

READ

Recognition and Enrichment
of Archival Documents



Transkriptionen mit Auszeichnungen (Tags) versehen

Version v1.9.1

Letztes Update dieser Anleitung 2.12.2019

Diese Anleitung zeigt, wie man bereits transkribierte Dokumente in Transkribus mit Tags (Auszeichnungen) versieht. Diese Funktion ermöglicht es, Wörter oder Phrasen als Personen, Orte, Abkürzungen oder andere Kategorie zu definieren. Ganz einfach kann man auch eigene Tag-Kategorien hinzufügen und nach bestimmten Tags in den Dokumenten suchen. Außerdem können die Tags exportiert werden. Informationen zum Tag Export finden Sie in dieser Anleitung: [How to Export Documents from Transkribus guide](#).

Laden Sie den Transkribus Expert Client herunter oder stellen Sie sicher, dass Sie die neueste Version verwenden:

- <https://transkribus.eu/>

Besuchen Sie das Transkribus Wiki für mehr Informationen und weitere How to Guides:

- <https://transkribus.eu/wiki/>

Transkribus und die zugrundeliegende Technologie werden durch die folgenden Projekten und Plattformen verfügbar gemacht:

- <https://read.transkribus.eu/>
- <https://transcriptorium.eu/>
- <https://github.com/transkribus/>

Kontakt

- Das Transkribus Team: email@transkribus.eu

Inhalt

Einführung	3
Tagging Interface	3
Eigene Tag Kategorien erstellen	4
Tags hinzufügen	5
Historische Zeichen und Abkürzungen	9
Unlesbarer Text	11
Streichungen	11
Geschwärzter Text	12
Nach Tags suchen	13
Metadaten	15
Editorial Declaration	16
Danksagung	16



Das READ Projekt wird durch das Horizon 2020 Forschungs- und Innovationsprogramm im Rahmen des Fördervertrags Nr. 674943 finanziert.

Einführung

Die Tagging-Funktion in Transkribus ermöglicht:

- Wichtige Wörter und Phrasen im Dokument mit Tags zu versehen.
- Nach einzelnen Tags oder Tag Kategorien zu suchen
- Die hinzugefügten Tags in verschiedenen Formaten zu exportieren, um sie außerhalb von Transkribus weiterverarbeiten zu können.

Tagging Interface

- Die Tagging Interface findet man unter dem „Metadata“ und dann „Textual“ Tab.

The screenshot shows the Transkribus tagging interface. The top navigation bar includes 'Server', 'Overview', 'Layout', 'Metadata', and 'Tools'. Under 'Metadata', the 'Textual' tab is selected. The main area is titled 'Tags of current Transcript' and contains a table with the following data:

	Tag	Value	Text	Properties
1	organization	Komität	erscheinenden Komittenscha	
2	abbrev	Komität	erscheinenden Komittenscha	expansion: Kommittens
3	abbrev	ernsth	Komittenschaften zur ernstl	expansion: ernsthaften
4	abbrev	empfc	ernsthaft überlegung empfo	expansion: empfohlen
5	textStyle	n	überlegung empfahl werde	strikethrough: true
6	textStyle	möge	empfohl werden möge, da-	strikethrough: true
7	abbrev	lezthir	jene bei beeden lezthinig k	expansion: lezthinigen
8	organization	Kongr	beeden lezthinig Kongresse	
9	organization	Stadt	Zusage, lokern, den Stadt, G	
10	abbrev	Stadt	Zusage, lokern, den Stadt, G	
11	organization	Gerich	lokern, den Stadt, Gerichten	
12	abbrev	u.	Stadt, Gerichten, u. Gemein	expansion: und
13	organization	Geme	Gerichten, u. Gemeinden be	
14	abbrev	erheyr	ledig - oder verheyrath Stan	expansion: verheyrather
15	abbrev	un-	ein art einer un- gezwunger	

Below the table is a 'Tags' panel with a 'Show all' checkbox (checked) and a 'Customize...' button. It contains a table of all tags:

Tag specification	Color	Shortcut
abbrev	Red	
add	Cyan	
Address	Green	
blackening	Black	
date	Blue	

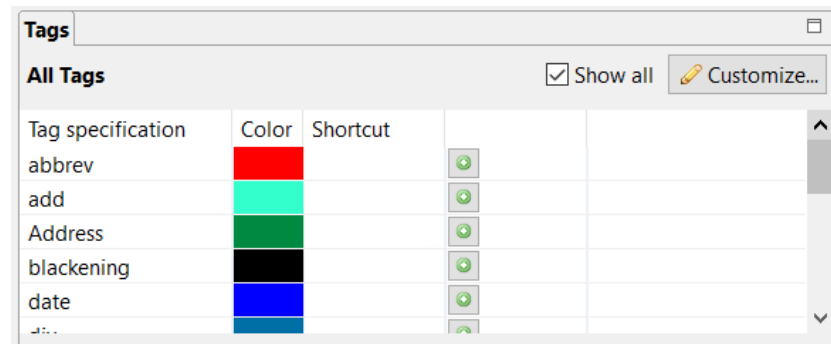
At the bottom of the panel, it says 'Props for tag: no tag selected' and there are buttons for 'Previous', 'Next', and 'Apply to selected'.

Darstellung 1 Der "Textual" Tab

- Wenn man bei „Show all“ ein Häkchen setzt, werden alle vordefinierten Tags angezeigt. Diese können direkt verwendet werden.



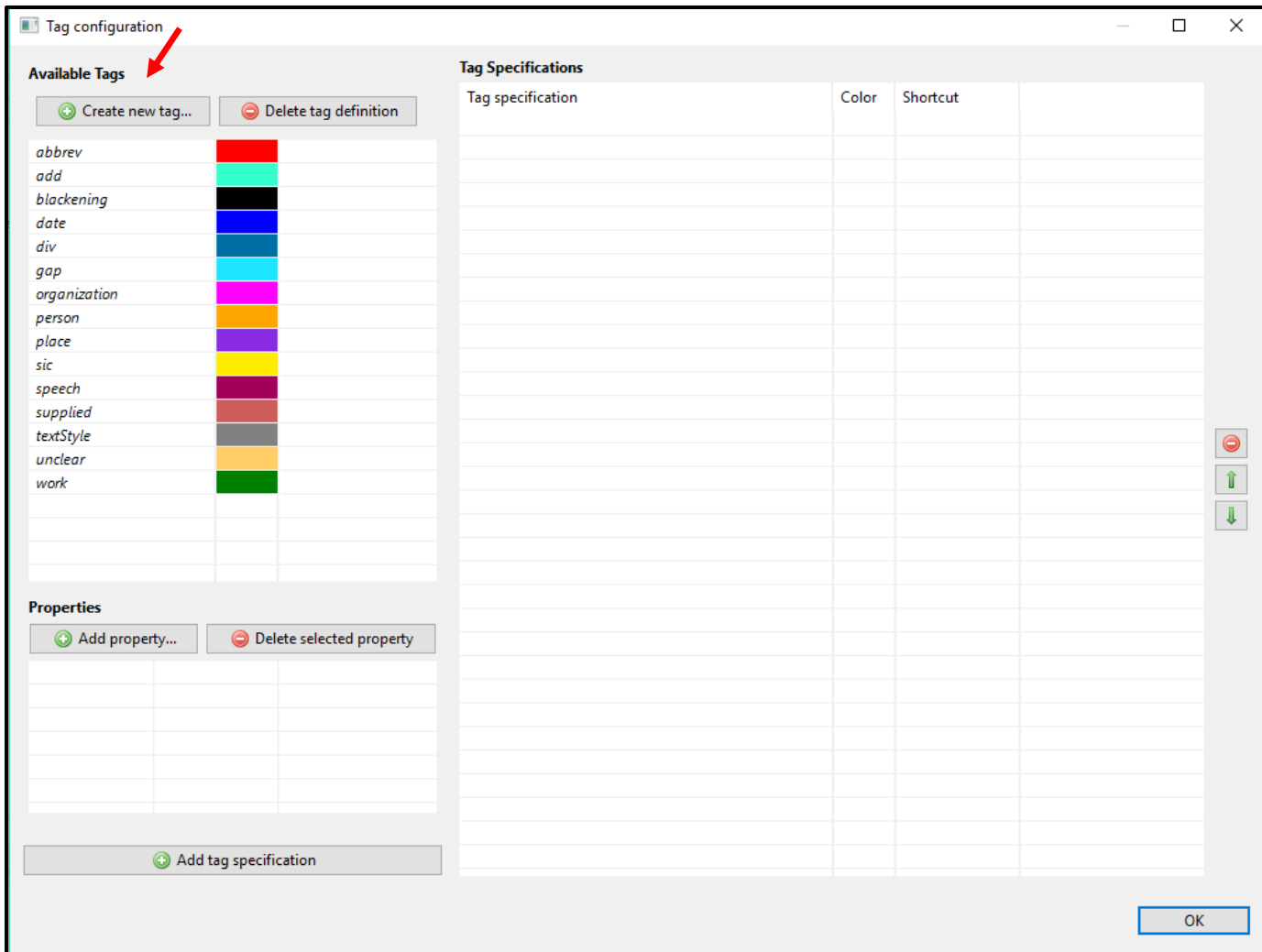
Darstellung 2 Alle vordefinierten Tags anzeigen



Darstellung 3 Vordefinierte Tags

Eigene Tag Kategorien erstellen

- Um Ihre eigene Tag Kategorie zu erstellen, klicken Sie auf „Customize“ im „Tags“ Tab, um das „Tag configuration“ Fenster zu öffnen.

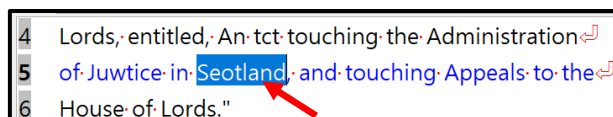


Darstellung 4 Eigene Tag Kategorien definieren

- Mit Klick auf „Create new tag“, kann man seine eigenen Tage Kategorien definieren.
- Wenn Sie eine eigene Kategorie definiert haben, erscheint diese auch, wenn Sie auf die „Show all“ Schaltfläche klicken.
- Im „Tag configuration“ Fenster erscheinen vordefinierte Tags kursiv und eigene in Normalschrift.

Tags hinzufügen

- Wenn Sie ein Wort oder eine Phrase mit einem Tag markieren möchten, haben Sie (mindestens) drei Möglichkeiten:
 - o Markieren Sie den Text im Text Editor und klicken Sie danach auf den grünen Kreis jener Tag Kategorie, die Sie gerne setzen möchten.

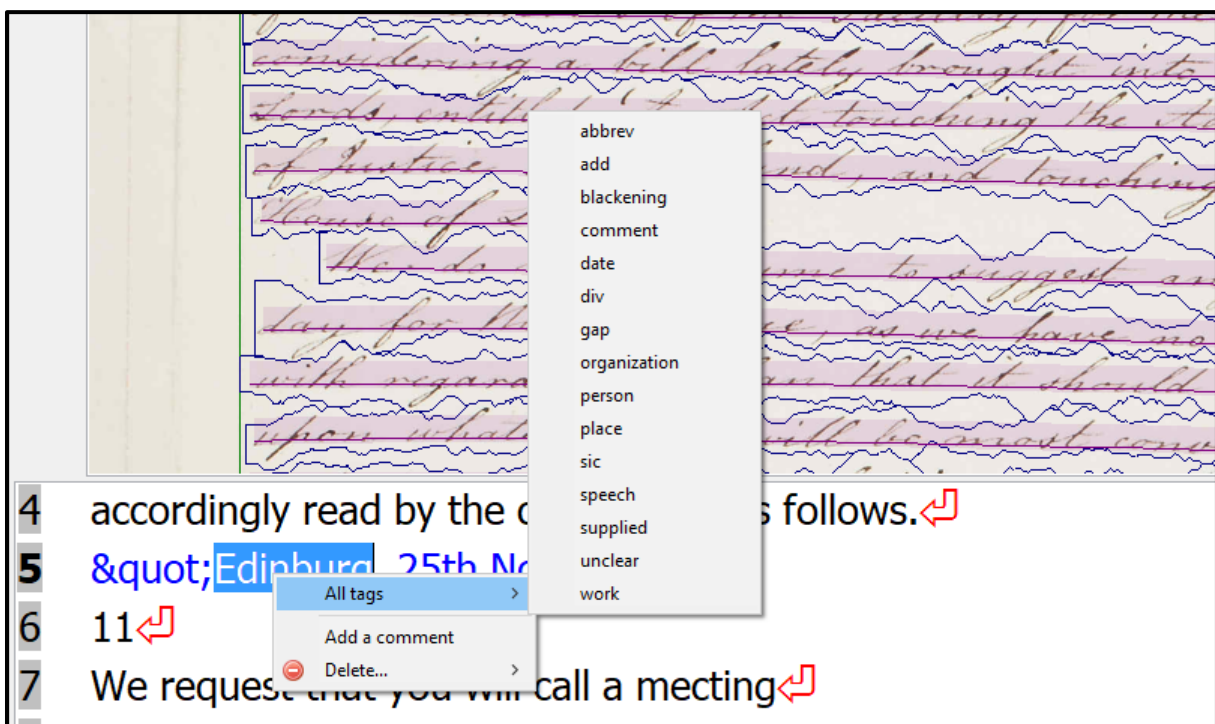


Darstellung 5 Das Wort markieren, das mit einem Tag versehen werden soll

Tag specification	Color	Shortcut	
abbrev	Red		<input type="checkbox"/>
add	Cyan		<input type="checkbox"/>
blackening	Black		<input type="checkbox"/>
date	Blue		<input type="checkbox"/>
div	Dark Blue		<input type="checkbox"/>
gap	Light Blue		<input type="checkbox"/>
organization	Magenta		<input type="checkbox"/>
person	Orange		<input type="checkbox"/>
place	Purple		<input checked="" type="checkbox"/>
sic	Yellow		<input type="checkbox"/>
speech	Maroon		<input type="checkbox"/>
supplied	Brown		<input type="checkbox"/>
unclear	Light Orange		<input type="checkbox"/>
work	Green		<input type="checkbox"/>

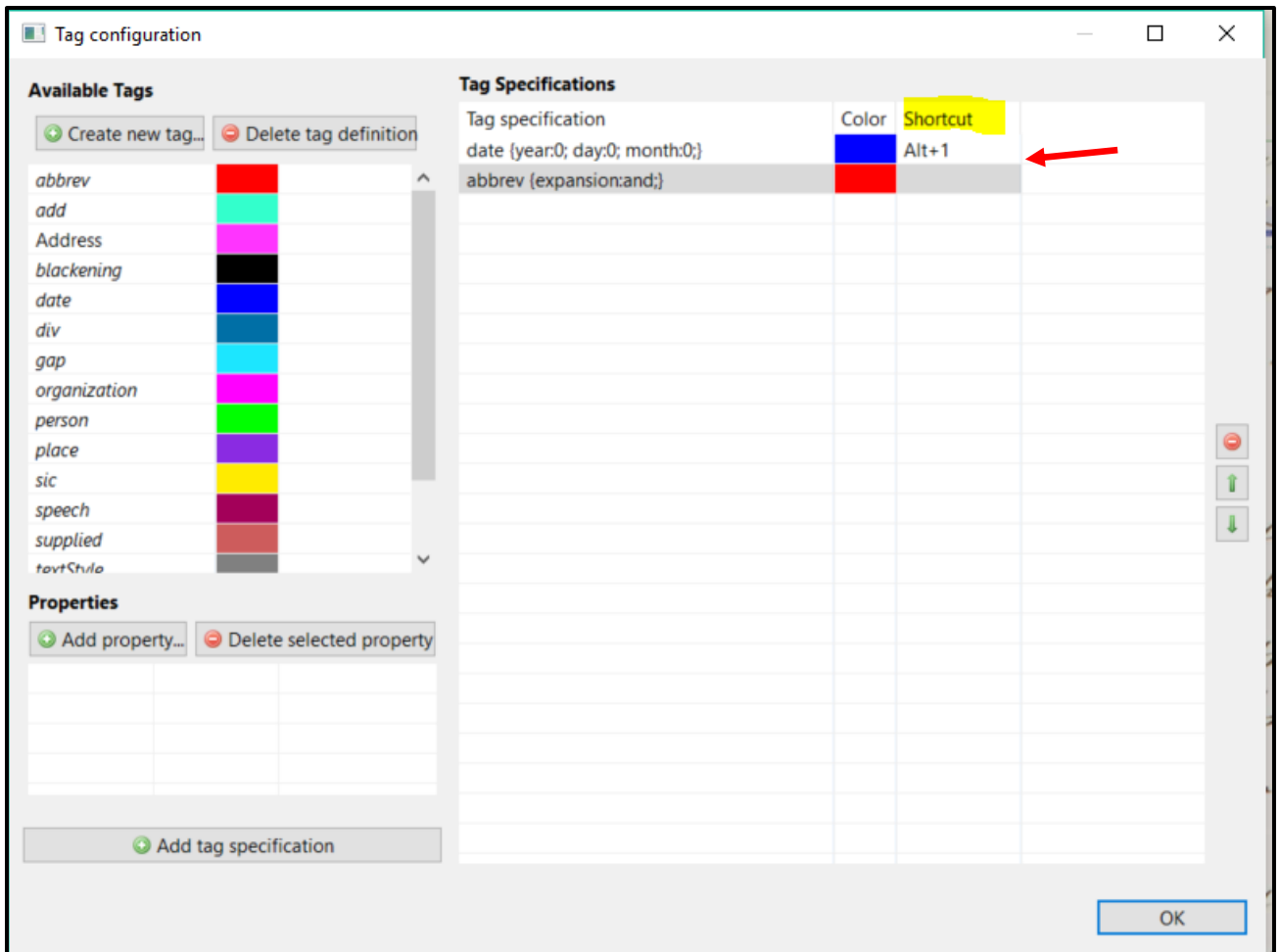
Darstellung 6 Die passende Tag Kategorie wählen

- Alternativ können Sie auch das Wort oder die Phrase markieren, dann rechts klicken und dann unter „All tags“ die passende Kategorie auswählen.



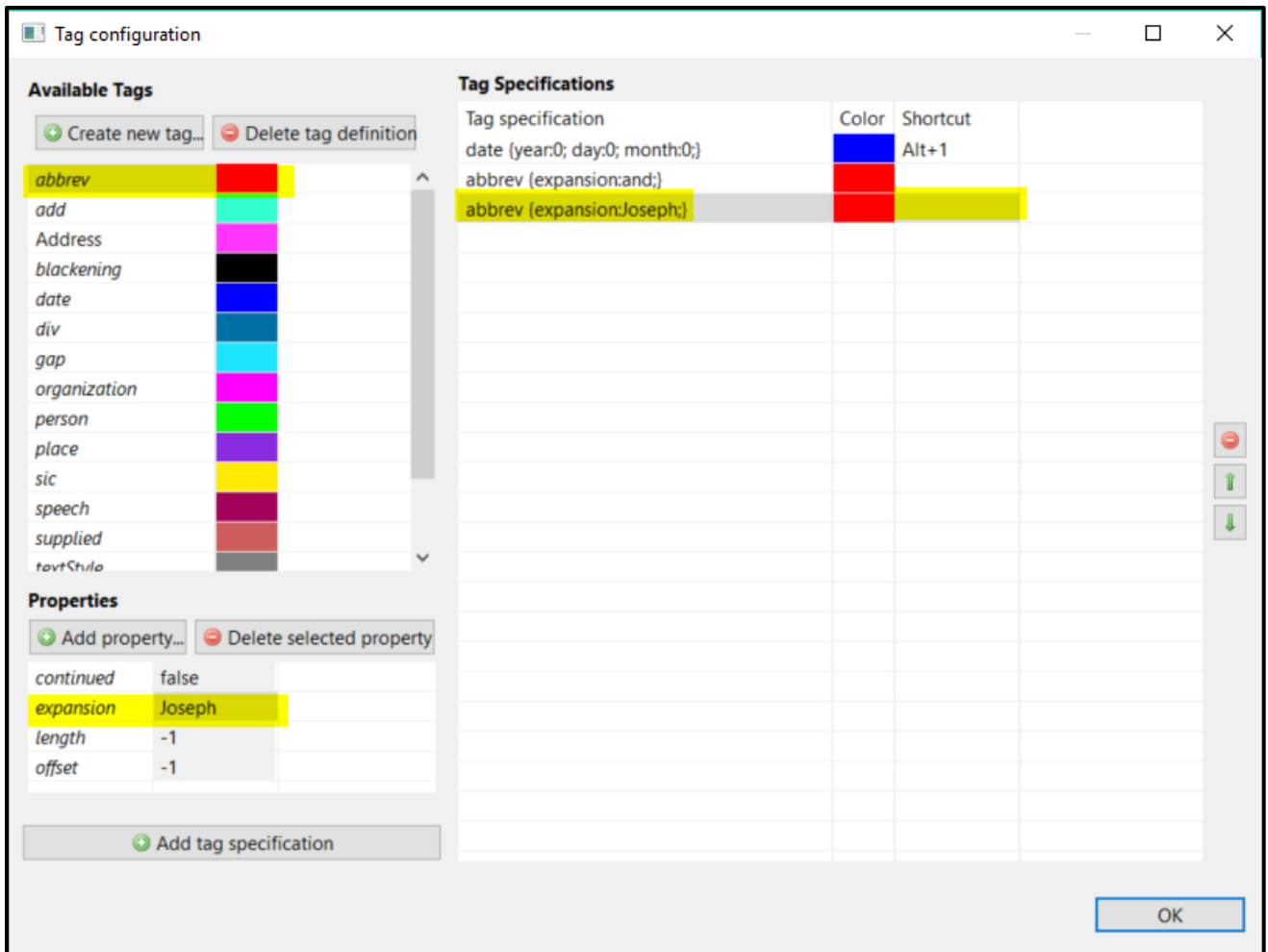
Darstellung 7 Mit rechtem Mausklick einen Tag setzen

- Wenn Sie gleiche Tag Kategorien öfters verwenden, können Sie für jene einen Shortcut setzen, das spart Zeit. Dazu gehen Sie in den „Textual“ Tab und dann in den „Tags“ Tab. Hier klicken Sie auf „Customize“. Im „Tag Specification“ Bereich können Sie nun die gewünschte Tastenkombination in der „Shortcut“ Spalte eintippen.



Darstellung 8 Shortcuts für regelmäßig verwendete Kategorien hinzufügen

- Außerdem können Shortcuts für die Properties von Tags hinzugefügt werden, zum Beispiel für die Erweiterung von Abkürzungen oder für standardisierte Ländernamen usw.
 - o Dazu klicken Sie auf „Customize“ im „Tags“ tab.
 - o Im „Tag configuration“ Fenster klicken Sie auf den gewünschten Tag. Die Details zu diesem Tag werden dann im „Properties“ Bereich gezeigt.
 - o Klicken Sie „Add property“, um die gewünschten Eigenschaften hinzuzufügen.
 - o Dann klicken Sie „Add tag specification“
 - o Nun erscheint der Tag und seine Eigenschaften (z.B. die Erweiterung für eine Abkürzung) im „Tag Specification“ Bereich.
 - o Fügen Sie den Shortcut hinzu, den Sie gerne verwenden möchten.
 - o Jetzt können Sie den Tag mit seinen Properties ganz einfach zu einem Wort oder einer Phrase hinzufügen, indem Sie es im Text Editor markieren und dann den Shortcut drücken.



Darstellung 9 Eine fixierte Erweiterung hinzufügen

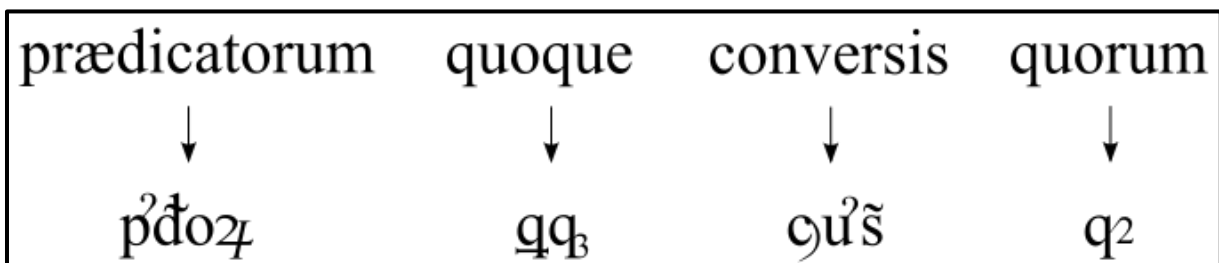
- Wenn Sie einen gesetzten Tag entfernen möchten, erreichen Sie das, indem Sie das Wort oder die Phrase markieren, dann rechts klicken und „Delete“ auswählen. Das Programm gibt Ihnen dann zwei Optionen:
 - o Nur den markierten Tag löschen
 - o Alle Tags dieser Collection löschen
- **Anmerkung:** Tags können auf Regionen-, Wörter- oder Zeichenbasis gesetzt werden. Anleitungen für Struktur-Tags finden Sie hier: [How to Use the Structural Tagging Feature.](#)
- Es gibt keine Begrenzung hinsichtlich der Anzahl von Tags in einem Dokument, man kann so viele, wie nötig hinzufügen.
- Im „Textual“ Tab gibt Transkribus einen Überblick über die von Ihnen gesetzten Tags.

Server Overview Layout Metadata Tools				
Document Structural Textual Comments				
Tags of current Transcript				
	Tag	Value	Text	Properties
11	abbrev	ds	Bertolini selbst für ds Pferd 2 f 30 Xr	
12	abbrev	f	für ds Pferd 2 f 30 Xr an- gebothen	
13	abbrev	Xr	ds Pferd 2 f 30 Xr an- gebothen hak	
14	abbrev	f:	habe, und daß um 20 f: zu wenig Ge	
15	organization	Magistr	satze, daß der Magistrat von von dei	
16	abbrev	löb.	hoffe, massen von löb. Kreisamt un	
17	organization	Kreisam	massen von löb. Kreisamt unterm 1	
18	date	14t 9br	Kreisamt unterm 14t 9br abhin und	
19	textStyle	t	Kreisamt unterm 14 t 9br abhin und	superscript:
20	abbrev	t	Kreisamt unterm 14 t 9br abhin und	
21	abbrev	9br	Kreisamt unterm 14t 9br abhin und	
22	organization	Militar	durchgehends von dem Militar Ver	
23	abbrev	u.	Vertheilungsmagazin besorget, u. vi	
24	abbrev	u	für den Staab u. für die Kompagnie	
25	abbrev	werd.	Kompagnien gefodert werd.	
26	abbrev	H'r	H'r v̇ Mayrl. No 1459.	
27	person	v̇ Mayrl	H'r v̇ Mayrl. No 1459.	
28	abbrev	v̇	H'r v̇ Mayrl. No 1459.	
29	abbrev	No	H'r v̇ Mayrl. No 1459.	
30	textStyle	o	H'r v̇ Mayrl. o 1459.	superscript:
31	organization	Gubern	Gubernial -Circulare d. d.	
32	abbrev	d. d	Gubernial-Circulare d. d. 4t, præ. 2i	
33	date	4t	Gubernial-Circulare d. d. 4t , præ. 2i	

Darstellung 10 Überblick

Historische Zeichen und Abkürzungen

- In historischen Dokumenten haben Abkürzungen oft eine weit größere Bedeutung, als in modernen Dokumenten. Es wurde viel abgekürzt, in der Hoffnung Zeit und/oder Papier zu sparen. In einigen Dokumenten sind mehr als 20-30% der Wörter abgekürzt, wie *Darstellung 11* zeigt.



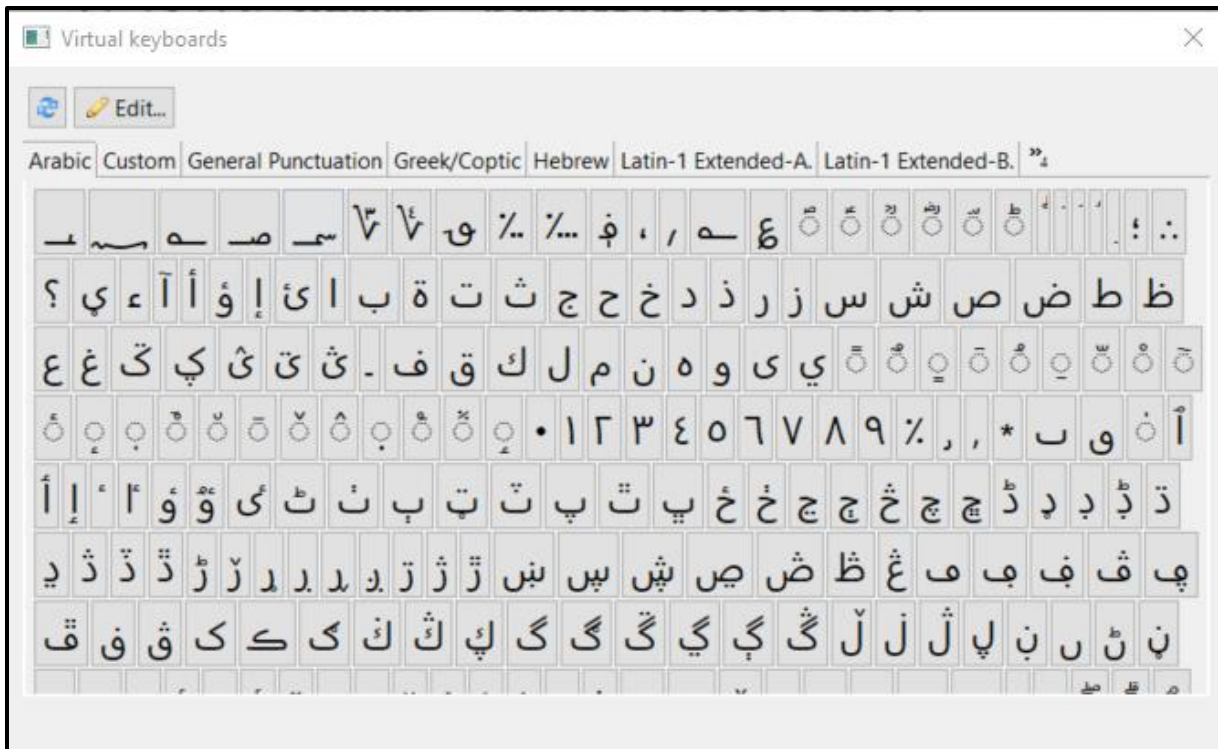
Darstellung 11 Beispiel für typische Abkürzungen in einem lateinischen Text aus dem Mittelalter (cf. Wikipedia: https://en.wikipedia.org/wiki/Scribal_abbreviation)

- Es gibt zwei Möglichkeiten, abgekürzten Text zu transkribieren:
 - o **Option 1:** Die Abkürzung in der üblichen Weise erweitern. Neuronale Netze sind oft schon fähig, Erweiterungen zu erkennen und zu reproduzieren. Zum Beispiel lateinische Präfixe und Suffixe wie „cum“, „con“ oder „us“ und „orum“ können einfach

- erlernt werden. Das heißt man muss nur eine erweiterte Version des Textes in der Transkription zur Verfügung stellen.
- **Option 2:** bitte halten Sie sich an die oben genannte Regel – solange Sie die Basiszeichen erkennen können – transkribieren Sie die Basiszeichen. Dies gilt vor allem für Historiker und all jene, die Interesse am „Inhalt“ eines Dokuments haben und jene, die Trainingsdaten für ein HTR Modell herstellen möchten.
 - Die richtige Transkription für die oben gezeigten Beispiele wäre also:
 - **pdor – qq – cus – qr**
 - **Anmerkung:** In Zukunft sollte es möglich sein, dass HTR Maschinen lernen, Abkürzungen automatisch zu erweitern (oder im Gegenzug die richtige Abkürzung für einen Erweiterung bereitstellen), um eine computergestützte Transkription möglich zu machen.
 - **Option 3:** wenn Sie Unicode Zeichen verwenden möchten, die den Graphemen im Originaldokument nahekommen, dann können Sie den Text mit Hilfe von Unicode transkribieren.
 - In diesem Fall sollte die Transkription des Beispiels folgendermaßen aussehen:
 - p[˚] (LATIN SMALL LETTER P COMBINING OGONEK ABOVE)
 - đ (LATIN SMALL LETTER D WITH MIDDLE TILDE)
 - o (LATIN SMALL LETTER O)
 - ŕ (LATIN SMALL LETTER RUM ROTUNDA; auch LATIN SMALL LETTER R ROTUNDA könnte in diesem Fall verwendet werden)
 - **Anmerkung:** Oft ist es nicht einfach zu entscheiden, welche diakritische oder Unicode Zeichen das richtige ist. Mehr Informationen dazu können Sie auf der MUF1 Website finden: <http://folk.uib.no/hnooh/mufi/>
 - In Transkribus findet man Unicode und andere Spezialzeichen im „Virtual Keyboard“ im Text Editor Menu.



Darstellung 12 „Virtual Keyboard“ Symbol



Darstellung 13 "Virtual keyboards" Fenster

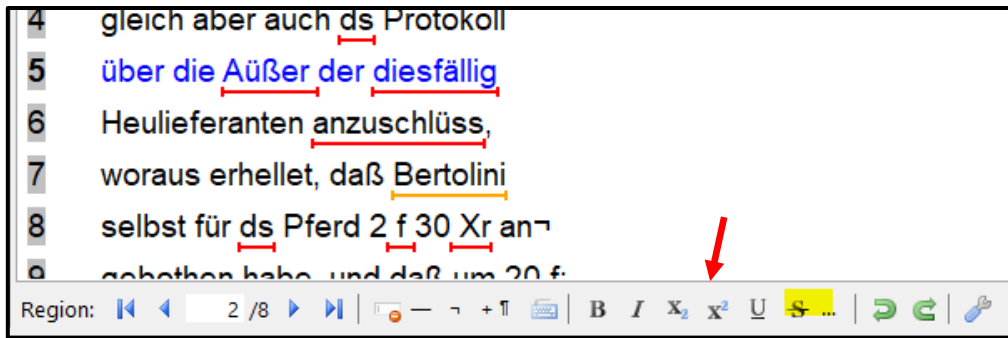
- Modelle mit gemischten Daten sind hier oft nützlich. Oft vorkommende historische Zeichen können zum Beispiel mit Unicode transkribiert werden, während Zeichen, die nur von einem Autor verwendet wurden, mit den Basiszeichen transkribiert werden können.
 - o **Beispiel:** ʒ (LATIN SMALL LETTER RUM ROTUNDA) wird regelmäßig in mittelalterlichen und modernen Texten verwendet. Deshalb wäre es sinnvoll diesen Buchstaben in eine HTR einzubringen, die ausschließlich auf mittelalterliche Dokumente und die Verarbeitung von großen Mengen solcher Dokumente ausgerichtet ist.

Unlesbarer Text

- Text, der nicht transkribiert werden kann, weil er nicht lesbar ist, kann mit den Tag „unclear“ oder „gap“ markiert werden.
- Wenn der Text nicht klar ist, markieren Sie ihn im Text Editor und setzen Sie den Tag „unclear“.
- Wenn der Text unlesbar ist, setzen Sie „gap“.
- Außerdem können Sie unter „Properties“ zum Beispiel Vorschläge für das unlesbare Wort hinzufügen.

Streichungen

- Wenn Sie durchgestrichene Passagen im Text entdecken, gibt es verschiedene Optionen:
 - o **Option 1:** Der durchgestrichenen Text ist **noch lesbar**, oder zumindest große Teile davon. In diesem Fall transkribieren Sie den Text so gut wie möglich und markieren ihn als „strike-through“. Den „strike-through“ button finden Sie im Text Editor Menu.

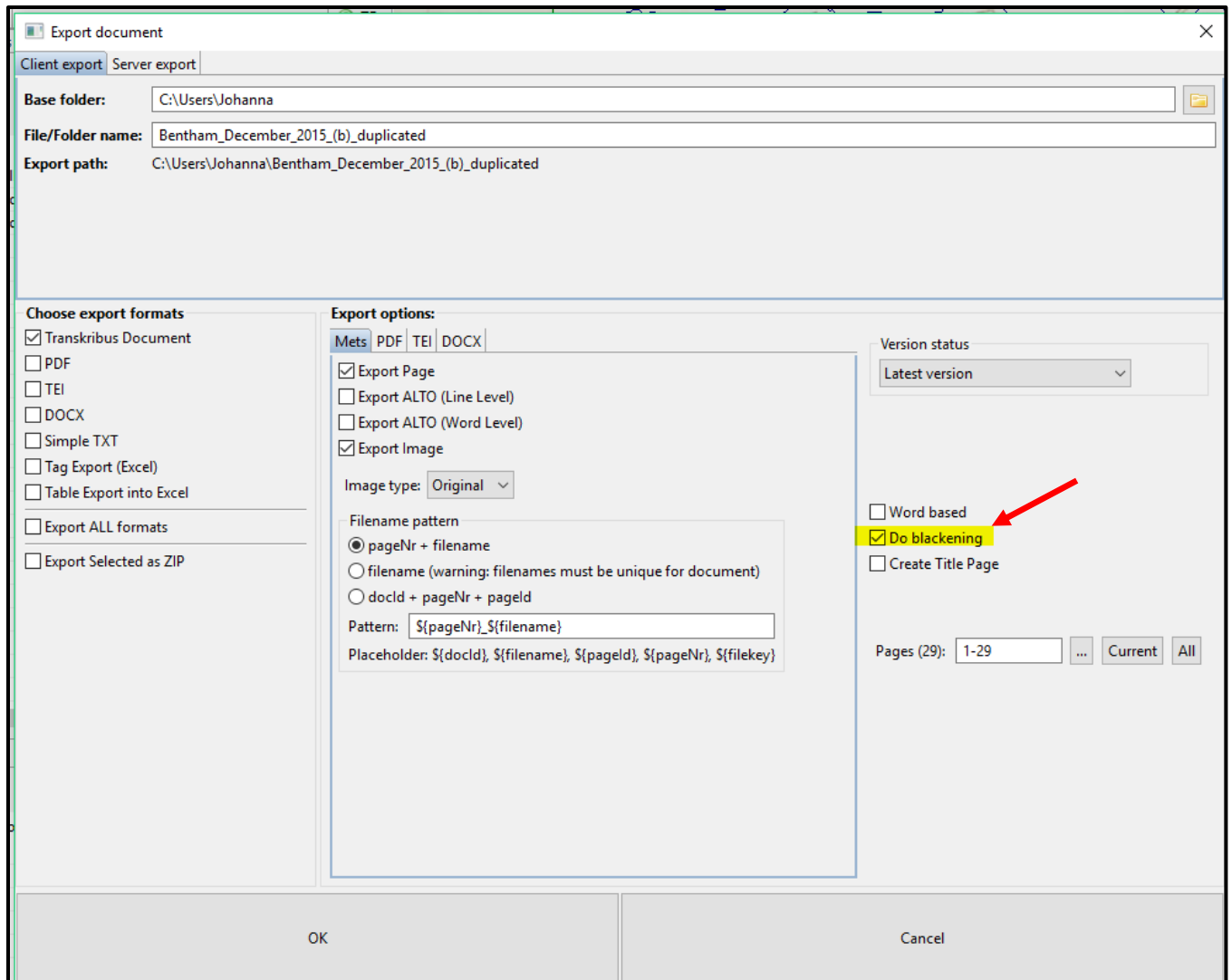


Darstellung 14 „Mark as strike through“ Schaltfläche

- **Anmerkung:** die HTR kann durchgestrichenen Text entziffern, umso mehr Beispiele sie hat, umso besser.
- **Option 2:** Der durchgestrichene Text ist nicht mehr lesbar, oder nur kleine Teile davon. In diesem Fall verwenden Sie den „gap“ Tag.

Geschwärtzter Text

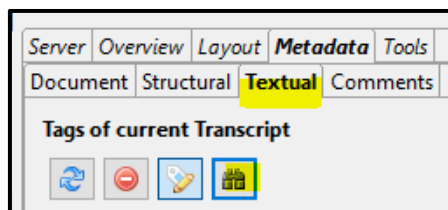
- Mit dem „blackening“ tag können sensible Informationen beim Export geschwärzt werden. Typischerweise wird dieser Tag genutzt, um in einem Dokument, das veröffentlicht wird, persönliche Daten zu schützen.
- Der „blackening“ Tag wird in Verbindung mit der „blackening region“ verwendet, die mit den Segmentierungs-Tools hinzugefügt wird.
- Um zu schwärzen, gehen Sie folgendermaßen vor:
 - Klicken Sie auf den grünen Kreis mit den drei Punkten im Canvas Menu und wählen Sie „Blackening“. Nun ziehen Sie die „Blackening region“ über das Wort oder den Abschnitt, den Sie gerne verbergen würden.
 - Um die geschwärtzten Bereiche zu zeigen, klicken Sie auf „Item visibility“ im Hauptmenü und wählen Sie dann „Render blackenings“.
 - Markieren Sie das entsprechende Wort/den entsprechenden Text im Text Editor und wählen Sie den „Blackeningj“ Tag, der Text wird dann durch [...] ersetzt.
 - Beim Export des Dokuments bitte sicherstellen, dass „Do blackening“ ausgewählt ist.
 - **Anmerkung:** In METS und TEI Dokumenten ist das Wort verborgen, es bleibt im Hintergrund aber bestehen. In den anderen Formaten verschwindet auch der Text im Hintergrund.



Darstellung 15 Beim Export „Do blackening“ auswählen um die Bereiche im exportierten Dokument zu schwärzen

Nach Tags suchen

- Wenn Sie nach bestimmten Tags im Dokument suchen möchten, funktioniert das mit dem Fernglas im „Textual“ Tab.



Darstellung 16 Fernglas Schaltfläche für die Suche nach Tags

- Im „Search for...“ Fenster können Sie Ihre Suche definieren
 - o Wählen Sie, wo Sie gerne suchen würden (current collection, current page...)
 - o Auf Zeilen- oder Wortbasis (line/word level)
 - o Unter „Name“ geben Sie den Namen des Tags ein.
 - o In das „Text“ Feld kommt der geschriebene Text.
 - o Am Schluss auf „Search!“ klicken.
 - o Die Suchresultate werden dann im unteren Bereich des Fensters angezeigt.

Search for...

Documents Fulltext (Solr) Tags KWS

Current collection Line level Case sensitive Exact match

Name: abbrev Text:

Properties to search:

Add property to search: Value: + -

Property	Value

Search!

Search results

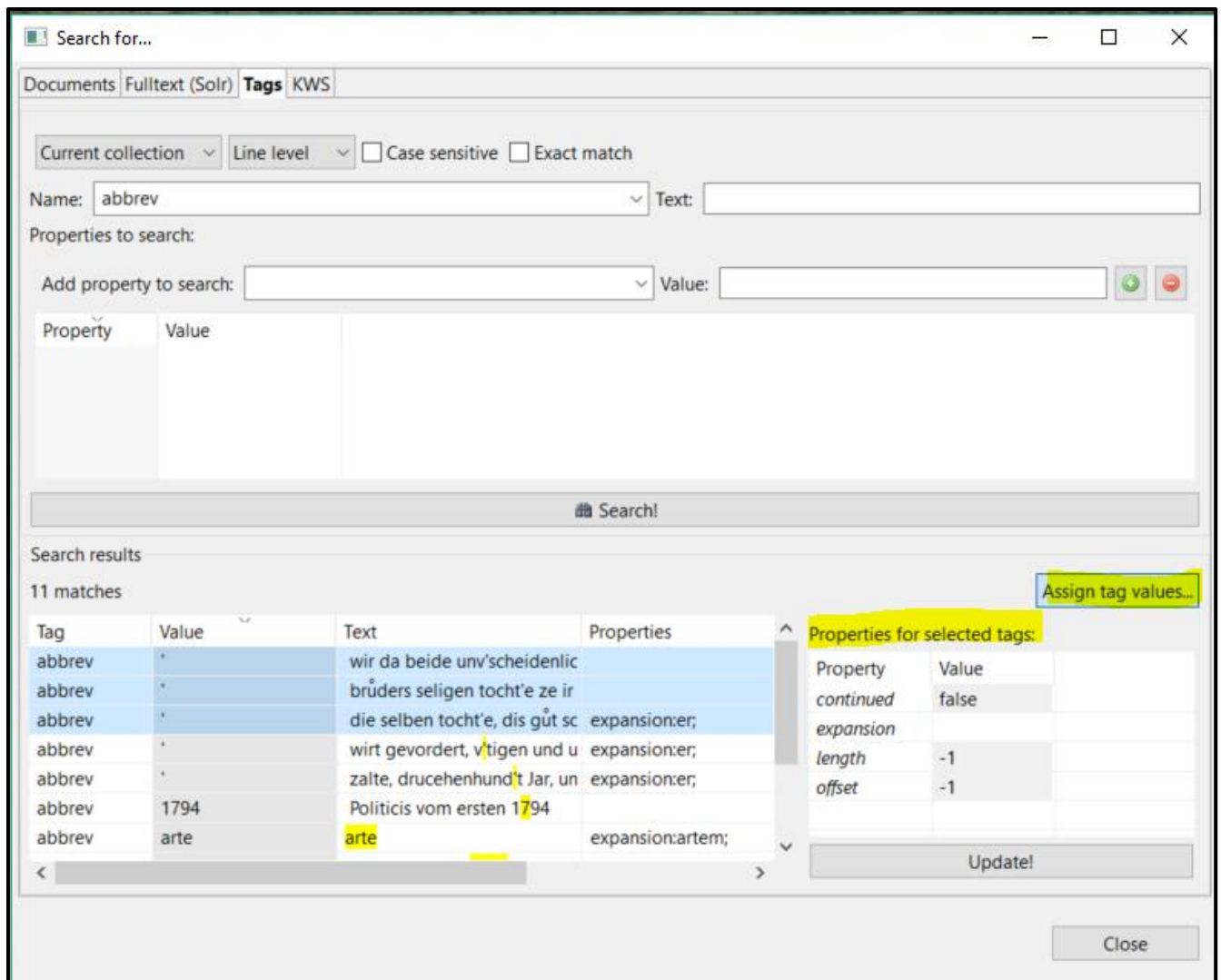
11 matches Assign tag values...

Tag	Value	Text	Properties	Doc	Page	Region
abbrev	turnkeys	more than two turnkeys from morni		3158	06	r111
abbrev	arte	arte	expansion:artem;	29667	01	line_1513
abbrev	obig	daß daß nebst obig Punkt den Anf	expansion:obiger	6679	01	r1007
abbrev	'	wir da beide unv/scheidenlich habe		20759	03	r1013
abbrev	'	swenne die vorgenaßen unser brüde	expansion:an;	20759	03	r1014
abbrev	'	brüders seligen tocht,e ze ir ta ger		20759	03	r1015
abbrev	'	danne fügen un schafften suln	expansion:nd;	20759	03	r1016
abbrev	'	die selben tocht,e, dis güt so	expansion:er;	20759	03	r1017
abbrev	'	wirt gevordert, vtigen und uf geber	expansion:er;	20759	03	r1018
abbrev	'	zalte, drucehenhundt Jar, und dar	expansion:er;	20759	03	r1035
abbrev	1794	Politicis vom ersten 1794		25278	01	r116

Close

Darstellung 17 "Search for.." Fenster für die Suche

- Um schnell und unkompliziert eine Erweiterung oder andere Eigenschaft zu einem immer wieder vorkommenden Wort hinzuzufügen gehen Sie folgendermaßen vor:
 - o Sortieren Sie die Suchresultate nach „Value“. Das geschieht ganz einfach durch das Klicken auf „Value“.
 - o Markieren Sie die gleichen Wörter indem Sie auf sie klicken, während Sie die „Strg“ Taste halten.
 - o Dann klicken Sie auf „Assign tag values...“ und geben die Eigenschaft (property) ein, die hinzugefügt werden soll.



Darstellung 18 Effektiver arbeiten, indem man mehreren Wörtern oder Phrasen die Eigenschaften auf einmal zuordnet

Metadaten

- Im Moment ist in Transkribus nur eine sehr einfache Beschreibung der Dokumente als Metadaten verfügbar, da wir davon ausgehen, dass in digitalen Editionen die meisten Metadaten am externen Server bleiben und auf das Dokument verlinkt werden. Jedes Dokument hat seine eigenen ID und kann auch über die REST Services erreicht werden (zur Verfügung gestellt von der Transkribus Plattform <https://transkribus.eu/wiki/>).
- Die folgenden Felder können derzeit im „Metadata“ und dann „Document“ Tab ausgefüllt werden:
 - o Title
 - o Author
 - o Uploaded
 - o Genre
 - o Writer
 - o Language
 - o Script type
 - o Date of writing
 - o Description

Editorial Declaration

- Da es immer mehrere Wege gibt, um eine korrekte Transkription zu produzieren, ist es wichtig, den gewählten Weg nachvollziehbar zu machen.
- Aus diesem Zweck gibt es die „Editorial Declaration“ Funktion in Transkribus. Man findet Sie im „Metadata“ und dann „Document“ Tab.
- Wie das Tagging System bietet die „Editorial Declaration“ eine Reihe von vordefinierten Funktionen und Optionen. Außerdem können Sie Ihre eigenen Beschreibungen verfassen und sie zusammen mit dem Dokument abspeichern.
- Besonders wichtig ist es, die Sonderzeichen und ihre Verwendung in der „Editorial Declaration“ zu listen

The screenshot shows the 'Metadata' tab in the Transkribus interface. The 'Document' sub-tab is active, displaying various metadata fields. A red arrow points to the 'Editorial Declaration...' button located below the 'Date of writing' section. The fields shown include Title, Author, Uploaded, Genre, Writer, Language (with a list of languages like Arabic, Basque, Bulgarian, Catalan, and Croatian), Script type, and Date of writing (with 'From' and 'To' date pickers).

Darstellung 19 „Editorial Declaration“ Schaltfläche

Danksagung

Wir möchten den vielen Nutzern danken, die mit ihrem Feedback zur Verbesserung der Transkribussoftware beigetragen haben.

Transkribus wird der Öffentlichkeit im Rahmen des H2020e Infrastrukturprojekts READ (Recognition and Enrichment of Archival Documents) zugänglich gemacht, das von der Europäischen Kommission finanziert wird.