

# Transkribus med The Swedish Lion

Riksarkivet har presenterat en AI-modell som kan tolka handskrivna texter på svenska från 1600-, 1700- och 1800-talen och uppnå en imponerande noggrannhetsgrad på 95%.

Den modellen finns tillgänglig via verktyget Transkribus.

Läs här om hur du använder Transkribus och får hjälp med dina svårtolkade texter.

[Transkribus](#) är en webbapplikation skapad av en österrikisk Kooperation “Read-Coop” med över 150 medlemmar såsom universitet, arkivinstitutioner, bibliotek med flera från över 30 olika länder.

Transkribus webbtjänst är under utveckling och fungerar för närvarande bäst i Google Chrome, men Windows Edge eller Macens standardwebbläsare går också bra.

**HTR** - Handwritten text recognition fritt översatt handskrivna texttolkning

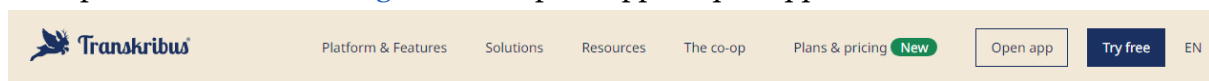
**Model** - Ett program som har tränats på en uppsättning data för att känna igen vissa mönster eller fatta vissa beslut utan ytterligare mänsklig inblandning. I det här fallet har det tränats på dokumentbilder.

Riksarkivet i Sverige har skapat modellen “The Swedish Lion” den AI-modellen kan tolka handskrivna texter på svenska från 1600-, 1700- och 1800-talen och uppnå en imponerande noggrannhetsgrad på 95%. Den har tränats med hjälp av data från texter som omfattar 3,3 miljoner textrader och totalt 15,6 miljoner ord.

Läs mer om Riksarkivets arbete [Ny banbrytande AI-modell för svenska historiska texter](#) i pressmeddelande från 2024-02-07

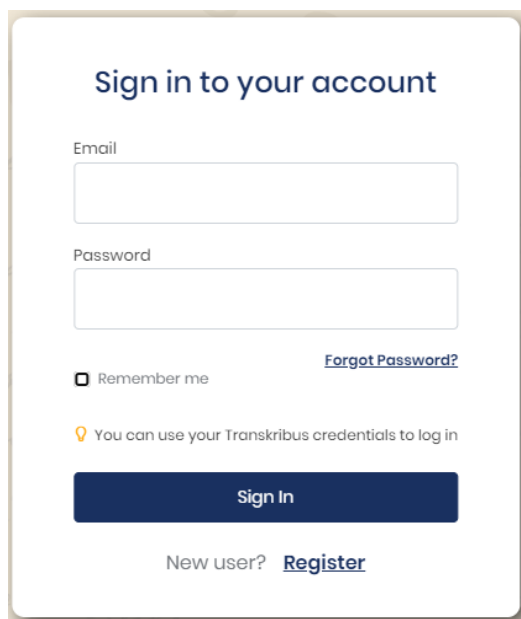
## Skapa ett konto i Transkribus

Gå in på [www.transkribus.org](http://www.transkribus.org) och klicka på knappen *Open app*.



Därifrån kan du skapa ett nytt konto.

Du klickar på länken Register längst ned i formuläret **Sign in to your account**.


A screenshot of a sign-in form titled "Sign in to your account". It contains two input fields for "Email" and "Password". Below the password field is a checkbox for "Remember me" and a link for "Forgot Password?". A note indicates that Transkribus credentials can be used for login. A dark blue "Sign In" button is positioned below the form. At the bottom, there is a link for "New user? Register".

Sign in to your account

Email

Password

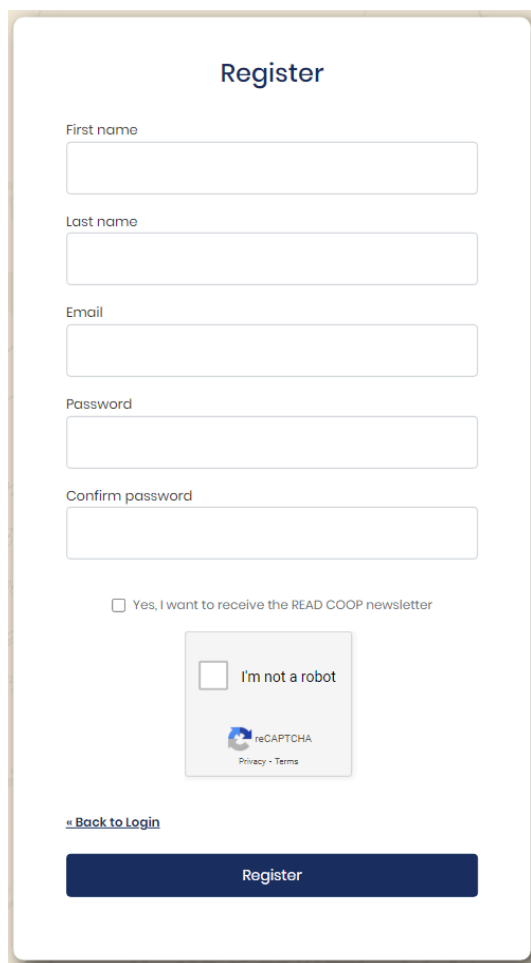
Remember me [Forgot Password?](#)

 You can use your Transkribus credentials to log in

Sign In

New user? [Register](#)

Därefter fyller du i formuläret **Register** för att registrera dig som användare i Transkribus.

A screenshot of a web registration form titled "Register". The form contains several input fields: "First name", "Last name", "Email", "Password", and "Confirm password". Below these fields is a checkbox labeled "Yes, I want to receive the READ COOP newsletter". Underneath is a reCAPTCHA widget with the text "I'm not a robot" and a small image of a robot. At the bottom left of the form is a link that says "Back to Login". At the bottom center is a dark blue button with the text "Register".

Hur man skapar ett konto och kommer igång finns också beskrivet här: <https://help.transkribus.org> som också innehåller mer hjälp till hur du hanterar Transkribus.

Det kostar inget att skapa ett eget konto, ladda upp bilder och använda Transkribus för manuell redigering. Däremot tar Transkribus betalt med så kallade "krediter" när man använder en HTR-modell för att maskinellt texttolka en bild. Varje användare får 100 krediter gratis per månad.

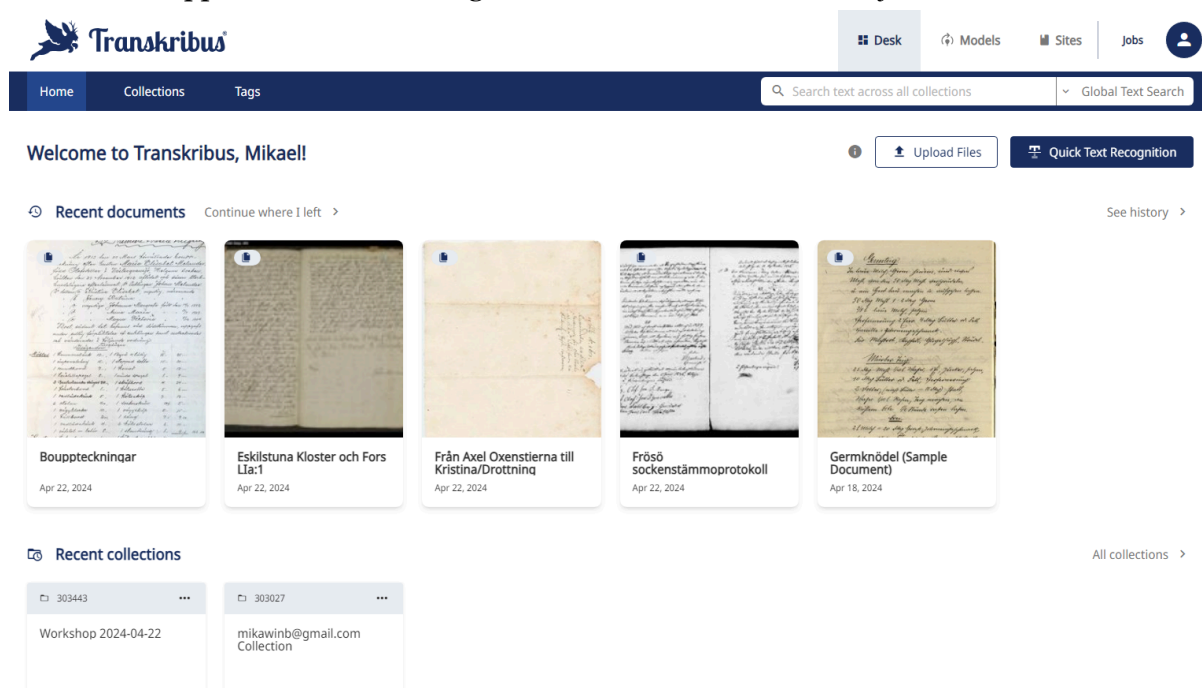
Behöver man mer går det att köpa till fler krediter.

[app.transkribus.org](https://app.transkribus.org)

Du arbetar sedan i applikationen [app.transkribus.org](https://app.transkribus.org) som du antingen når via samma knapp Open app på Transkribus startside eller direkt med länken [app.transkribus.org](https://app.transkribus.org)

Du loggar in med ditt eget skapade konto med den e-postadress du angav och det lösenord du satte på kontot i formuläret för **Sign in to your account**.

Transkribus appens startsida ser ungefär ut såhär och innehåller följande.



**Home** (Hem) - Länk för att komma tillbaka till startsidan i applikationen.

**Collection** (Samlingar) - Gruppering av ett eller flera dokument i form av olika samlingar med dokument. Mer om Collections nedan.

**Tags** (Taggar)- Taggar som du kan använda för att markera olika valfria aspekter av ditt dokument för att göra det hanterbart för dig.

**Desk** - Den här startsidan.

**Models** - Innehåller det texttolkningsmodeller som finns i Transkribus samt möjlighet att skapa egna modeller (överkurs).

**Sites** (anläggningar) - Publicerade sajter med transkriberade dokument. Du kan också göra din egen sajt här. (överkurs).

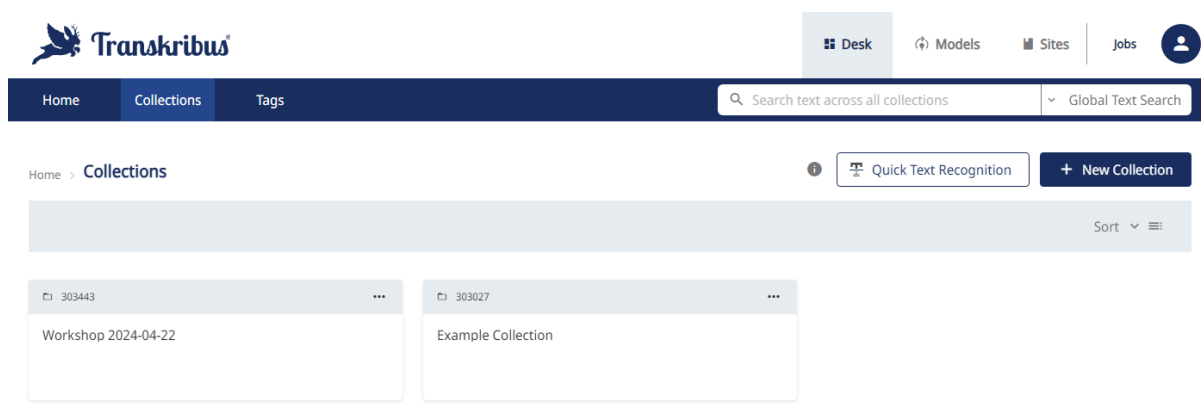
**Jobs** (Jobb) - Lista med dina beställda serverarbeten och dess status.

**Ditt konto (figuren)** - Här kan du hantera ditt konto. Här finns också möjligheten att byta språk till svenska. Allt är dock inte översatt, inte heller så bra i vissa fall, så engelska är nästan att föredra.

## Förbereda dokument

### Skapa *Collection*

Logga in och skapa en ny *Collection* (samling) som motsvarar ett arbete som du vill göra, ett projekt eller ett visst arkiv, som kan innehålla flera mappar/volymer med bilder. Det går att byta namn på en *Collection* efteråt. Man kan också dela sin *Collection* med andra användare.

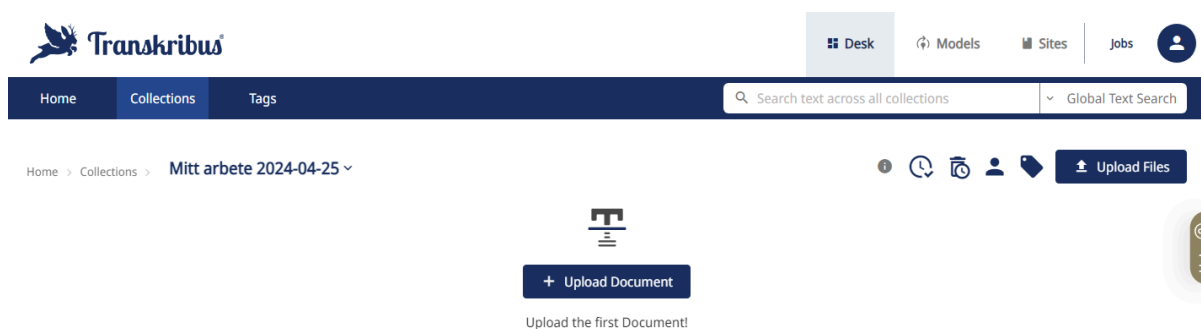


Du får namnge din *Collection* med ett lämpligt beskrivande namn.

### Skapa *Document*

Ladda därefter upp de bilder du vill få transkriberade.  
Det går att lägga till fler bilder till din *Collection* senare.

Du kan antingen lägga till dina bilder direkt eller backa och lägga till bilder vid ett senare tillfälle. Klicka i sådana fall på din *Collection* och välj **Upload Files** och ange *Document Title*. *Document* motsvarar en mapp som man samlar bilder under, till exempel en inbunden volym med flera skannade sidor. Det kan vara bilder du fotat av själv, eller bilder från till exempel Riksarkivets söktjänst.



Döper *Document* med en lämplig titel.

### File Upload ✕

Document Title

Häggdångers kyrkoarkiv - Sockenprotokoll - 1846-1862

Drag & Drop your files or [Browse](#)

**Note:** Accepted file formats are JPEG/JPG (10 MB), PNG (10 MB), and PDF (200 MB) with maximum 3000 pages. [More info.](#)

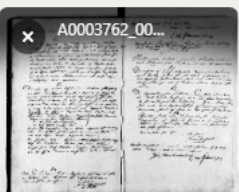
Dokument tillagda som jag vill ha uppladdade.

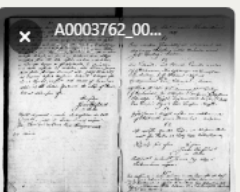
### File Upload ✕

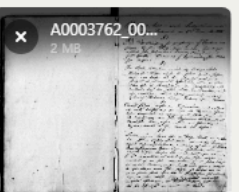
Document Title

Häggdångers kyrkoarkiv - Sockenprotokoll - 1846-1862

Drag & Drop your files or [Browse](#)

✕ A0003762\_00...  


✕ A0003762\_00...  


✕ A0003762\_00...  


**Note:** Accepted file formats are JPEG/JPG (10 MB), PNG (10 MB), and PDF (200 MB) with maximum 3000 pages. [More info.](#)

Klicka på knappen **Submit** som finns längst ned i fönstret.

I *Jobs* går det att se att arbetet pågår med uppladdning av dina bilder.

Home > **Jobs** 🔍 Job ID...

Title	Search	User	State (All)	Date created	Date started	Description	Nr of	Job-ID	Action
Document Upload		mikawinb@gmail.com	RUNNING...	Apr 25, 2024, 10:35	Apr 25, 2024, 10:35	Uploaded PAGE XML for p...		9066131	...
Eskilstuna Kloster och Fors Lla:1		mikawinb@gmail.com	FINISHED	Apr 22, 2024, 14:33	Apr 22, 2024, 14:33	Done, duration: 3m 17s 7...	1-7	9019790	...
Bouppteckningar		mikawinb@gmail.com	FINISHED	Apr 22, 2024, 14:32	Apr 22, 2024, 14:32	Done, duration: 26s 857ms	2	9019769	...

Du kan behöva uppdatera webbsidan (tangenta: F5) för att se om uppladdningen eller andra arbeten är färdiga.

Home > **Jobs** 🔍 Job ID...

Title	Search	User	State (All)	Date created	Date started	Description	Nr. of	Job-ID	±1	Action
Häggdångers kyrkoarkiv - Sockenprotokoll - ...	Document Upload	mikawinb@gmail.com	FINISHED	Apr 25, 2024, 10:35	Apr 25, 2024, 10:35	Done, duration: 9s 749ms		9066131		...
Eskilstuna Kloster och Fors Lta:1	Text Recognition	mikawinb@gmail.com	FINISHED	Apr 22, 2024, 14:33	Apr 22, 2024, 14:33	Done, duration: 3m 17s 7...	1-7	9019790		...

## Texttolka dokumentbilder

Det är nu det roliga arbetet börjar när du ska börja tolka bildens text. Klicka dig in i din *Collection* och det *Document* som du skapat där du laddat upp de bilder du vill arbeta med.

Markera en eller flera bilder genom att klicka i den lilla rutan nedtill vänster vid respektive bild.

Klicka på knappen **Recognize** som finns ovanför bilderna.



Home > Collections > Mitt arbete 2024-04-25 > Häggdångers kyrkoarkiv - Sockenprotokoll - 1846-1862 ▾



## Första gången

✕
Text Recognition
Layout

Häggdångers kyrkoarkiv - Sockenprotokoll - 1846-1862
✕

Start Recognition

Language Model
⚙️ Advanced Settings
Credits needed: -3.00

Available (Personal | Collection)
86 | 0

★ Favorite Models 0

🌐 Public Models 187

🔒 Private Models 0

Filter

🔍 Search ...

Languages

🔍 Search


Handwritten or Printed

Centuries

0 — 21

Name	Words	Language
Search		
The Text Titan I (Super model)		GER, DUT, FRE, FIN, SWE, ENG
The German Giant I	15 420 976	GER
The Dutchess I	11 693 499	DUT
Transkribus Print M1	5 068 310	GER, ENG, DUT, FRE, SWE, FIN, POL, ITA, SPA, CZE, SLO, SL
Transkribus French Model 1	1 933 011	FRE
Latin Incunabula (Reichenau)	672 762	LAT
Latin/German Bilingual Incunabula (Reichenau)	113 794	LAT, GER
German Incunabula (Reichenau)	60 559	GER
Tibetan Modern U-chen Print 0.1	9 822	TIB
Ukrainian Wikisource	89 442	UKR
T&S_H1_M1	135 584	GER
HABE_All GT 24	323 776	BAS
15th Century Spanish Gothic Hybrid Script (model B)	41 435	SPA
Pinkas Brody Model	159 013	YID, HEB
14th Century Spanish Gothic Hybrid (model A)	34 394	SPA

Scholar Featured ID: 51170



### The Text Titan I (Super model)

Created by Transkribus Apr 5, 2023

🔍 Languages GER, DUT, FRE, FIN

📄 Training Set Size

📊 CER (Accuracy) 2.95%

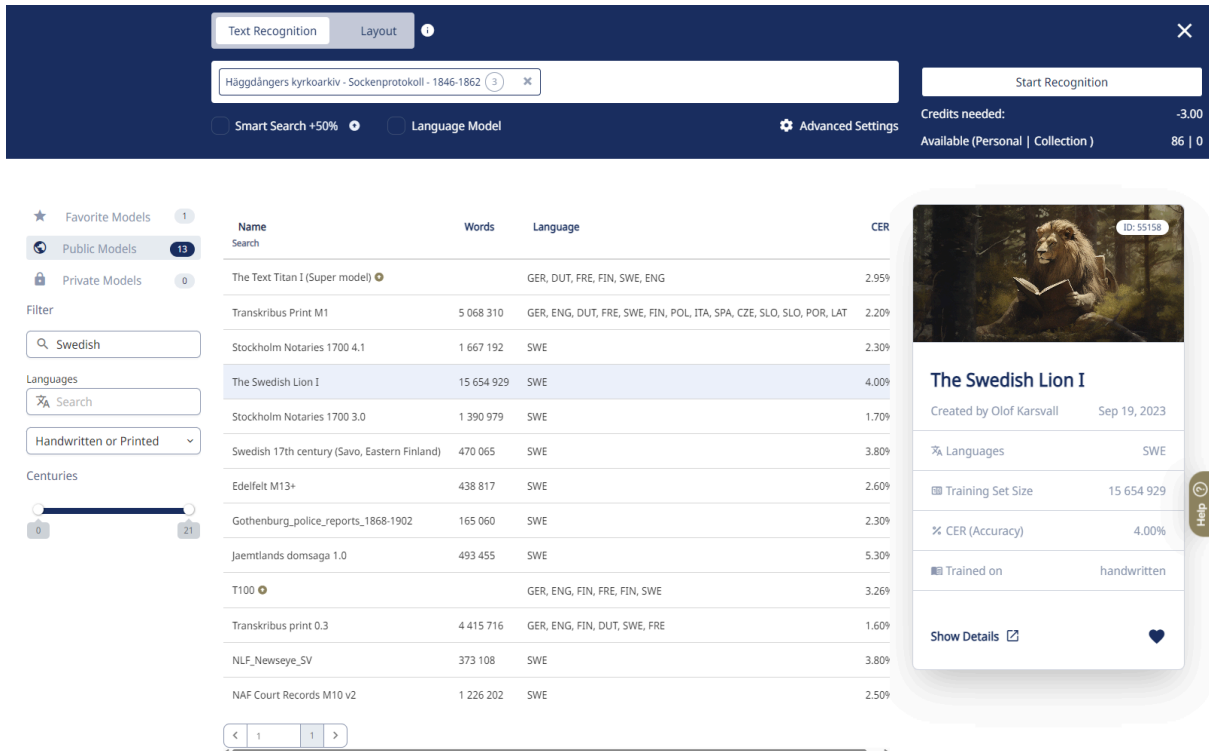
📅 Centuries 16-21

🎓 Trained on handwritten

Show Details
❤️

Sök efter “*Swedish*” eller bläddra i listan bland alla modeller som finns tillgängliga. Markera ***The Swedish Lion I*** i listan. Det är den modellen som Riksarkivet har arbetat med att ta fram. Du ser detaljer om vald modell till höger. Klicka gärna i det lilla hjärtat längst ned till höger i rutan med detaljer om vald modell. På så vis sparar du den modellen som en favorit så att du hittar den lättare nästa gång.

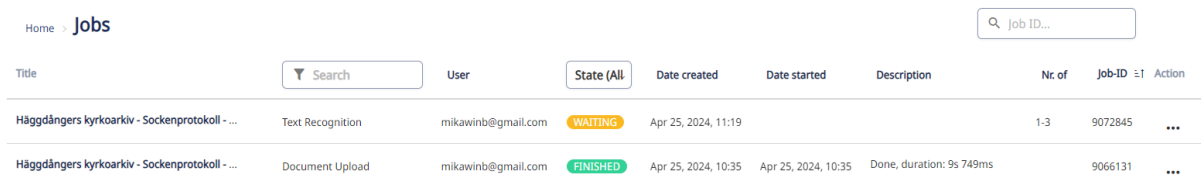




Name	Words	Language	CER
The Text Titan 1 (Super model)		GER, DUT, FRE, FIN, SWE, ENG	2.95%
Transkribus Print M1	5 068 310	GER, ENG, DUT, FRE, SWE, FIN, POL, ITA, SPA, CZE, SLO, POR, LAT	2.20%
Stockholm Notaries 1700 4.1	1 667 192	SWE	2.30%
The Swedish Lion 1	15 654 929	SWE	4.00%
Stockholm Notaries 1700 3.0	1 390 979	SWE	1.70%
Swedish 17th century (Savo, Eastern Finland)	470 065	SWE	3.80%
Edelfelt M13+	438 817	SWE	2.60%
Gothenburg_police_reports_1868-1902	165 060	SWE	2.30%
Jaemtlands domsaga 1.0	493 455	SWE	5.30%
T100		GER, ENG, FIN, FRE, FIN, SWE	3.26%
Transkribus print 0.3	4 415 716	GER, ENG, FIN, DUT, SWE, FRE	1.60%
NLF_Newseye_SV	373 108	SWE	3.80%
NAF Court Records M10 v2	1 226 202	SWE	2.50%

Klicka på **Start Recognition** uppe till höger.  
Det går nu åt en kredit per dokument som du vill ha texttolkad.

Jobs listan dyker upp.

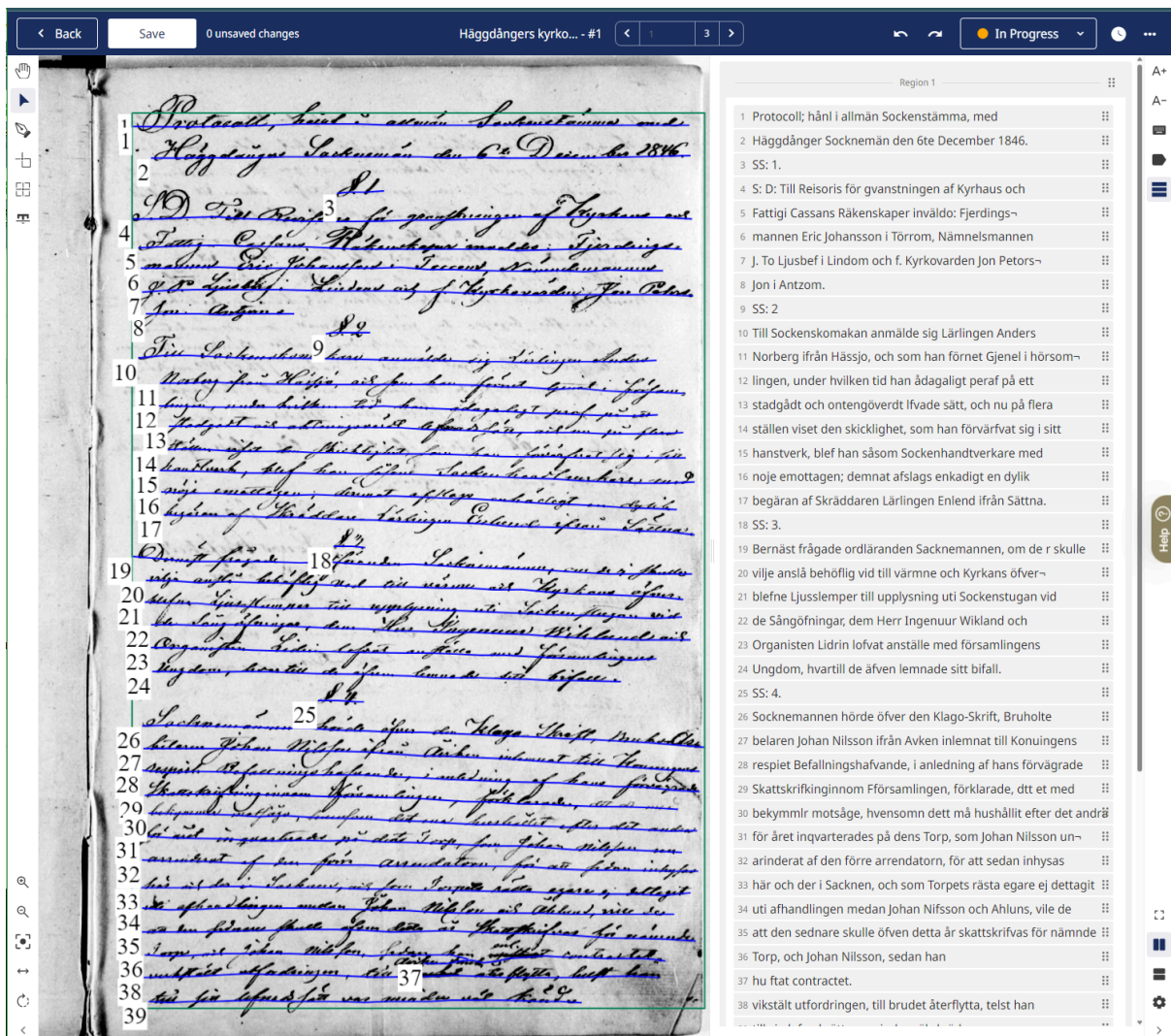


Title	Search	User	State (All)	Date created	Date started	Description	Nr. of	Job-ID	Action
Häggdångers kyrkoarkiv - Sockenprotokoll - ...	Text Recognition	mikawinb@gmail.com	WAITING	Apr 25, 2024, 11:19			1-3	9072845	...
Häggdångers kyrkoarkiv - Sockenprotokoll - ...	Document Upload	mikawinb@gmail.com	FINISHED	Apr 25, 2024, 10:35	Apr 25, 2024, 10:35	Done, duration: 9s 749ms		9066131	...

Uppdatera sidan för att se status på ditt arbete.

När det är klart (Finished) klickar du på rubriken för ditt arbete och därefter på respektive bild i din *Collection*, eller klicka på de tre prickarna ute till höger under rubriken **Action** och välj **Open**.

Du ska då få upp en bild med dokumentet och den tolkade texten till höger.



## Redigera den tolkade texten

HTR-modellen kommer inte att lyckas med att tolka all text exakt rätt. Den tolkade texten till höger går dock att redigera så det blir som du tycker att det egentligen står i dokumentet.

Grundprincipen är en eller flera "Regions" – för en textsida eller de textdelar som man vill hålla isär (t.ex. löptext, rubrik, paginering, marginalanteckningar) – och inom varje "Region" finns "Lines" för varje textrad. Det finns en rad olika snabbkommandon för att redigera Regions och Lines.

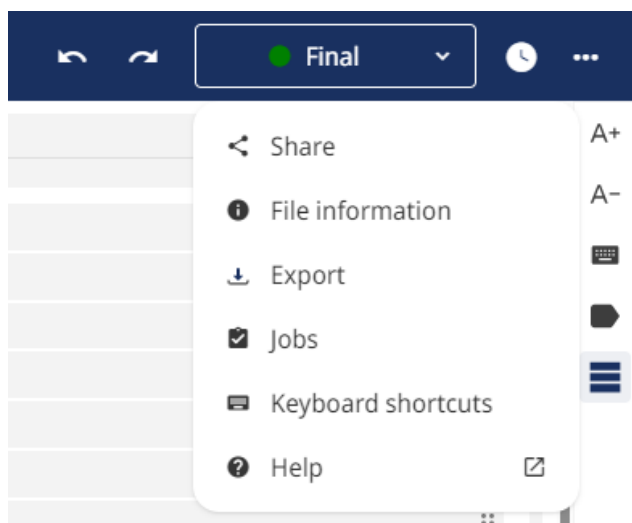
Bekanta dig med dessa i hjälpen: <https://help.transkribus.org/manual-layout-editing>.

Uppe till vänster i menyraden har du knappen **Save** för att spara dina ändringar i den ändrade texten. Klicka på den så att dina ändringar sparas.

Uppe till höger har du en vallista med olika statusar på ditt arbete. Ett nytolkat dokument får statusen "In Progress", du kan sen själv välja bland ytterligare några andra statusar.

## Dela med dig av den tolkade texten

När du känner att den tolkade texten är mer eller mindre "korrekt" utefter vad du har för behov och syfte med texttolkningen så kan du dela med dig av texten på olika sätt.



**Share** - Här skapas en publicering på webben och du får en URL länk som du kan skicka till andra.

**File information** - Du kan lägga in olika egenskaper på dokumentet.

**Export** - Du kan exportera dels bilddokumentet och dels den tolkade texten. Den tolkade texten kan exporteras dels som XML och dels som textfil.

**Jobs** - Du öppnar jobblistan.

**Keyboard shortcuts** - Lista med tillgängliga tangentbords kommandon.

**Help** - Transkribus hjälp.

## Riksarkivet - HTRFLOW

### [HTRFLOW - a Hugging Face Space by Riksarkivet](#)

HTRFLOW är Riksarkivets egen applikation för texttolkning via "The Swedish Lion" modellen. Den är något enklare i sitt utförande. Den har dock inte samma kapacitet för tillfället som Transkribus har så det kan bli väntetid och ta längre tid. Den är främst gjord för demonstrationssyfte i olika sammanhang för att förklara processen i texttolkningen.

HTRFLOW har två olika lägen, ett snabbspår och ett som går steg för steg enligt modellen med att först analysera hela dokumentet och hitta olika textregioner. I steg två genomförs textradssegmentering och slutligen i tredje steget körs själva text igenkänningen.

I den här applikationen ser du tydligare hur verktyget arbetar och vilket resultat den får fram i de olika stegen.