

## INSTRUCTIONS FOR PREPARING TEXTS FOR ADDITION TO TESSERAE

### FOUR STEPS

Tesserae can accept texts for addition to its intertextual search engine provided they have the proper provenance and format.

The steps in preparing a text are to:

- 1) identify the most suitable text available
- 2) format the text properly
- 3) send the formatted text and provenance document for review, and
- 4) check the returned Tesserae text for errors against the edition it was based on.

### IDENTIFICATION OF TEXT AND PROVENANCE DOCUMENT

A statement must be included indicating the electronic source of the text and edition. The statement should also confirm that there are no intellectual property problems that would conflict with posting it on the site. This information should be in a document separate from the text itself.

Provenance information ideally includes the editor, publisher, and year published of a paper edition represented by the electronic version. In some cases, this may require digging. For example, in the case of a Project Gutenberg text of Hamlet, we would like the following statement:

“Source: Project Gutenberg public domain text  
(<http://www.gutenberg.org/cache/epub/1524/pg1524.txt>), downloaded October 2, 2012. Based on edition by [editor], published by [publisher] [year].”

Filling in the blanks on the edition might be possible using the header information:

```
Hamlet, Prince of Denmark  
  
by William Shakespeare [Collins edition]  
  
November, 1998 [Etext #1524]  
Most recently updated: January 26, 2009
```

“Collins edition” gives a clue. Is the publisher Harper Collins? Who is the editor? When was it published?

In all cases, we should make a good faith effort to get the edition information and include as much as we can.

## PROPER FORMAT

Tesserae texts need to be a simple format: plain text files (.txt extension), with markers of the text location down the left side, separated from the text by tabs. These markers are not used for matching, only for identifying locations to the user after a set of matches is produced.

The following is an example of the beginning of a Tesserae text file, for Vergil's *Aeneid*.

```
<verg. aen. 1.1> Arma virumque cano, Troiae qui primus ab oris
<verg. aen. 1.2> Italiam, fato profugus, Laviniaque venit
<verg. aen. 1.3> litora, multum ille et terris iactatus et alto
<verg. aen. 1.4> vi superum saevae memorem Iunonis ob iram;
<verg. aen. 1.5> multa quoque et bello passus, dum conderet urbem,
<verg. aen. 1.6> inferretque deos Latio, genus unde Latinum,
<verg. aen. 1.7> Albanique patres, atque altae moenia Romae.
```

The abbreviation and location for the work are given in front of each line. In this case, since the *Aeneid* is an epic in 12 books, each book of the epic is a separate file, but the book and line number are still given for each line, since this exact information will be displayed to the user viewing matches.

The markers will differ in the case of works composed of many shorter poems, as in the following example from Catullus.

```
<cat. 1.1> Cui dono lepidum novum libellum
<cat. 1.2> arido modo pumice expolitur?
<cat. 1.3> Corneli, tibi; namque tu solebas
<cat. 1.4> meas esse aliquid putare nugas,
<cat. 1.5> iam tum cum ausus es unus Italorum
<cat. 1.6> omne aevum tribus explicare chartis,
<cat. 1.7> doctis, Iuppiter, et laboriosis!
<cat. 1.8> quare habe tibi quidquid hoc libelli
<cat. 1.9> qualecumque, quod, o patrona virgo,
<cat. 1.10> plus uno maneat perenne saeclo.

<cat. 2.1> Passer, deliciae meae puellae,
<cat. 2.2> quicum ludere, quem in sinu tenere,
<cat. 2.3> cui primum digitum dare adpetenti
<cat. 2.4> et acris solet incitare morsus,
<cat. 2.5> cum desiderio meo nitenti
<cat. 2.6> carum nescio quid libet iocari
<cat. 2.7> (et solaciolum sui doloris,
<cat. 2.8> credo, ut tum gravis adquiescat ardor),
<cat. 2.9> tecum ludere sicut ipsa possem
<cat. 2.10> et tristis animi levare curas!
<cat. 2.11> Tam gratum est mihi quam ferunt puellae
<cat. 2.12> pernici aureolum fuisse malum,
<cat. 2.13> quod zonam solvit diu ligatam.
```

In this case, the markers indicate first the poem number then the line number. Note also that according to convention, there is no indication of specific work for Catullus.

Whereas we have several works from Vergil, so the work must be indicated in the marker, the extant poems of Catullus were transmitted together in one manuscript, making an indication of the poetic work unnecessary. For classical texts, such conventional abbreviations can be found in the opening pages of the Oxford Classical Dictionary. Whatever the segmenting conventions are for a given work, these should be reflected in the location markers in the Tesseract text.

## HOW TO FORMAT A TEXT OR PREPARE IT FOR FORMATTING

Whoever is submitting a text to be incorporated must decide on what the segmentation of the text will be, and how it will be represented in the markers. Tesseract will soon be adding prose texts, so an example this process would be the addition of Cicero's *Pro Archia*. The following is taken from the web version of the text at the Perseus website (the primary source for Tesseract classical texts), converted to plain text with footnotes removed manually.

1. si quid est in me ingeni, iudices, quod sentio quam sit exiguum, aut si qua exercitatio dicendi, in qua me non infitior mediocriter esse versatum, aut si huiusce rei ratio aliqua ab optimarum artium studiis ac disciplina profecta, a qua ego nullum confiteor aetatis meae tempus abhorruisse, earum rerum omnium vel in primis hic A. Licinius fructum a me repetere prope suo iure debet. nam quoad longissime potest mens mea respicere spatium praeteriti temporis et pueritiae memoriam recordari ultimam, inde usque repetens hunc video mihi principem et ad suscipiendam et ad ingrediendam rationem horum studiorum exstitisse. quod si haec vox huius hortatu praeceptisque conformata non nullis aliquando saluti fuit, a quo id accepimus quo ceteris opitulari et alios servare possemus, huic profecto ipsi, quantum est situm in nobis, et opem et salutem ferre debemus. [2] ac ne quis a nobis hoc ita dici forte miretur, quod alia quaedam in hoc facultas sit ingeni neque haec dicendi ratio aut disciplina, ne nos quidem huic uni studio penitus umquam dediti fuimus. etenim omnes artes quae ad humanitatem pertinent habent quoddam commune vinclum et quasi cognatione quadam inter se continentur.

This is section 1 of the speech, a fact we would know from the paratext on the webpage, also indicated by the "1" at the beginning of the text. The "[2]" in brackets indicates the second part of section 1. We find no "[1]" indicating the first part of section 1 because according to convention the section number indicates both the beginning of the whole section and the beginning of its first part.

To make the text usable to Tesseract, we must indicate each of these sections in some consistent way. Here we would mark the text as follows.

```
<cic. arch. 1.1> si quid est in me ingeni, iudices, quod sentio quam
sit exiguum, aut si qua exercitatio dicendi, in qua me non infitior
mediocriter esse versatum, aut si huiusce rei ratio aliqua ab
optimarum artium studiis ac disciplina profecta, a qua ego nullum
confiteor aetatis meae tempus abhorruisse, earum rerum omnium vel in
primis hic A. Licinius fructum a me repetere prope suo iure debet.
nam quoad longissime potest mens mea respicere spatium praeteriti
temporis et pueritiae memoriam recordari ultimam, inde usque
repetens hunc video mihi principem et ad suscipiendam et ad
```

ingrediendam rationem horum studiorum exstitisse. quod si haec vox huius hortatu praeceptisque conformata non nullis aliquando saluti fuit, a quo id accepimus quo ceteris opitulari et alios servare possemus, huic profecto ipsi, quantum est situm in nobis, et opem et salutem ferre debemus.

<cic. arch. 1.2> ac ne quis a nobis hoc ita dici forte miretur, quod alia quaedam in hoc facultas sit ingeni neque haec dicendi ratio aut disciplina, ne nos quidem huic uni studio penitus umquam dediti fuimus. etenim omnes artes quae ad humanitatem pertinent habent quoddam commune vinculum et quasi cognatione quadam inter se continentur.

The previous text location markers (“1”, “[2]”) were deleted, and the Tesserae markers with appropriate location information inserted separated from the text by a tab.

In this case, the reformatting of the text was performed manually, but in practice for whole texts it can be performed automatically. If those who are adding texts have the necessary computational skills to perform this operation and deliver a properly formatted text, this would be most helpful. But users without such proficiency need only prepare the text in some consistent way with adequate markers. To return to an example above, the following would be an adequate system for representing *Aeneid* book 1, assuming first that each book of the epic is a separate text file.

Start at <verg. aen. 1.1> and number continuously by line to end of file.

Arma virumque cano, Troiae qui primus ab oris  
 Italiam, fato profugus, Laviniaque venit  
 litora, multum ille et terris iactatus et alto  
 vi superum saevae memorem Iunonis ob iram;  
 multa quoque et bello passus, dum conderet urbem,  
 inferretque deos Latio, genus unde Latinum,  
 Albanique patres, atque altae moenia Romae.

Catullus would require somewhat more information.

Start at <cat. 1.1> and increase the second (line) number by one for each new line. After two consecutive line breaks, increase the first (poem) number by one to indicate the start of a new poem. Proceed with numbering this way through the document.

Cui dono lepidum novum libellum  
 arido modo pumice expolitum?  
 Corneli, tibi; namque tu solebas  
 meas esse aliquid putare nugas,  
 iam tum cum ausus es unus Italorum  
 omne aevum tribus explicare chartis,  
 doctis, Iuppiter, et laboriosis!  
 quare habe tibi quidquid hoc libelli  
 qualecumque, quod, o patrona virgo,  
 plus uno maneat perenne saeclo.

Passer, deliciae meae puellae,  
 quicum ludere, quem in sinu tenere,

cui primum digitum dare adpetenti  
 et acris solet incitare morsus,  
 cum desiderio meo nitenti  
 carum nescio quid libet iocari  
 (et solaciolum sui doloris,  
 credo, ut tum gravis adquiescat ardor),  
 tecum ludere sicut ipsa possem  
 et tristis animi levare curas!  
 Tam gratum est mihi quam ferunt puellae  
 pernici aureolum fuisse malum,  
 quod zonam solvit diu ligatam.

In fact, this sequential numbering scheme is inadequate for Catullus, because it does not account for particularities of the text reflected in the edition being followed here (by E.T. Merrill, 1893, from Perseus). For example, a section of Catullus 64 runs as follows, as correctly reproduced in the current Tesserae text:

```
<cat. 64.21>      tum Thetidi pater ipse iugandum Pelea sensit.
<cat. 64.22>      o nimis optato saeculorum tempore nati
<cat. 64.23>      heroes, salvete, deum genus, o bona matrum
<cat. 64.23b>     progenies, salvete iterum
<cat. 64.24>      vos ego saepe meo, vos carmine compellabo,
```

An additional, unfinished line from the manuscript is represented here as “23b”, but would not be accounted for in the instructions above, rendering the numbering of this and all subsequent lines in the poem erroneous. A set of instructions that included information on this irregularity is thus required, such as:

Start at <cat. 1.1> and increase the second (line) number by one for each new line. After two consecutive line breaks, increase the first (poem) number by one to indicate the start of a new poem. Proceed with numbering this way through the document.

Exceptions: 1) after 64.23, the following line should be 64.23b, then continue to 64.24 and onward regularly.

There are other irregularities in Catullus’s text that would also need to be noted, but these instructions should suffice to indicate how such circumstances should be addressed in the text instructions.

## SENDING A TEXT FOR REVIEW AND CHECKING IT

Once your text has the proper format, send it and the provenance document to the Tesserae team. We are happy to see your text added, but ask for your patience in reviewing texts, as we have limited resources for managing the project. Once we have reviewed the text and put it into final format, we will return it to you and ask you to check it against the original edition for accuracy. After you indicate the text is fully accurate or send a final revised text, we will incorporate it as soon as we can.

Thank you very much for your help, and we hope our efforts will speed your inquiry.