

## Nachträgliche Bearbeitung, der aus doc konvertierten html- Dateien der Transkriptionen

(Vereinheitlichung, Vereinfachung, Glättung der Dateien)

A)

- ersten Zeilen löschen (Angaben über Bestandssignaturen)
- Folioangaben der Revision prüfen (RRev, URev)
- „[-]“ löschen (nur meine Dateien)
- Linkeinträge für Fussnoten überprüfen (Dateiname zu #anno\_trans-htm)

1.1.)

Zeilenumbrüche <br/> innerhalb von Tabellenzellen einfügen anstelle von  
Paragraphentrennung

suchen	<p>\s*<p
ersetzen	<p> <p
xpath	//td

1.2.)

Paragraphen zwischen Tabellen

suchen	<p .*/p>
ersetzen	<p class="abstand"/>
xpath	//body/table/following-sibling::p[1]

suchen	<p .*/p>
ersetzen	
xpath	//body/p

suchen	<p class="abstand"/>
ersetzen	<p class="abstand">&nbsp;</p>
xpath	

2.)

Klassen definieren

2.1.)

- Zeilen/Spalten: Klassen löschen

suchen	(<tr.*) class="[^"]*"
ersetzen	\$1
xpath	//tr

suchen	(<td.*) class="[^"]*"
ersetzen	\$1
xpath	//td

2.2.)

- erste Zeile: Metadaten

suchen	<tr
ersetzen	<tr class="first"
xpath	//tr[1]

### 2.3.)

- erste Spalte: Metadaten (Seitenangabe)

suchen	<td
ersetzen	<td class="seite"
xpath	//td[1]

### 2.4.)

- zweite Spalte: Metadaten (Zeile)

suchen	<td
ersetzen	<td class="zeile"
xpath	//td[2]

### 2.5.)

- Tabelle: class umbenennen

suchen	class="MsoNormalTable"
ersetzen	class="TableTranskription"
xpath	//table

### 2.6.)

(in Layoutansicht eines HTML-Editors bearbeiten und Menüs nutzen)

- erste Seitenüberschrift: <td class="head\_page\_first"
- wiederholte Seitenüberschrift: <td class="head\_page\_repeat"
- Kapitelüberschrift: <td class="head\_chapter"
- Spalte, Morgen/Rthl: <td class="morgen" (Tip: ganze Tabelle (alles) markieren und „ruten“ später überschreiben )
- Spalte, zentriert, normale Schriftgrößen (Kartensymbole, Heufuder, u.a.): <td class="zentriert\_12"
- Spalte, zentriert, kleine Schriftgröße: <td class="zentriert\_10"
- Spalte, Ruten/ß: <td class="ruten"
- Spalte, „transport, Summe“: <td class="summe"

### 3.1) [Verlinkungen überprüfen:]

- nach „links“ und „rechts“ suchen und überprüfen, ob richtig eingetragen (ein Link ohne Zeilenumbruch)
- Folioangaben in getrennte Zellen umkopieren, falls notwendig

### 4.1)

- Kopien der bisherigen bearbeiteten Dateien sichern

### 4.2)

- Formatierung: Breitenangaben löschen

suchen	width:[^;]*;
ersetzen	
xpath	//@style

suchen	width:[^"]*"
ersetzen	"

xpath	//@style
-------	----------

suchen	width="[^"]*"
ersetzen	
xpath	//table

5.)

- Formatierung: Höhenangaben löschen

suchen	height:[^;]*;
ersetzen	
xpath	//@style

suchen	height:[^"]*"
ersetzen	"
xpath	//@style

suchen	height="[^"]*"
ersetzen	
xpath	//table

6.)

- Formatierung: alle padding-Angaben löschen

suchen	padding[^;]*;
ersetzen	
xpath	//@style

suchen	padding[^"]*"
ersetzen	"
xpath	//@style

7.)

- Formatierung: alle border „none“-Angaben löschen

suchen	border[^;]* none;
ersetzen	
xpath	//@style

8.)

- Formatierung: alle „windowtext“ border - Angaben löschen

suchen	windowtext
ersetzen	
xpath	//@style

9.)

- Formatierung: alle „1.0pt“ border - Angaben löschen

suchen	1.0pt
ersetzen	1px
xpath	//td/@style

10.)

- Formatierung: alle „1.5pt“ border - Angaben löschen

suchen	1.5pt
ersetzen	4px
xpath	//td/@style

11.)

- Formatierung: alle „2.25pt“ border - Angaben löschen

suchen	2.25pt
ersetzen	4px
xpath	//td/@style

12.a)

- Formatierung: alle valign-Einträge löschen

suchen	valign="[^\"]*"
ersetzen	
xpath	//td

12.b)

- Formatierung: alle align-Einträge löschen

suchen	align="[^\"]*"
ersetzen	
xpath	//td

13.)

- Formatierung: alle leeren style-Angaben löschen

suchen	*style= *" *"?>
ersetzen	>
xpath	

14.)

- Formatierung: alle überflüssigen Leerzeichen in style löschen

suchen	style=" *
ersetzen	style="
xpath	

suchen	;*
ersetzen	;
xpath	//td/@style

suchen	;*"
ersetzen	;"
xpath	

suchen	;*;
ersetzen	;
xpath	

15.)

- Formatierung: alle leerer Attribute löschen

suchen	style=""
ersetzen	

xpath	
-------	--

suchen	title=""
ersetzen	
xpath	

16.)

- Element: <p> innerhalb tabel löschen

suchen	<p[^>]*>
ersetzen	
xpath	//table

suchen	</p>
ersetzen	
xpath	

17.)

- Element: <span> innerhalb tabel löschen

suchen	<span[^>]*>
ersetzen	
xpath	//table

suchen	</span>
ersetzen	
xpath	

18.)

- Element: <b> innerhalb tabel löschen

suchen	<b>
ersetzen	
xpath	//table

suchen	</b>
ersetzen	
xpath	

19.)

- Entity: &nbsp; innerhalb tabel löschen

suchen	&nbsp;
ersetzen (1 Leerzeichen)	
xpath	//table

20.)

- Element: <br /> vor </td> innerhalb tabel löschen

suchen	<br */>\s*</td>
ersetzen	</td>
xpath	//td

21.)

Oxygen: Dokument formatieren Strg+Umschalt+P

22.)

- überflüssige Leerzeichen im Text löschen (nur td betreffend)
- Leerzeichen vor und nach <a>, <sup>, <img> dabei

suchen	<td([^\<]*)> *
ersetzen	<td\$1>
xpath	//table

suchen	*</td>
ersetzen	</td>
xpath	//table

23.)

- leeres Element <td/> mit Leerzeichen füllen

suchen	\s*</td>
ersetzen	</td>
xpath	//table

suchen	<td([^\<]*)> *</td>
ersetzen	<td\$1>&nbsp;</td>
xpath	//table

24.)

- @style in <table> löschen

suchen	<table (.*) style="[^"]*">
ersetzen	<table \$1>
xpath	//table

25.)

- **per Hand löschen:**

suchen	<a name
ersetzen	
xpath	

26.)

Oxygen: Dokument formatieren Strg+Umschalt+P

27.)

- Ergebnis mit doc-Datei vergleichen und zusätzlich notwendige Formatierungen nachtragen wie
  - durchgestrichener Text: <strike> (doc mit „Suchen“ durchsuchen)
  - unterstrichener Text: <u>
  - fetter Text: <b>
  - zentriert <td align="center">
  - rechts <td align="right">
  - Abstände in px: <td width= (Abstände für Transport o. Summe etwa 300, 350, 400, 450px usw.)
  - Schriftgröße: <span style="font-size:10.0pt">

28.)

in beiden Browsern Ergebnis testen

- evtl. @width-Einträge (nicht in @style) in <td> vornehmen um Abstände anzugleichen  
(wenn sich Spalten mit großen Abständen über eine Zeile verteilen, dann letzte Zelle der  
zeile mit width bearbeiten);

29.)

- Rahmenfarbe bei Seitenüberschriftwdhungen „black“