# ETD Batch Processing – Part 1 Prepare ProQuest XML for IRO Ingest

## Copy PDFs to SDRC

This needs to be done by someone with access to this location. This stepped moved up in the process simply to be adjacent to other things currently done by Wendy.

* + 1. Open L:\sdrc.lib.uiowa.edu\ETDUploads\etd\ in Windows Explorer
    2. Open the current semester’s PDF folder in a different Windows Explorer window
    3. Copy all the PDFs from the etd folder to the SDRC folder

## Edit Bepress Conversion XSL

Note: Search for XXX in the xsl to find all the places that need modification

### Copy the master file elsewhere

* + 1. Copy the conversion file (ETD-ProQuestXML2bepressXML.xsl) from L:\etd.lib.uiowa.edu\etd\xsl-master to Desktop to your desk or documents (or L drive or One Drive).  
       Note: You can delete the file after the transformation is done (or after the files have been uploaded to IRO)
    2. Open local version of the xsl in Oxygen (so that the edits for each batch saved will not appear in the master file)

### Add the graduation month and season

* + 1. Replace XXX in the publication-date for the correct month; in summer check PDF for July or August
    2. Add the season (replace XXX)

### Add list of optimized files

* + 1. Copy all filenames in the “pdf” “large” sub-folder :
       1. Select all the files (Ctrl- A)
       2. Shift+right-click + copy as path
    2. Paste into Excel (showing the file names of pdfs that have been optimized)
       1. Use find and replace to delete the path and folder name (everything up through large\)
       2. Copy formula =CONCATENATE("DISS\_content/DISS\_binary='",A1,"' or ") into cell B1 and fill down (to fill down, paste in B1 move cursor to lower Right corner of cell and double click on + )
    3. Combine into a single statement
       1. In cell C1, add formula =CONCATENATE(TRANSPOSE(B1:B29))   
          [replace B29 with the last cell with a value]
       2. Do NOT press Enter
       3. Select the TRANSPOSE(B1:B29) portion and press F9 (This replaces the Transpose section with the result)
       4. Remove curly brackets { and }, and the final “ or ”, leave final “.

(Microsoft made a change sept. 2017 so step e was added. It is adding quotations around each cells value)

* + - 1. Select entire product including = then press F9.
      2. NOW PRESS ENTER
      3. See <http://chandoo.org/wp/2014/01/13/combine-text-values-quick-tip/> for more information
    1. *Old method: Copy into Word for cleanup*
       1. *Copy text from column B and paste special in Word as unformatted text*
       2. *Remove paragraph breaks using Find ^p and Replace [blank]*
       3. *Remove the final “or” and preceding space*
       4. *Highlight text, and copy for pasting into XSL*
    2. Insert optimization data into XSL file
       1. In Oxygen search the open ProquestXML2bepressXML.xsl file for the string "DISS\_content/DISS\_binary='XXX PASTE HERE'"> (it will appear ***twice,*** once in the comment and once in the full-text URL)
       2. Insert list of optimized files between the DOUBLE quotation marks in both instances, replacing all text currently between the double quotation marks

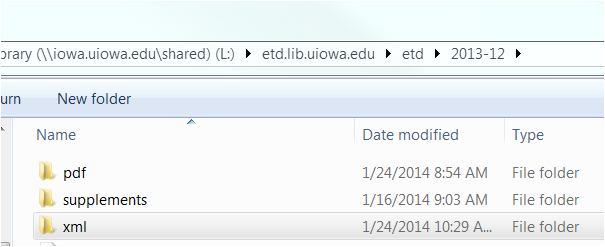
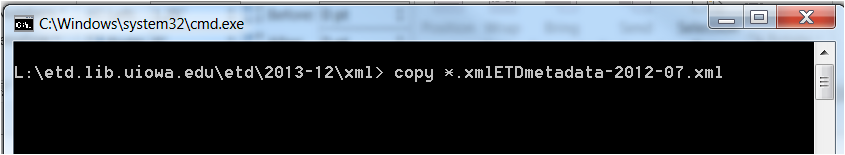
### Define the start of file numbering sequence

* + 1. Determine the number of the last thesis already added to bepress
    2. Find xsl:value-of select="position()+XXX in the conversion file by searching for “position(
    3. Replace the number here with the file number of the **last thesis** added to bepress

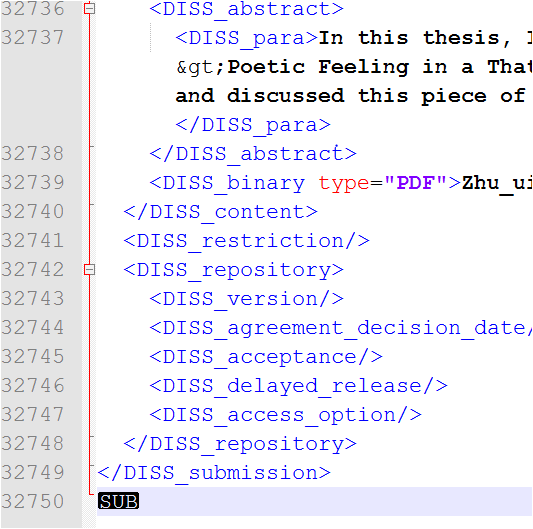
### Save the xsl copy (being careful not to alter the master)

## Combine XML Into a Single File

### Concatenate the XML files

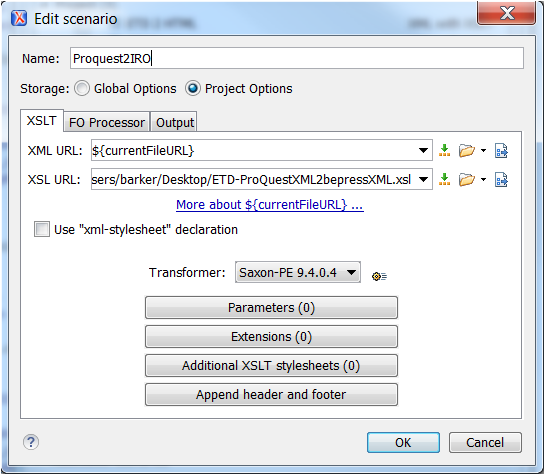
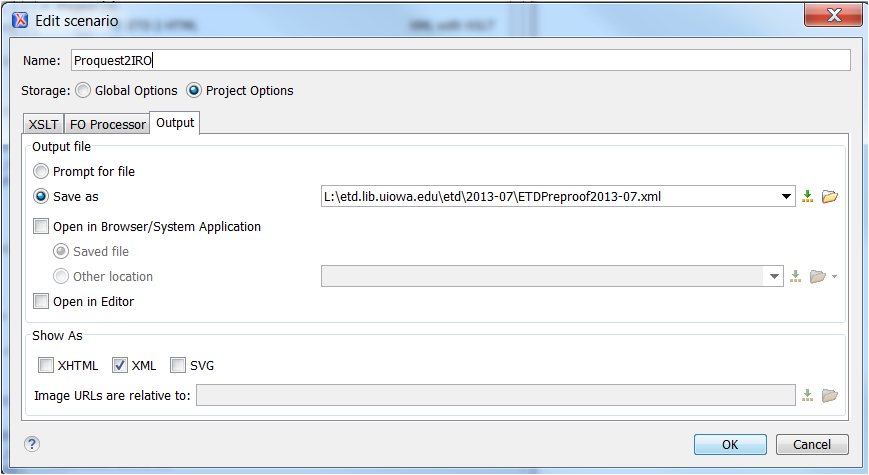
* + 1. Select xml folder in L:/etd.lib.uiowa.edu/[2013-12]
    2. Shift and right-click choose “open command window here”
    3. Type: copy \*.xml XXX.xml (replacing XXX with the file name you want, e.g. ETDmetadata-2012-07.xml ). Press enter. Close command window when it has finished running.  
       
    4. Move your file from its spot inside L:/etd.lib.uiowa.edu/(2013-12)/xml up   
       one level to L:/etd.lib.uiowa.edu/(2013-12)

### Correct structural elements of XML so that it is properly formed

* + 1. Open your new ETDmetadata file in Notepad++ or Textpad. Use Find and Replace, remove all of the <?xml version="1.0" encoding="UTF-8"?> headers and then put back one occurrence at the top. (You can copy this from the first line of the file, do the replace and then paste back in). There should now only be one occurrence of <?xml in the file.
    2. Add <xml> on second line, after the <?xml...> element and close at the end with </xml>.
    3. Make sure the first line begins <?xml and the last line is </xml> (you never want blank lines at the top or bottom of an xml file)
    4. Make sure to delete any line at the bottom that says “sub”  
       
    5. Save file (as ETDmetadata[date].xml )
    6. Open in Oxygen and fix any red-light errors (e.g. remove extra space at beginning/end ; no final period) and save.

## Create a Pre-Proof File Structured for IRO

### Transform the metadata file to Bepress format

* + 1. Open in Oxygen both the metadata xml and the revised xsl file(xsl from desktop)
    2. With the xml file selected to view over xsl file, click the wrench and select the ETDProQuest-IRO option
    3. Edit, with XSL URL pointing to the revised xsl file (on Desktop, not the master file version)
    4. Make sure Transformer is set to SaxonPE  
       
    5. Name the output file (click on tab) something like ETDpreproof-2009-07.xml  
        
    6. Click “Apply Associated” (file will automatically be saved—see message at bottom of screen Red or green note will tell you if you are successful or not.)
    7. Open new file, format-and-indent 
    8. File may show red errors

### Double check the transformation

* + 1. Check the value in the first and last “label” field to make sure it is what you expected. Make sure they are unused numbers and the numbers you expected. It is far easier to recreate the file now than to fix these problems later
    2. Check one of the optimized files to make sure it has transformed correctly, both note and url form.
    3. Double check date and season on one.

### Fix errors with disciplines

Unauthorized disciplines show as red errors. They all begin “FIX and NOTIFY Wendy:” followed by the department name.

* + 1. If our local department has an & in it, the transformation won’t work because Wendy doesn’t know how to get it to work. You will need to fix these manually. Do not notify Wendy. E.g. change “FIX and NOTIFY Wendy: Rehabilitation &amp; Counselor Education” to “Vocational Rehabilitation Counseling”
    2. With one such as “FIX and NOTIFY Wendy: Anatomy &amp; Cell Biology” you may add a second discipline ex:   
        <disciplines>  
        <discipline>Cell Anatomy</discipline>  
        <discipline>Cell Biology</discipline>  
        </disciplines>
    3. Mapping from our local department names to bepress disciplines is in this spreadsheet: <https://sharepoint.uiowa.edu/sites/libraries/cat/Work%20Documentation/IRO%20Metadata/UIowa-depts-mapped-to-bepress-discipline.xlsx?Web=1>
    4. If the mapping is incorrect, let Wendy know (may need to check IRO collections)
    5. If mapping is absent (new department), check official bepress list of disciplines at <http://www.bepress.com/disciplines.xsd>
    6. Notify Amanda to add new mapping to xsl so that next time it will happen automatically. See also step “VII Review list of departments”
    7. You will need to pick something appropriate for interdisciplinary studies (you can include more than one) or you can omit the discipline
    8. Save the XML once disciplines have been fixed

From MAY 2017 to Summer 2018 – for the corrections below, if you find no occurrences of the problem, add a note in a comment. After Summer 2018 we should be caught up on embargoes so all future etds will be freshly submitted (i.e. no technical changes in submission in a single batch). We are probably checking things that we don’t need to be now that we are getting UTF-8 files from ProQuest. By noting what hasn’t appeared in the files in over 5 semesters, we can better judge what steps we may be able to omit in the future.

### Fix errors with entities

This step will identify entities that were improperly formatted by the graduate and therefore did not convert to Unicode. (An entity should begin with an & and end with ; - the final ; is sometimes omitted.)

In XML, there are 5 characters (predefined entities) used by the code so if you want the actual character, you must use either a character or numeric entity. Either form will work correctly; they do not need to be altered.

* + - 1. Ampersand - &amp; OR &#38;
      2. Less than sign - &lt; OR &#60;
      3. Greater than sign - &gt; OR &#62;
      4. Double quotation mark &quot; OR &#34;
      5. Apostrophe - &apos; OR &#39;
    1. Find and replace &amp; with & (no space following)
    2. Find and replace & [**followed by a space**] with &amp; [**followed by a space**] – this will correct most occurrences of an actual &
    3. Fix any that are really & that are not followed by spaces (i.e. change & back to &amp;)
    4. Correct any red errors. See [Appendix 2](#_Appendix_2:_Common)
    5. Correct italics
       1. Use <em> </em> (will not display in html file that we create for proofreading)
       2. Look for &lt;em&gt; and replace with <em>
       3. Look for &lt;/em&gt; and replace with </em>
       4. Look for &lt;italics&gt; and replace with <em>
       5. Look for &lt;/italics&gt; and replace with </em>
       6. Look for &lt;i&gt; and replace with <em>
       7. Look for &lt;/i&gt; and replace with </em>
       8. If you don’t replace the same number of occurrences, check the <em> </em> pairings in the file
    6. It causes an upload problem if < as a character does not follow a space. See the next step for super and subscripts.
       1. Ctrl+F
       2. Click regular expression
       3. &lt;[0-9.b-z]
       4. Find all
       5. Add a space after the semicolon
       6. DO change this in the html formatted note in the comments (the search above does not include &lt;a because this is in the comments and is fine)
       7. Repeat the search with &#60;[0-9.b-z]
       8. Note the author and label for all those which have had the abstract altered so that IRO can be corrected after upload (i.e. to remove the space). The XML file should also be corrected before conversion to MARC.
    7. Do a find all for &# - these are numeric entities that didn’t convert to UTF-8 so we may want to fix them now, e.g. &#945; is α
    8. When a character is unclear, check in the pdf abstract. When a graduate avoids a complex character with words, leave as is (5 - 15 solar radii. Vs 5−15RSol [circle with dot in it]
    9. Save all error fixes

### Fix superscript and subscript in the abstract as submitted

* + 1. <sup> and <sub> will upload and display correctly in the IRO abstract, but will not convert correctly to MARC. Be sure to look for this format as well: &lt;sub&gt;2&lt;/sub&gt;
    2. If a Unicode version of the character exists, manually change to that character.
       1. The characters are spread out over all the Unicode charts. Look for additional characters with the [Unicode character name index](http://www.unicode.org/charts/charindex.html). Always be careful to retain the case of the letter (in both Latin and Greek)
       2. See [Appendix 1](#_Appendix_1:_Superscripts) below for all the characters
    3. If a superscript or subscript version of a character doesn’t exist, leave the abstract with the html tags so that it displays correctly in IRO. For example, change &lt;sup&gt;-/-&lt;/sup&gt; to <sup>-/-</sup>.

### Fix other formatting errors in abstract (RARE)

We do not want to introduce errors to the abstract as we try to improve it. We also should be cognizant of the amount of time perfecting an abstract will take vs. moving more efficiently though the process. We largely accept the abstract as it was submitted by the author. Do not seek out every possible difference. The author chose to submit the abstract as we got it and we should trust their publishing decision.

* + 1. Sometimes the abstract submitted to ProQuest/IRO is different than the one in PDF (either intentional or due to later modifications). Accept the version of the abstract submitted by the graduate to ProQuest.
    2. Sometimes a public abstract has been included, not the formal one. If the formal abstract is full of LATEX formatting, leave the public abstract in place. If the formal abstract would be easy to add, copy the public abstract to its own field and add the formal abstract
    3. If you happen to notice a public abstract, you may add it to the public abstract field. Do not seek them out; they will be added eventually.
    4. Save all error fixes

### Look for non-standard/untested characters

This step looks for characters that have not been tested for upload. It also can identify a Unicode character used incorrectly (e.g. a line from a box used for an em-dash) and spot other “funny characters” to replace with Unicode character (or numeric entity). This step will strip all letters, numbers and keyboard-accessible-symbols from the xml file (find and replace all).

* + 1. Save the file.
    2. Make a copy.
    3. Open the copy in Notepad++
    4. Do a replace with nothing, with regular expression checked
       1. \w (all standard characters including Beta– could do as [a-z] and [0-9] instead)
       2. \n (line feed)
       3. \t (tab)
       4. [`~@ #$%^&\*()-\_=+{}\|;:'",/’“”±—°–≤≥©®¹²³ºª⁰ⁱ⁴⁵⁶⁷⁸⁹⁺⁻⁼⁽⁾ⁿ₀₁₂₃₄₅₆₇₈₉₊₋₌₍₎ₐₑₒₓₔ →√−×· ‐]
    5. This may leave a few other characters. If they turn out to be fine (through testing to IRO demo site and test upload to OCLC), you can add them to the replace string above. They may look like a safe character but be a problem (such as ─ / &#9472; / [http://www.fileformat.info/info/unicode/char/2500/index.htm](http://www.fileformat.info/info/unicode/char/2500/index.htm%20) used instead of em-dash or ˚ (ring above modifier letter) instead of a degree sign) The non-breaking space (  00A0)may still be in the results but it is OK.
    6. You can copy the characters to word, highlight a single character, ALT-X to see what the Unicode character is.
    7. You can look at <http://demo.uiowa2.bepress.com/test_etd/> and <http://demo.uiowa2.bepress.com/achv2010/1/> to find characters we have tested that are OK but not in the string to copy paste above (since we haven’t run into them in an actual etd)
    8. Also remember that the character may be fine but it simply does not display in Notepad++ or in Oxygen (or on the web) due to a lack of a font. These are NOT problems; the data is fine.
    9. Copy the problematic character and search for it in the XML file.
    10. Make necessary correction (or test)
    11. Save all error fixes
    12. Delete the copied file.

### Fix simple problems with LATEX encoding

Make these changes when they are not embedded with a lot of other LATEX

* + 1. Find and replace double grave `` with “ (&#8220; left double quotation mark)
    2. Find and replace double apostrophe ‘’ with ” (&#8221; right double quotation mark)
    3. Find and replace ` with ‘ (&#8216; (left single quotation mark)). Look for the pair and change to ’ (right single quotation mark)
    4. Search for $\$ to find dollar signs, e.g. $\$ 1.2$ vs $1:2
    5. Change \% to %
    6. Look for {\ to find italics: {\it TEXT}
    7. Look for \ used for diacritics and change to Unicode: e.g., Flatt\'e –
    8. LATEX uses -- for an en dash (but many people use – for either an en or em dash) and it uses --- for an emdash. The double hyphens can be ignored. Change the triple hyphen to —
    9. Save all error fixes

### ~~Optionally: Fix problems with LATEX encoding (for mathematics)~~

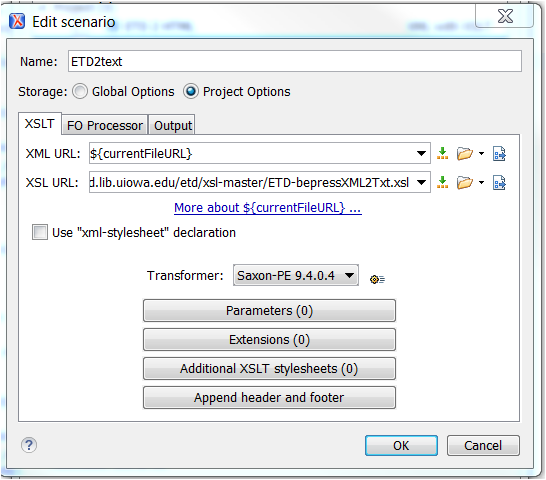
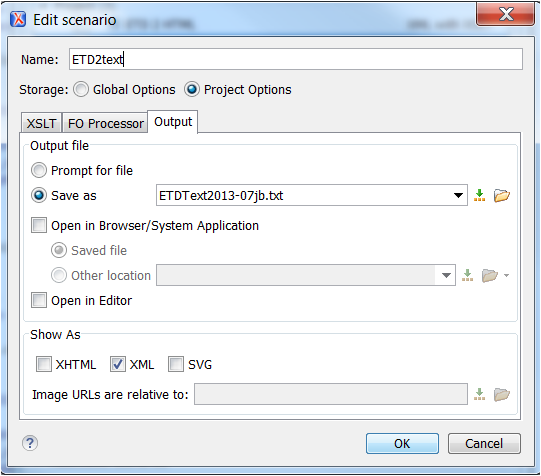
[~~http://www.rpi.edu/dept/arc/docs/latex/latex-intro.pdf~~](http://www.rpi.edu/dept/arc/docs/latex/latex-intro.pdf) ~~for more info~~

~~Be sure not to introduce errors into the abstract when making modification. Do not spend a lot of time on this; people that can understand the content of the abstract can understand LATEX so it is acceptable to leave it as is.~~As of 21 July 2017 CMD meeting, we will NOT be fixing most LATEX coding.

* + 1. ~~Refer to the end of w3schools.com~~ [~~http://www.w3schools.com/charsets/default.asp~~](http://www.w3schools.com/charsets/default.asp) ~~for list of entities~~
    2. ~~Mathematical symbols are included at the end of~~ [~~http://www.w3schools.com/charsets/ref\_utf\_math.asp~~](http://www.w3schools.com/charsets/ref_utf_math.asp)
    3. ~~Search for $ to identify problems - $ indicates entering math mode~~
    4. ~~Some problems can be fixed with global replaces, e.g. $beta replaced by the beta symbol or the numeric entity~~
    5. ~~Pay attention to case (check the pdf if in doubt)~~
    6. ~~Save all error fixes~~

### Identify additional problems with abstracts and use of all capital letters

#### Convert XML to text

* + - 1. Open the master file ETD-bepressXML2Txt.xsl in Oxygen
      2. With cursor in the open Preproof xml file, click on the wrench
      3. Select ETD2Text option
      4. Edit, with XSL URL pointing to the master xsl file   
         
      5. Output file should point to the current ETD batch (year/month) and named something like ETDtxtDEC2009.txt  
         
      6. Apply Assoc. (file will automatically be saved)
      7. Red or green note at bottom of screen will tell you if you are successful or not.

#### Check for errors in Excel

* + - 1. Open blank worksheet, select “from text” from Data menu
      2. Navigate to the text file and import
      3. Change file origin from Windows (ANSI) to Unicode (UTF-8)
      4. Accept other defaults and after it has rendered save file as something like ETDtxtAug2011.txt
      5. Review text for oddities in Excel, then find and edit oddities in the Preproof file
         1. TITLE field: remove any formatting in the title by searching column for angle brackets e.g. <italic> <em> <bold> <sup> <sub> or other (these will be part of the title as indexed so html display codes should not be in that field)
         2. Advisor, Department, Language: check for non-standard terminology or formatting (e.g. all caps)
         3. ABSTRACT field: find and replace things like n/a, no abstract required, not necessary, or even the heading ABSTRACT
         4. Save all edits you just made to the preproof file.

#### Notify Institutional Repository Librarian that Excel file has been made

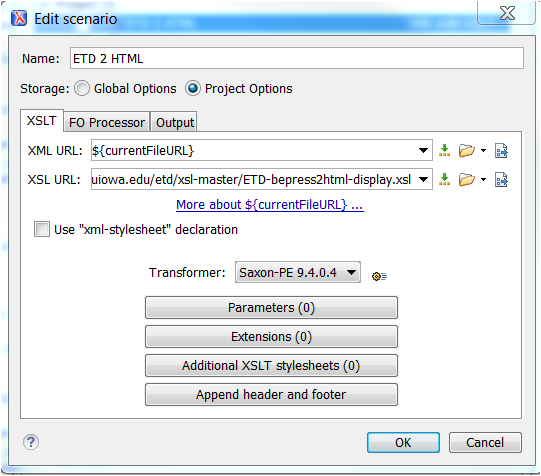
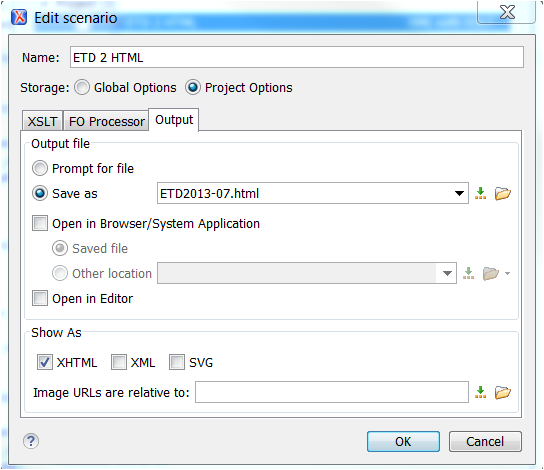
* + - 1. Copy file to have only author, title, dept, graduation date, and embargo date
      2. Note graduates with Theatre Arts portfolios
      3. Note graduate with DMA recordings
      4. Check departments for the titles with supplemental items
         1. Note film studies graduates with videos
         2. Note any other creative works
      5. Send list to Erin at the Grad College to triple check the embargo dates
      6. Ask Erin about Film studies or other creative works with supplemental content that is NOT embargoed so that we can determine if the AV can be posted openly, limited to campus, or not posted at all

### Other comments

* + 1. Symbol fonts do not upload correctly—will require a global replace with the correct Unicode value
    2. Correct Unicode characters may not display in Oxygen or Notepad++. This is fine. However, if the correct Unicode character does not display in a standard html page due to a lack of an appropriate font, we may want to change the character to something that will display (or leave as is since that is how it was submitted and eventually html fonts will add more characters). We have not run into this situation.
    3. Some problems may not be visible until after upload. A problematic character in an abstract will stop the rest of that paragraph from uploading.
    4. Another way to fix problems in abstract would be to copy the text from the pdf into Word, make the necessary replacements, then copy the corrected abstract into the revise form in IRO and also into Connexion. Be careful not to replace an intentionally simplified abstract meant for display on a website with a complex one with a lot of formatting issues.

## Create HTML file

### Convert preproof file to html

* + 1. With the preproof file open in Oxygen
    2. Open master file ETD-bepress2html-display.xsl in Oxygen
    3. With cursor in the preproof xml file, click on the wrench
    4. Select ETD2HTML option and click Edit  
       
    5. Output file should be Saved in ETD folder as something like ETD-DEC09.html  
       
    6. Apply Assoc. (will automatically save file)
    7. Red or green note at bottom of screen will tell you if you are successful or not.

### Spot and fix formatting errors

* + 1. Open the HTML file and look for abstracts with paragraph breaks in the middle of a sentence or run on paragraphs.
    2. Delete or add the paragraph breaks in the preproof file.
       1. Highlight a paragraph.
       2. Replace <p> with nothing on selected lines
       3. Replace </p> with nothing on selected lines
       4. Highlight paragraph and then CTRL+E.
       5. Type p in the box to specify tag.
       6. Click OK
       7. Repeat as needed.
    3. Fix any remaining character issues (note: Title display is more important than Abstract for funny characters)
    4. Check against pdf file if necessary

### Convert to a clean html file

* + 1. Once all problems have been fixed, make a new html file and delete the old
    2. Make a final check of the new HTML file

## Prepare XML files for Copy Catalogers

* + 1. Copy html file to L:\librarydata\all staff\for CTS-ETDs
    2. Using XML Notepad, divide the preproof file into separate sections as equal in size as possible for each copy cataloger and place in For CTS-ETDs
       1. Divide the total number of PDFs by 3.
       2. Add the split to the number in label field to find the numbers to split on
       3. When making a new file, be sure to begin with <?xml version="1.0" encoding="UTF-8"?> and the entire <documents> field.
       4. Close with </documents>
    3. Notify copy catalogers of file location and instruct them to create a working ETD folder on their H-drive with shortcut to Desktop (this will ensure that any wipeouts can be restored)
    4. Set a due date for them to send files from their H-drive folder

## Review list of Departments

### New/renamed degree

* + 1. Add it to <https://sharepoint.uiowa.edu/sites/libraries/cat/Work%20Documentation/IRO%20Metadata/UIowa-depts-mapped-to-bepress-discipline.xlsx?Web=1>
    2. Modify xsl for disciplines
    3. Modify MARC xml if necessary (if not a simple Department of…)
    4. Ask catalogers which set it should go into

### Full list of degrees

* + 1. Let the cataloger’s know how many they will be getting and in what departments are in the science set. If a cataloger would like to add a specific subject to their list, modify the MARC.
    2. Use VLOOKUP as the following: =VLOOKUP(I2,'[etd-degree-division-for-marc-cataloging.xlsx]Sheet1'!$A:$B,2,FALSE) and see <https://sharepoint.uiowa.edu/sites/libraries/cat/Work%20Documentation/ETD's%2C%20Theses%2C%20Dissertations/etd-degree-division-for-marc-cataloging.xlsx?Web=1> for the cataloging division
    3. Use PIVOTTABLE to get account of title by cataloger groups

## Appendix 1: Superscripts and Subscripts

### Most common superscripts

<http://www.unicode.org/charts/PDF/U0080.pdf>

|  |  |
| --- | --- |
| ¹ | SUPERSCRIPT ONE |
| ² | SUPERSCRIPT TWO |
| ³ | SUPERSCRIPT THREE |
| º | MASCULINE ORDINAL INDICATOR |
| ª | FEMININE ORDINAL INDICATOR |

### Most other superscripts/subscripts

<http://unicode.org/charts/PDF/U2070.pdf>

|  |  |
| --- | --- |
| ⁰ | SUPERSCRIPT ZERO |
| ⁱ | SUPERSCRIPT LATIN SMALL LETTER I [Unicode charts incorrectly have regular i] |
| ⁴ | SUPERSCRIPT FOUR |
| ⁵ | SUPERSCRIPT FIVE |
| ⁶ | SUPERSCRIPT SIX |
| ⁷ | SUPERSCRIPT SEVEN |
| ⁸ | SUPERSCRIPT EIGHT |
| ⁹ | SUPERSCRIPT NINE |
| ⁺ | SUPERSCRIPT PLUS SIGN |
| ⁻ | SUPERSCRIPT MINUS |
| ⁼ | SUPERSCRIPT EQUALS SIGN |
| ⁽ | SUPERSCRIPT LEFT PARENTHESIS |
| ⁾ | SUPERSCRIPT RIGHT PARENTHESIS |
| ⁿ | SUPERSCRIPT LATIN SMALL LETTER N |
| ₀ | SUBSCRIPT ZERO |
| ₁ | SUBSCRIPT ONE |
| ₂ | SUBSCRIPT TWO |
| ₃ | SUBSCRIPT THREE |
| ₄ | SUBSCRIPT FOUR |
| ₅ | SUBSCRIPT FIVE |
| ₆ | SUBSCRIPT SIX |
| ₇ | SUBSCRIPT SEVEN |
| ₈ | SUBSCRIPT EIGHT |
| ₉ | SUBSCRIPT NINE |
| ₊ | SUBSCRIPT PLUS SIGN |
| ₋ | SUBSCRIPT MINUS |
| ₌ | SUBSCRIPT EQUALS SIGN |
| ₍ | SUBSCRIPT LEFT PARENTHESIS |
| ₎ | SUBSCRIPT RIGHT PARENTHESIS |
| ₐ | LATIN SUBSCRIPT SMALL LETTER A |
| ₑ | LATIN SUBSCRIPT SMALL LETTER E [displays in Oxygen as superscript 1] |
| ₒ | LATIN SUBSCRIPT SMALL LETTER O |
| ₓ | LATIN SUBSCRIPT SMALL LETTER X |
| ₔ | LATIN SUBSCRIPT SMALL LETTER SCHWA |

### Problem characters

Note that these characters, while correct, do not display in Oxygen, Word, or in a browser due to the lack of a font to view them:

|  |  |
| --- | --- |
| ₕ | LATIN SUBSCRIPT SMALL LETTER H |
| ₖ | LATIN SUBSCRIPT SMALL LETTER K |
| ₗ | LATIN SUBSCRIPT SMALL LETTER L |
| ₘ | LATIN SUBSCRIPT SMALL LETTER M |
| ₙ | LATIN SUBSCRIPT SMALL LETTER N |
| ₚ | LATIN SUBSCRIPT SMALL LETTER P |
| ₛ | LATIN SUBSCRIPT SMALL LETTER S |
| ₜ | LATIN SUBSCRIPT SMALL LETTER T |

### A few additional ones

<http://www.unicode.org/charts/PDF/U1D00.pdf>

|  |  |
| --- | --- |
| ᵢ | LATIN SUBSCRIPT SMALL LETTER I |
| ᵣ | LATIN SUBSCRIPT SMALL LETTER R |
| ᵤ | LATIN SUBSCRIPT SMALL LETTER U |
| ᵥ | LATIN SUBSCRIPT SMALL LETTER V |
| ᵦ | GREEK SUBSCRIPT SMALL LETTER BETA |
| ᵧ | GREEK SUBSCRIPT SMALL LETTER GAMMA |
| ᵨ | GREEK SUBSCRIPT SMALL LETTER RHO |
| ᵩ | GREEK SUBSCRIPT SMALL LETTER PHI |
| ᵪ | GREEK SUBSCRIPT SMALL LETTER CHI |

## Appendix 2: Common ISO-8859-1 Numeric and Entity Names

Most of these were pulled from this table:  
<http://www.freeformatter.com/html-entities.html#iso88591-symbols>

### ASCII Characters (Printable)

|  |  |  |  |
| --- | --- | --- | --- |
| Character | Entity Name | Entity Number | Description |
|  | &nbsp; | &#32; | Space |
| ! |  | &#33; | Exclamation mark |
| " |  | &#34; | Quotation mark |
| # |  | &#35; | Number sign |
| $ |  | &#36; | Dollar sign |
| % |  | &#37; | Percent sign |
| & | &amp; | &#38; | Ampersand |
| ' |  | &#39; | Apostrophe |
| ( |  | &#40; | Opening/Left Parenthesis |
| ) |  | &#41; | Closing/Right Parenthesis |
| \* |  | &#42; | Asterisk |
| + |  | &#43; | Plus sign |
| , |  | &#44; | Comma |
| - |  | &#45; | Hyphen |
| . |  | &#46; | Period |
| / |  | &#47; | Slash |
| 0 |  | &#48; | Digit 0 |
| 1 |  | &#49; | Digit 1 |
| 2 |  | &#50; | Digit 2 |
| 3 |  | &#51; | Digit 3 |
| 4 |  | &#52; | Digit 4 |
| 5 |  | &#53; | Digit 5 |
| 6 |  | &#54; | Digit 6 |
| 7 |  | &#55; | Digit 7 |
| 8 |  | &#56; | Digit 8 |
| 9 |  | &#57; | Digit 9 |
| : |  | &#58; | Colon |
| ; |  | &#59; | Semicolon |
| < | &lt; | &#60; | Less-than |
| = |  | &#61; | Equals sign |
| > | &gt; | &#62; | Greater than |
| ? |  | &#63; | Question mark |
| @ |  | &#64; | At sign |
| A |  | &#65; | Uppercase A |
| B |  | &#66; | Uppercase B |
| C |  | &#67; | Uppercase C |
| D |  | &#68; | Uppercase D |
| E |  | &#69; | Uppercase E |
| F |  | &#70; | Uppercase F |
| G |  | &#71; | Uppercase G |
| H |  | &#72; | Uppercase H |
| I |  | &#73; | Uppercase I |
| J |  | &#74; | Uppercase J |
| K |  | &#75; | Uppercase K |
| L |  | &#76; | Uppercase L |
| M |  | &#77; | Uppercase M |
| N |  | &#78; | Uppercase N |
| O |  | &#79; | Uppercase O |
| P |  | &#80; | Uppercase P |
| Q |  | &#81; | Uppercase Q |
| R |  | &#82; | Uppercase R |
| S |  | &#83; | Uppercase S |
| T |  | &#84; | Uppercase T |
| U |  | &#85; | Uppercase U |
| V |  | &#86; | Uppercase V |
| W |  | &#87; | Uppercase W |
| X |  | &#88; | Uppercase X |
| Y |  | &#89; | Uppercase Y |
| Z |  | &#90; | Uppercase Z |
| [ |  | &#91; | Opening/Left square bracket |
| \ |  | &#92; | Backslash |
| ] |  | &#93; | Closing/Right square bracket |
| ^ |  | &#94; | Caret |
| \_ |  | &#95; | Underscore |
| ` |  | &#96; | Grave accent |
| a |  | &#97; | Lowercase a |
| b |  | &#98; | Lowercase b |
| c |  | &#99; | Lowercase c |
| d |  | &#100; | Lowercase d |
| e |  | &#101; | Lowercase e |
| f |  | &#102; | Lowercase f |
| g |  | &#103; | Lowercase g |
| h |  | &#104; | Lowercase h |
| i |  | &#105; | Lowercase i |
| j |  | &#106; | Lowercase j |
| k |  | &#107; | Lowercase k |
| l |  | &#108; | Lowercase l |
| m |  | &#109; | Lowercase m |
| n |  | &#110; | Lowercase n |
| o |  | &#111; | Lowercase o |
| p |  | &#112; | Lowercase p |
| q |  | &#113; | Lowercase q |
| r |  | &#114; | Lowercase r |
| s |  | &#115; | Lowercase s |
| t |  | &#116; | Lowercase t |
| u |  | &#117; | Lowercase u |
| v |  | &#118; | Lowercase v |
| w |  | &#119; | Lowercase w |
| x |  | &#120; | Lowercase x |
| y |  | &#121; | Lowercase y |
| z |  | &#122; | Lowercase z |
| { |  | &#123; | Opening/Left curly brace |
| | |  | &#124; | Vertical bar |
| } |  | &#125; | Closing/Right curly brace |
| ~ |  | &#126; | Tilde |

### ISO-8859-1 Characters

|  |  |  |  |
| --- | --- | --- | --- |
| Character | Entity Name | Entity Number | Description |
| À | &Agrave; | &#192; | Capital a with grave accent |
| Á | &Aacute; | &#193; | Capital a with acute accent |
| Â | &Acirc; | &#194; | Capital a with circumflex accent |
| Ã | &Atilde; | &#195; | Capital a with tilde |
| Ä | &Auml; | &#196; | Capital a with umlaut |
| Å | &Aring; | &#197; | Capital a with ring |
| Æ | &AElig; | &#198; | Capital ae |
| Ç | &Ccedil; | &#199; | Capital c with cedilla |
| È | &Egrave; | &#200; | Capital e with grave accent |
| É | &Eacute; | &#201; | Capital e with acute accent |
| Ê | &Ecirc; | &#202; | Capital e with circumflex accent |
| Ë | &Euml; | &#203; | Capital e with umlaut |
| Ì | &Igrave; | &#204; | Capital i with grave accent |
| Í | &Iacute; | &#205; | Capital i with accute accent |
| Î | &Icirc; | &#206; | Capital i with circumflex accent |
| Ï | &Iuml; | &#207; | Capital i with umlaut |
| Ð | &ETH; | &#208; | Capital eth (Icelandic) |
| Ñ | &Ntilde; | &#209; | Capital n with tilde |
| Ò | &Ograve; | &#210; | Capital o with grave accent |
| Ó | &Oacute; | &#211; | Capital o with accute accent |
| Ô | &Ocirc; | &#212; | Capital o with circumflex accent |
| Õ | &Otilde; | &#213; | Capital o with tilde |
| Ö | &Ouml; | &#214; | Capital o with umlaut |
| Ø | &Oslash; | &#216; | Capital o with slash |
| Ù | &Ugrave; | &#217; | Capital u with grave accent |
| Ú | &Uacute; | &#218; | Capital u with acute accent |
| Û | &Ucirc; | &#219; | Capital u with circumflex accent |
| Ü | &Uuml; | &#220; | Capital u with umlaut |
| Ý | &Yacute; | &#221; | Capital y with acute accent |
| Þ | &THORN; | &#222; | Capital thorn (Icelandic) |
| ß | &szlig; | &#223; | Lowercase sharp s (German) |
| à | &agrave; | &#224; | Lowercase a with grave accent |
| á | &aacute; | &#225; | Lowercase a with acute accent |
| â | &acirc; | &#226; | Lowercase a with circumflex accent |
| ã | &atilde; | &#227; | Lowercase a with tilde |
| ä | &auml; | &#228; | Lowercase a with umlaut |
| å | &aring; | &#229; | Lowercase a with ring |
| æ | &aelig; | &#230; | Lowercase ae |
| ç | &ccedil; | &#231; | Lowercase c with cedilla |
| è | &egrave; | &#232; | Lowercase e with grave accent |
| é | &eacute; | &#233; | Lowercase e with acute accent |
| ê | &ecirc; | &#234; | Lowercase e with circumflex accent |
| ë | &euml; | &#235; | Lowercase e with umlaut |
| ì | &igrave; | &#236; | Lowercase i with grave accent |
| í | &iacute; | &#237; | Lowercase i with acute accent |
| î | &icirc; | &#238; | Lowercase i with circumflex accent |
| ï | &iuml; | &#239; | Lowercase i with umlaut |
| ð | &eth; | &#240; | Lowercase eth (Icelandic) |
| ñ | &ntilde; | &#241; | Lowercase n with tilde |
| ò | &ograve; | &#242; | Lowercase o with grave accent |
| ó | &oacute; | &#243; | Lowercase o with acute accent |
| ô | &ocirc; | &#244; | Lowercase o with circumflex accent |
| õ | &otilde; | &#245; | Lowercase o with tilde |
| ö | &ouml; | &#246; | Lowercase o with umlaut |
| ø | &oslash; | &#248; | Lowercase o with slash |
| ù | &ugrave; | &#249; | Lowercase u with grave accent |
| ú | &uacute; | &#250; | Lowercase u with acute accent |
| û | &ucirc; | &#251; | Lowercase u with circumflex accent |
| ü | &uuml; | &#252; | Lowercase u with umlaut |
| ý | &yacute; | &#253; | Lowercase y with acute accent |
| þ | &thorn; | &#254; | Lowercase thorn (Icelandic) |
| ÿ | &yuml; | &#255; | Lowercase y with umlaut |

### ISO-8859-1 Symbols

|  |  |  |  |
| --- | --- | --- | --- |
| Symbol | Entity Name | Entity Number | Description |
|  | &nbsp; | &#160; | Non-breaking space |
| ¡ | &iexcl; | &#161; | Inverted exclamation mark |
| ¢ | &cent; | &#162; | Cent |
| £ | &pound; | &#163; | Pound |
| ¤ | &curren; | &#164; | Currency |
| ¥ | &yen; | &#165; | Yen |
| ¦ | &brvbar; | &#166; | Broken vertical bar |
| § | &sect; | &#167; | Section |
| ¨ | &uml; | &#168; | Spacing diaeresis |
| © | &copy; | &#169; | Copyright |
| ª | &ordf; | &#170; | Feminine ordinal indicator |
| « | &laquo; | &#171; | Opening/Left angle quotation mark |
| ¬ | &not; | &#172; | Negation |
| ­ | &shy; | &#173; | Soft hyphen |
| ® | &reg; | &#174; | Registered trademark |
| ¯ | &macr; | &#175; | Spacing macron |
| ° | &deg; | &#176; | Degree |
| ± | &plusmn; | &#177; | Plus or minus |
| ¹ | &#185; | &sup1; | superscript 1 |
| ² | &sup2; | &#178; | Superscript 2 |
| ³ | &sup3; | &#179; | Superscript 3 |
| ´ | &acute; | &#180; | Spacing acute |
| µ | &micro; | &#181; | Micro |
| ¶ | &para; | &#182; | Paragraph |
| ¸ | &cedil; | &#184; | Spacing cedilla |
| ¹ | &cedil; | &#185; | Superscript 1 |
| º | &ordm; | &#186; | Masculine ordinal indicator |
| » | &raquo; | &#187; | Closing/Right angle quotation mark |
| ¼ | &frac14; | &#188; | Fraction 1/4 |
| ½ | &frac12; | &#189; | Fraction 1/2 |
| ¾ | &frac34; | &#190; | Fraction 3/4 |
| ¿ | &iquest; | &#191; | Inverted question mark |
| × | &times; | &#215; | Multiplication |
| ÷ | &divide; | &#247; | Divide |

### Math Symbols

|  |  |  |  |
| --- | --- | --- | --- |
| Symbol | Entity Name | Entity Number | Description |
| ∀ | &forall; | &#8704; | For all |
| ∂ | &part; | &#8706; | Part |
| ∃ | &exist; | &#8707; | Exist |
| ∅ | &empty; | &#8709; | Empty |
| ∇ | &nabla; | &#8711; | Nabla |
| ∈ | &isin; | &#8712; | Is in / Element Of |
| ∉ | &notin; | &#8713; | Not in |
| ∋ | &ni; | &#8715; | Ni |
| ∏ | &prod; | &#8719; | Product |
| ∑ | &sum; | &#8721; | Sum |
| − | &minus; | &#8722; | Minus |
| ∗ | &lowast; | &#8727; | Asterisk (Lowast) |
| √ | &radic; | &#8730; | Square root |
| ∝ | &prop; | &#8733; | Proportional to |
| ∞ | &infin; | &#8734; | Infinity |
| ∠ | &ang; | &#8736; | Angle |
| ∧ | &and; | &#8743; | And |
| ∨ | &or; | &#8744; | Or |
| ∩ | &cap; | &#8745; | Cap |
| ∪ | &cup; | &#8746; | Cup |
| ∫ | &int; | &#8747; | Integral |
| ∴ | &there4; | &#8756; | Therefore |
| ∼ | &sim; | &#8764; | Similar to |
| ≅ | &cong; | &#8773; | Congurent to |
| ≈ | &asymp; | &#8776; | Almost equal |
| ≠ | &ne; | &#8800; | Not equal |
| ≡ | &equiv; | &#8801; | Equivalent |
| ≤ | &le; | &#8804; | Less or equal |
| ≥ | &ge; | &#8805; | Greater or equal |
| ⊂ | &sub; | &#8834; | Subset of |
| ⊃ | &sup; | &#8835; | Superset of |
| ⊄ | &nsub; | &#8836; | Not subset of |
| ⊆ | &sube; | &#8838; | Subset or equal |
| ⊇ | &supe; | &#8839; | Superset or equal |
| ⊕ | &oplus; | &#8853; | Circled plus |
| ⊗ | &otimes; | &#8855; | Circled times |
| ⊥ | &perp; | &#8869; | Perpendicular |
| ⋅ | &sdot; | &#8901; | Dot operator |

### Greek Letters

|  |  |  |  |
| --- | --- | --- | --- |
| Letter | Entity Name | Entity Number | Description |
| Α | &Alpha; | &#913; | Alpha |
| Β | &Beta; | &#914; | Beta |
| Γ | &Gamma; | &#915; | Gamma |
| Δ | &Delta; | &#916; | Delta |
| Ε | &Epsilon; | &#917; | Epsilon |
| Ζ | &Zeta; | &#918; | Zeta |
| Η | &Eta; | &#919; | Eta |
| Θ | &Theta; | &#920; | Theta |
| Ι | &Iota; | &#921; | Iota |
| Κ | &Kappa; | &#922; | Kappa |
| Λ | &Lambda; | &#923; | Lambda |
| Μ | &Mu; | &#924; | Mu |
| Ν | &Nu; | &#925; | Nu |
| Ξ | &Xi; | &#926; | Xi |
| Ο | &Omicron; | &#927; | Omicron |
| Π | &Pi; | &#928; | Pi |
| Ρ | &Rho; | &#929; | Rho |
| Σ | &Sigma; | &#931; | Sigma |
| Τ | &Tau; | &#932; | Tau |
| Υ | &Upsilon; | &#933; | Upsilon |
| Φ | &Phi; | &#934; | Phi |
| Χ | &Chi; | &#935; | Chi |
| Ψ | &Psi; | &#936; | Psi |
| Ω | &Omega; | &#937; | Omega |
| α | &alpha; | &#945; | alpha |
| β | &beta; | &#946; | beta |
| γ | &gamma; | &#947; | gamma |
| δ | &delta; | &#948; | delta |
| ε | &epsilon; | &#949; | epsilon |
| ζ | &zeta; | &#950; | zeta |
| η | &eta; | &#951; | eta |
| θ | &theta; | &#952; | theta |
| ι | &iota; | &#953; | iota |
| κ | &kappa; | &#954; | kappa |
| λ | &lambda; | &#955; | lambda |
| μ | &mu; | &#956; | mu |
| ν | &nu; | &#957; | nu |
| ξ | &xi; | &#958; | xi |
| ο | &omicron; | &#959; | omicron |
| π | &pi; | &#960; | pi |
| ρ | &rho; | &#961; | rho |
| ς | &sigmaf; | &#962; | sigmaf |
| σ | &sigma; | &#963; | sigma |
| τ | &tau; | &#964; | tau |
| υ | &upsilon; | &#965; | upsilon |
| φ | &phi; | &#966; | phi |
| χ | &chi; | &#967; | chi |
| ψ | &psi; | &#968; | psi |
| ω | &omega; | &#969; | omega |
| ϑ | &thetasym; | &#977; | Theta symbol |
| ϒ | &upsih; | &#978; | Upsilon symbol |
| ϖ | &piv; | &#982; | Pi symbol |

### Miscellaneous HTML entities

|  |  |  |  |
| --- | --- | --- | --- |
| Symbol | Entity Name | Entity Number | Description |
| Œ | &OElig; | &#338; | Uppercase ligature OE |
| œ | &oelig; | &#339; | Lowercase ligature OE |
| Š | &Scaron; | &#352; | Uppercase S with caron |
| š | &scaron; | &#353; | Lowercase S with caron |
| Ÿ | &Yuml; | &#376; | Capital Y with diaeres |
| ƒ | &fnof; | &#402; | Lowercase with hook |
| ˆ | &circ; | &#710; | Circumflex accent |
| ˜ | &tilde; | &#732; | Small Tilde |
|  | &ensp; | &#8194; | En space |
|  | &emsp; | &#8195; | Em space |
|  | &thinsp; | &#8201; | Thin space |
| ‌ | &zwnj; | &#8204; | Zero width non-joiner |
| ‍ | &zwj; | &#8205; | Zero width joiner |
| ‎ | &lrm; | &#8206; | Left-to-right mark |
| ‏ | &rlm; | &#8207; | Right-to-left mark |
| – | &ndash; | &#8211; | En dash |
| — | &mdash; | &#8212; | Em dash |
| ‘ | &lsquo; | &#8216; | Left single quotation mark |
| ’ | &rsquo; | &#8217; | Right single quotation mark |
| ‚ | &sbquo; | &#8218; | Single low-9 quotation mark |
| “ | &ldquo; | &#8220; | Left double quotation mark |
| ” | &rdquo; | &#8221; | Right double quotation mark |
| „ | &bdquo; | &#8222; | Double low-9 quotation mark |
| † | &dagger; | &#8224; | Dagger |
| ‡ | &Dagger; | &#8225; | Double dagger |
| • | &bull; | &#8226; | Bullet |
| … | &hellip; | &#8230; | Horizontal ellipsis |
| ‰ | &permil; | &#8240; | Per mille |
| ′ | &prime; | &#8242; | Minutes (Degrees) |
| ″ | &Prime; | &#8243; | Seconds (Degrees) |
| ‹ | &lsaquo; | &#8249; | Single left angle quotation |
| › | &rsaquo; | &#8249; | Single right angle quotation |
| ‾ | &oline; | &#8254; | Overline |
| € | &euro; | &#8364; | Euro |
| ™ | &trade; | &#8482; | Trademark |
| ← | &larr; | &#8592; | Left arrow |
| ↑ | &uarr; | &#8593; | Up arrow |
| → | &rarr; | &#8594; | Right arrow |
| ↓ | &darr; | &#8595; | Down arrow |
| ↔ | &harr; | &#8596; | Left right arrow |
| ↵ | &crarr; | &#8629; | Carriage return arrow |
| ⌈ | &lceil; | &#8968; | Left ceiling |
| ⌉ | &rceil; | &#8969; | Right ceiling |
| ⌊ | &lfloor; | &#8970; | Left floor |
| ⌋ | &rfloor; | &#8971; | Right floor |
| ◊ | &loz; | &#9674; | Lozenge |
| ♠ | &spades; | &#9824; | Spade |
| ♣ | &clubs; | &#9827; | Club |
| ♥ | &hearts; | &#9829; | Heart |
| ♦ | &diams; | &#9830; | Diamond |