

xhtml-Konvertierung und Validierung der Transkriptionen.

im DFG-Projekt Schwedische Landesaufnahme von Pommern 1692-1709

1. Inhalte in html konvertieren

Zuerst die Word-Vorlage aus dem Pfad P:\08 Schwedische Landesaufnahme\08WebTexte\word_Textvorlagen\ öffnen. Dann Microsoft Expression Web öffnen und eine neue Datei erstellen. Inhalt aus der Word Datei komplett markieren und in eine neue html Datei ziehen (nicht kopieren/einfügen). Link zum Stylesheet einbinden: `<link rel="stylesheet" type="text/css" href=".././darstellung.css" />` Das jeweilige Programm und Prozedere sind notwendig, weil sonst die Tabellen nicht korrekt erhalten bleiben.

2. Strukturieren

Fußnoten mitsamt des umschließenden DIV-Containers kopieren, entfernen und in die neue html-Datei einfügen. iFrame-Vorlage öffnen (siehe Anm. 1) und Namen der Orte austauschen. Die iFrame-Vorlage als NAME_trans.htm abspeichern. Die Inhaltsdatei als NAME_trans_txt.htm abspeichern. Die Fußnoten als NAME_trans_anno.htm abspeichern

Hinweis: Die folgenden Arbeitsschritte 3-5 können parallel bearbeitet werden.

3. Tabellen korrigieren

Teilweise werden bei der Konvertierung Tabellenlinien falsch, bzw. lückenhaft gezogen. Hier eignet es sich am Besten, die Wordvorlage und die html-Version nebeneinander zu öffnen und die Tabellen abzugleichen. In der darstellung.css-Datei existieren eine Reihe von Korrektursätzen (fix1-11), mit denen die Linien korrigiert werden können. Dazu in Expression Web: "Bereiche/ Formatvorlagen verwalten/" anzeigen, die fehlerhaften Tabellenzellen markieren und mit Rechtsklick auf fix1-11 (je nach Problem) "Formatvorlage übernehmen" auswählen.

4. Text korrigieren

Die Überschriften der einzelnen Tabellen werden oft nicht korrekt umgebrochen, ebenso wie einzelne Tabellenköpfe oder Arealgrößen. Die Zeilenumbrüche in Kopfzeilen müssen manuell korrigiert werden. Die Ausrichtung innerhalb einer Tabellenzelle (links, mitte, rechts) wird oft nicht korrekt übernommen. Die Ausrichtung muss ebenfalls manuell korrigiert werden.

5. Bilder einfügen

Bei der Konvertierung werden die vorhandenen Grafiken durch ein Microsoft-eigenes Vektorsystem ersetzt. Dieses ist nicht xhtml-valide und führt zu Fehlermeldungen (fehlerhafter Code siehe Anm. 2). Die automatisch generierten Bilder sind daher mitsamt des umgebenden Codes durch folgende

entsprechenden Grafiken zu ersetzen: P:\08 Schwedische Landesaufnahme\08WebTexte\Symbole
Hinweis: Die separaten NAME_trans_anno.htm-Dateien enthalten oft ebenfalls Bilder.

6. Validieren

Die Konvertierung aus Word führt zu einer Reihe überflüssiger Codefragmente welche die Validierung verhindern. Die automatische Entfernung in Expression Web: Extras/html optimieren/html entfernen/word html funktioniert jedoch nur rudimentär. Der Quellcode muss daher manuell von überflüssigen Codefragmenten befreit werden. Mit der Funktion Extras/Kompatibilitätsreports und der Einstellung "xhtml-Strict" nach Fehlern suchen und diese entfernen. Dazu in Expression Web innerhalb des Quellcodes mit strg+f das suchen/ersetzen-Fenster öffnen. In das "suchen nach"-Feld den nicht validen Code hineinkopieren, das "ersetzen durch"-Feld bleibt leer. Am besten lässt sich der Code durch die Nutzung regulärer Ausdrücke beseitigen (siehe Anm. 3). Vorher ein Backup der Datei anlegen. Dann "alle ersetzen" (Zur Zeitersparnis: dies auch in mehreren parallel geöffneten Dokumenten möglich). Sämtliche Eventualitäten für überflüssigen Code können hier nicht behandelt werden, zur Endkontrolle daher mit Oxygen die Datei validieren und die letzten Restfragmente entfernen.

Anm. 1

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">

<head>
<meta content="text/html; charset=utf-8" http-equiv="Content-Type" />
<title>Transkription: #Name</title>
<link rel="stylesheet" type="text/css" href=" ../darstellung.css" />
</head>

<body>
<div id="divcontain">
<div id="divhaupt">
<iframe src="katzenow_trans_txt.htm" frameborder="0" name="hauptframe"
target="fussnoten">
<p>Ihr Browser kann leider keine eingebetteten Frames anzeigen.</p>
</iframe>
</div>

<div id="divfuss">
<iframe src="katzenow_trans_anno.htm" frameborder="0" name="fussnoten" scrolling="no"
target="hauptframe">
<p>Ihr Browser kann leider keine eingebetteten Frames anzeigen.</p>
</iframe>
```

```
</div>
</div>

</body>
</html>
```

Anm. 2

Beispiel: Nicht valider Bilder-Code (If bis EndIf)

```
<!--[if gte vml 1]><v:shapetype id=""_x0000_t75"" coordsize=""21600,21600"" o:spt=""75""
o:preferrelative=""t"" path=""m@4@5l@4@11@9@11@9@5xe"" filled=""f"" stroked=""f"">

  <v:stroke jointstyle=""miter"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
  <v:formulas>
    <v:f eqn=""if lineDrawn pixelLineWidth 0"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""sum @0 1 0"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""sum 0 0 @1"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""prod @2 1 2"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""prod @3 21600 pixelWidth"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""prod @3 21600 pixelHeight"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""sum @0 0 1"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""prod @6 1 2"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""prod @7 21600 pixelWidth"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""sum @8 21600 0"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""prod @7 21600 pixelHeight"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
    <v:f eqn=""sum @10 21600 0"" xmlns:v=""urn:schemas-microsoft-com:vml""/>
  </v:formulas>

  <v:path o:extrusionok=""f"" gradientshapeok=""t"" o:connecttype=""rect""
xmlns:v=""urn:schemas-microsoft-com:vml""/> <o:lock v:ext=""edit"" aspectratio=""t""
xmlns:o=""urn:schemas-microsoft-com:office:office""/> </v:shapetype><v:shape
id=""_x0000_i1025"" type=""#_x0000_t75"" style='width:21.75pt; height:9.75pt'> <v:imagedata
src=""file:///C:\TEMP\msohtmlclip1\01\clip_image001.jpg"" o:title=""134 Symbol 12pt""
xmlns:v=""urn:schemas-microsoft-com:vml""/>

</v:shape><![endif]--><![if !vml]>

  <img border=""0"" height=""13"" src=""file:///C:/TEMP/msohtmlclip1/01/clip_image002.jpg""
v:shapes=""_x0000_i1025"" width=""29""><![endif]>
```

Anm. 3

Reguläre Ausdrücke

<o:p>|<\/o:p>|<\/!if !supportFootnotes>|<hr align=\"left\" size=\"1\" width=\"33%\" \>|<\/!endif>| style=\"mso-element: footnote\" | style=\"mso-spacerun: yes\" | style=\"mso-bidi-font-style: normal\" | font-family: \"Times New Roman\";\"serif\"; mso-fareast-font-family: \"Times New Roman\"; mso-ansi-language: DE; mso-fareast-language: DE; mso-bidi-language: AR-SA | style=\"mso-element: footnote-list\" | style=\"mso-special-character: footnote\" | style=\"mso-footnote-id: ftn[0-9]\" | style=\"mso-footnote-id: ftn[0-9][0-9]\" | lang=\"SV\" | style=\"mso-ansi-language: SV\" | clear=\"all\" | style=\"mso-bidi-font-size: 1[0-9].0pt\" | style=\"mso-bidi-font-weight: normal\" | lang=\"EN-GB\" | style=\"mso-ansi-language: EN-GB\" | style=\"mso-bidi-font-weight: bold\" | style=\"mso-bookmark: OLE_LINK2\" | style=\"mso-footnote-id: ftn[0-9]\" | style=\"mso-footnote-id: ftn[0-9][0-9]\" | style=\"mso-special-character: footnote\" | style=\"mso-bidi-font-weight: normal\" |<\/!if !supportFootnotes>|<\/!endif>| style=\"mso-spacerun: yes\" | style=\"font-family: \"Courier New\";\" | font-family: \"Arial\";\"sans-serif\"; |<st1:place w:st=\"on\">|<\/st1:place> | style=\"mso-bookmark: OLE_LINK[0-9]\" | style=\"mso-bookmark: OLE_LINK[0-9][0-9]\" | style=\"font-family: \"Sylfaen\";\"serif\"; mso-bidi-font-family: Sylfaen\" | style=\"font-family: \"Arial\";\"sans-serif\"; color: blue\" | mso-padding-alt: [0-9]cm [0-9].[0-9]pt [0-9]cm [0-9].[0-9]pt | style=\"mso-bidi-font-weight: bold\" | style=\"font-family: \"Arial\";\"sans-serif\";\" | mso-table-layout-alt: fixed; | style=\"mso-bidi-font-size: 1[0-9].0pt\" | mso-bidi-font-size: 1[0-9].0pt; | mso-fareast-font-family: \"Times New Roman\"; | mso-bidi-font-weight: bold | lang=\"EN-US\" | mso-ansi-language: EN-US | style=\"mso-special-character: line-break; page-break-before: always\" | mso-cell-special: placeholder; | clear=\"all\" | clear=\"all\" | mso-border-bottom-alt: double windowtext 2\.25pt; | mso-border-bottom-alt: solid #969696 \.5pt; | mso-border-top-alt: solid #969696 \.5pt; | mso-border-right-alt: double black 2\.25pt; | mso-border-right-alt: solid windowtext \.5pt; | mso-border-right-alt: solid #969696 \.5pt; | mso-border-left-alt: solid windowtext \.5pt; | mso-border-top-alt: double windowtext 2\.25pt; | mso-border-left-alt: double windowtext 2\.25pt; | mso-border-bottom-alt: solid windowtext 1\.0pt; | mso-border-right-alt: double windowtext 2\.25pt; | mso-border-left-alt: solid windowtext 1\.0pt; | mso-border-bottom-alt: solid windowtext \.5pt; | mso-border-top-alt: solid windowtext \.5pt; | mso-border-right-alt: solid black \.5pt; | mso-border-left-alt: solid windowtext 1\.0pt; | mso-border-top-alt: solid windowtext 1\.0pt; | mso-border-top-alt: solid windowtext 1\.0pt; | mso-border-bottom-alt: double black 2\.25pt; | mso-border-bottom-alt: solid #999999 \.5pt; | nowrap

mso-border-top-alt: solid #999999 \.5pt; | mso-border-bottom-alt: dot-dash windowtext \.5pt; | mso-border-left-alt: double windowtext 1\.5pt; | mso-border-top-alt: dot-dash windowtext \.5pt; | mso-border-left-alt: solid #969696 \.5pt; | mso-border-top-alt: double windowtext 1\.5pt; | mso-border-left-alt: solid silver \.5pt; | mso-border-top-alt: solid silver \.5pt; | mso-border-right-alt: solid silver \.5pt; | mso-border-bottom-alt: solid silver \.5pt; | mso-border-bottom-alt: solid black \.5pt; | mso-border-bottom-alt: double windowtext 1\.5pt; | mso-border-bottom-alt: solid windowtext 1\.5pt; | mso-border-top-alt: solid windowtext 1\.5pt; | mso-border-top-alt: double black 2\.25pt; | mso-border-left-alt: double black 2\.25pt; | mso-border-left-alt: solid #999999 \.5pt; | mso-border-right-alt: solid #999999 \.5pt; | mso-border-alt: solid #969696 \.5pt; | mso-border-bottom-alt: dot-dot-dash windowtext \.5pt; | mso-border-right-alt: double

windowtext 1\.5pt\;|mso-border-alt\: solid windowtext \.5pt\;|mso-border-left-alt\: solid black \.5pt\;|mso-border-right-alt\: solid windowtext \.75pt\;|mso-border-color-alt\: windowtext\;|mso-border-style-alt\: solid\;|mso-border-left-alt\: solid windowtext \.75pt\;|mso-border-top-alt\: \.5pt\;|mso-border-left-alt\: \.75pt\;|mso-border-bottom-alt\: \.5pt\;|mso-border-right-alt\: \.75pt\;|mso-border-right-alt\: solid windowtext\;|mso-border-bottom-alt\: dot-dash windowtext\;|mso-border-left-alt\: \#999999\;|mso-border-top-alt\: \#999999\;|mso-border-bottom-alt\: windowtext\;|mso-border-right-alt\: windowtext\;|mso-border-width-alt\: \.5pt\;|lang\="EN-GB"|lang\="SV"|lang\="DE"|mso-ansi-language\: EN-GB|mso-ansi-language\: DE|mso-ansi-language\: SV|

mso-bidi-language\: AR-SA|mso-fareast-language\: DE;|mso-yfti-irow\: [0-9][0-9]\;|mso-yfti-firstrow\: yes\;|mso-yfti-lastrow\: yes\;|mso-yfti-irow\: [0-9]\;|mso-footnote-id\: ftn[0-9]\;|mso-footnote-id\: ftn[0-9][0-9]\;

\\>[

zu

[

tab-stops\: [0-9][0-9]\.[0-9][0-9]pt [0-9][0-9]\.[0-9][0-9]cm [0-9][0-9]\.[0-9][0-9]cm [0-9][0-9]\.[0-9][0-9]cm [0-9][0-9]\.[0-9][0-9]cm

tab-stops\: [0-9][0-9]\.[0-9][0-9]pt [0-9]\.[0-9]cm [0-9]\.[0-9]cm [0-9][0-9]\.[0-9]cm [0-9][0-9]\.[0-9]cm

<s>

#zu

#und

</s>

#zu

<![if !supportMisalignedColumns]>

#undfolgende

text-underline: single;

#zu

text-decoration:underline;

<u> zu

</u> zu

align="right"

#zu

style="text-align: right"

#sowie

#center, left

<s> zu

</s> zu