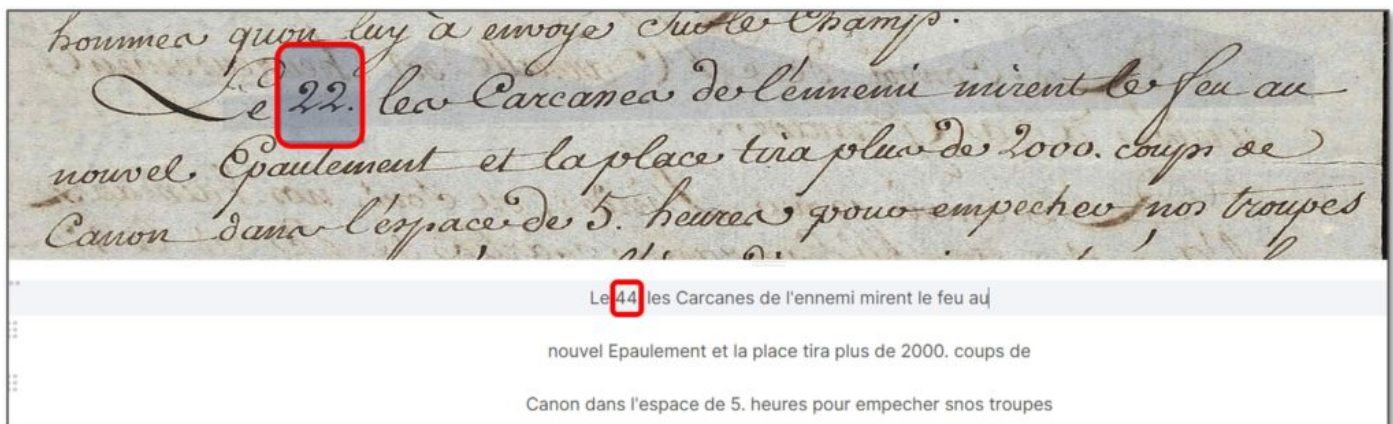


Figure 3. Correct result found thanks to Smart Search

Difference between Smart Search and Fuzzy Search

(<https://readcoop.eu>)

Fuzzy Search (<https://readcoop.eu/glossary/fuzzy-search/>) and Smart Search may seem similar, but in reality, the techniques behind them are different, thus also the results. Fuzzy Search allows one to find approximately matching words in addition to exact matches: it is useful with misspellings and spelling variations. However, Fuzzy Search only goes through the words in the transcription and retrieves results that differ by only one or two letters from the searched term. On the contrary, Smart Search performs the search among both the transcribed words and the several less confident variants, which may differ considerably from the word accepted as the best match.

For instance, suppose to search for the name “Tommaso” in a transcription generated by a PyLaia HTR model with a 20% CER. Fuzzy Search returns two results: it correctly detects the name also when the HTR model transcribes it as “Sommaso” (one letter incorrect) and “Sommato” (two letters incorrect). On the other hand, Smart Search browses both the best matches and their alternatives and finds a new occurrence of “Tommaso,” which appears as “Dominato” in the transcription. Fuzzy Search is not able to find this result because “Dominato” is too different from “Tommaso”; Smart Search, on the contrary, succeeds in finding it because “Tommaso” is one of the less confident variants recognised by the HTR model and stored at the time of the text recognition.

It is also possible to select both Smart Search and Fuzzy Search to combine both and do a fuzzy search within the variants.

The COOP

About us (<https://readcoop.eu/about/>)

Join us! (<https://readcoop.eu/join/>)

Our Members (<https://readcoop.eu/members/>)

Success Stories (<https://readcoop.eu/success-stories/>)

Work with us (<https://readcoop.eu/work-with-us/>)

Products & Services

Transkribus (<https://readcoop.eu/transkribus/>)

Transkribus lite (<https://transkribus.eu/lite>)

(<https://readcoop.eu>)

Read&search (<https://readcoop.eu/readsearch/>)

ScanTent (<https://readcoop.eu/scantent/>)



Account ▼



Useful information

News (<https://readcoop.eu/news/>)

Download Transkribus (<https://readcoop.eu/transkribus/download/>)

Public Models (<https://readcoop.eu/transkribus/public-models/>)

Payment and shipping (<https://readcoop.eu/payment-and-shipping/>)

Helpful resources

Resource center (<https://readcoop.eu/transkribus/resources/>)

How-to Guides (<https://readcoop.eu/transkribus/resources/how-to-guides/>)

Getting started with Transkribus (<https://readcoop.eu/transkribus/start/>)

FAQs (<https://readcoop.eu/transkribus/questions/>)

Videos (<https://readcoop.eu/transkribus/resources/videos/>)



Copyright © 2023 READ-COOP SCE

[Terms & conditions \(https://readcoop.eu/terms-and-conditions/\)](https://readcoop.eu/terms-and-conditions/)

[Privacy Policy \(https://readcoop.eu/privacy-policy/\)](https://readcoop.eu/privacy-policy/)

[Contact \(https://readcoop.eu/contact/\)](https://readcoop.eu/contact/)

[Imprint \(https://readcoop.eu/imprint/\)](https://readcoop.eu/imprint/)



(<https://readcoop.eu>)