Beyond the Facsimile: Early English Books Online (EEBO) and the Text Creation Partnership

Welzenbach, Rebecca

http://hdl.handle.net/2027.42/90981
Beyond the Facsimile: Early English Books Online (EEBO) and the Text Creation Partnership

Enriching Scholarship

May 7, 2012
Goals

• Understand the relationship between Early English Books Online and Text Creation Partnership
• Develop ability to “read” an interface: to think about, ask questions of, and make sense of what is being revealed or hidden by websites for digitized resources
• Understand what the Text Creation Partnership produces and some of the ways it can be used—as well as what it cannot do
• Share ideas for using these resources in research and teaching
Agenda

• Introductions (5 min)
• Background: What is the Text Creation Partnership? What is its relationship to EEBO? (10 min)
• Walk through/demo different ways to search the texts (25 min)
  • ProQuest’s EEBO
  • UM’s EEBO-TCP
  • Differences, similarities, preferences?
• Behind the scenes (25 min):
  • What does the TCP really produce?
  • What are the challenges and goals of this work? What purpose does it serve?
  • document analysis exercise
• Ideas/examples for teaching and research (20 min)
• Critical thinking: gaps and limitations of these resources (10 min)
• Questions/discussion (10 min)
• Evaluation (5 min)
Introductions

Who am I?
Rebecca Welzenbach, Text Creation Partnership Project Outreach Librarian

Who are you?

• Name
• Affiliation
• Do you already use EEBO or TCP? In what capacity?
• Any specific questions, concerns, or things you hope will be covered today?

Rebecca Welzenbach | tcp-info@umich.edu | textcreationpartnership.org | @TCPStream
Background: EEBO

• Contains more than 125,000 titles listed in Pollard & Redgrave’s Short-Title Catalogue (1475-1640) and Wing’s Short-Title Catalogue (1641-1700) and their revised editions, as well as the Thomason Tracts (1640-1661) and the Early English Books Tract Supplement

• Commercial product developed by ProQuest

• Purchased or licensed by libraries

• Grew out of Early English Books microfilm series

• Consists of scanned microfilm and (now) direct scans from books

• EEBO alone lets you search the MARC catalog records for each book and view the pages of the book. It does not provide full-text search capabilities

• Page images + searchable metadata

For more information: http://eebo.chadwyck.com.proxy.lib.umich.edu/about/about.htm
Background: EEBO-TCP

• More than 40,000 texts done so far. Goal is to complete ~70,000, or one edition of each unique work in the EEBO catalog

• Each page is keyed by hand (OCR doesn’t work; double/triple-keying) and mark-up added

• Product is a large set of searchable encoded text files that may be matched up with page images in EEBO for search and display, or put to other uses entirely

• Not-for-profit, library-based consortial project

• Libraries pay to support this work; gain immediate access to the output

• Text files belong to partner library institutions and will ultimately be made freely available to everyone

• Model has been replicated with ECCO- and Evans-TCP

For more information: http://textcreationpartnership.org

Rebecca Welzenbach | tcp-info@umich.edu | textcreationpartnership.org | @TCPStream
Demo: ProQuest’s EEBO Interface

http://eebo.chadwyck.com.proxy.lib.umich.edu/

Rebecca Welzenbach | tcp-info@umich.edu | textcreationpartnership.org | @TCPStream
Demo: ProQuest’s EEBO Interface

How does the Variant Spellings feature work?

The Variant spellings box appears on the Basic Search, Advanced Search and Periodicals Search screens. It is checked by default.

If you type a search term in the Keyword(s) box and the Variant spellings box is checked when you submit your search, you will automatically retrieve all instances of your search term and its early modern variant forms in EEBO. For example, if the box for Variant spellings is checked and you type the word murder in the Keyword(s) field, when you submit your search you will retrieve all occurrences of the word murder and its early modern variants murther, murder, murdr and murther.

This will also work if you check the box and type search terms in other fields like Title keyword(s) and Imprint.

If you type a phrase in the Keyword(s) search field (for example Keyword(s): "so foul and fair a day") and the Variant spellings box is checked when you submit your search, your results will include instances of the phrase where the spelling varies (such as so foule and faire a day). Similarly, if you type a series of terms connected by Boolean or proximity operators in this field (for example jealous and green-eyed), your search will include all available spelling and typographic variants of each term (such as jealious, greene-eyed and green-eyd).

When typing a search expression that includes Truncation and wildcard operators (e.g. Keyword(s): je?lo?*s*), you should uncheck the Variant spellings and Variant forms boxes.
Demo: ProQuest’s EEBO Interface


Rebecca Welzenbach | tcp-info@umich.edu | textcreationpartnership.org | @TCPStream
Demo: ProQuest’s EEBO Interface

Catalog record only

Catalog record, images, illustrations, full text from TCP

Catalog record, images, illustrations

---

   The principal navigations, voyages, traffiques and discoveries of the english nation made by sea or over-land, to the remotest and farthest distant quarters of the earth, at any time within the compass of these 1500. yeeres: divided into three severall volumes, according to the positions of the regions, wherunto they were directed. ... / by Richard Hakluyt Master of Artes, and sometime student of Christ-Church in Oxford, . Imprinted at London ; By George Bishop, Ralph Newberie and Robert Barker., 1590-1600.  
   Date: 1598-1600  
   Bib name / number: STC (2nd ed.) / 12626  
   Physical description: 3 v.  
   Copy from: Folger Shakespeare Library

   The principal navigations, voyages, traffiques and discoveries of the English nation made by sea or over-land, to the remotest and farthest distant quarters of the earth, at any time within the compass of these 1600. yeeres: divided into three severall volumes, according to the positions of the regions, wherunto they were directed. The first volume contains the worthy discoveries, &c. of the English ... The second volume comprehended the principal navigations ... to the south and south-east parts of the world ... By Richard Hakluyt preacher, and sometime student of Christ-Church in Oxford, . Imprinted at London ; By George Bishop, Ralph Newberie, and Robert Barker., Anno 1599-1600  
   Date: 1599-1600  
   Bib name / number: STC (2nd ed.) / 12626a  
   Physical description: 3 v. ([24], 619, [1]); [16], 312, 204; [16], 888 p., folded plate.  
   Copy from: Henry E. Huntington Library and Art Gallery

   The principal navigations, voyages and discoveries of the English nation made by sea or over land, to the most remote and farthest distant quarters of the earth at any time within the compass of these 1500. yeeres: dividded into three severall parts, according to the positions of the regions wherunto they were directed. ... Whereunto is added the last most renowned English navigation, round about the whole globe of the earth. By Richard Hakluyt Master of Artes, and sometime student of Christ-church in Oxford, . Imprinted at London ; By George Bishop and Ralph Newberie, deputys to Christopher Barker, printer to the Queens most excellent Maiestie, 1589.  
   Date: 1589  
   Bib name / number: STC (2nd ed.) / 12825  
   Copy from: Henry E. Huntington Library and Art Gallery
Demo: ProQuest’s EEBO Interface

TO THE RIGHT HONORABLE MY SINGULAR GOOD LORD, THE LORD CHARLES HOWARD, Erle of Nottingham, Baron of Effingham, Knight of the noble Order of the Garter, Lord high Adjudict of England, Ireland, and Wales, &c. one of her Maiesties most honourable prouier Counsell.

Right Honorable and my very good Lord, after I had long since published in Print many Navigations and Discoveries of Strangers in divers languages, as well here at London, as in the city of Paris, during my late stayes abode in France, with the worthie Knight Sir Edward Stafford your brother in law, her Maiesties most prudent and careful Ambassador with the French King, and had有色 on old matters and other matters to the sweet study of the history of Cognosce, that I began at the beginnes, to understand the methods of history, which might commend our nation for their high courage and singular active in the Search and Discoveries of the most unknowne quarters of the world. Henceforth seeing no man to stop, to undertake the recording of so many memorable actions, and many man to follow his private affairs: the modest use of my country desirous all difficulties, and as it were with a sharp good provided me and thrust me forward into this most troublesome and painful action. And after great charges and infinite cares, after many wanderings, toils, and travailes, and working out of my weak body, at length, I have collected these several Volumes of the English Navigations, Travellres, and Discoveries, to strange, remote, and far distant countreys. Which works of mine have not included within the compass of many other done in these latter dayes, so that like or nothing worthly of memory have beene performed by any ancient ages, but remaining still by the savor of many helper 


Rebecca Welzenbach | tcp-info@umich.edu | textcreationpartnership.org | @TCPStream
Demo: University of Michigan’s EEBO-TCP Interface

Basic Search
Boolean Search
Proximity Search
Bibliographic Search
Word Index (Phase I)
Word Index (Phase II)
Browse (Phase I)
Browse (Phase II)

About EEBO-TCP
Demo EEBO-TCP
comments to:
UMDL Help
Demo: University of Michigan’s EEBO-TCP Interface

Rebecca Welzenbach | tcp-info@umich.edu | textcreationpartnership.org | @TCPStream
Demo: University of Michigan’s EEBO-TCP Interface

THE FIRST VOLUME OF THE principal Navigations, Voyages, Traffiques, and Discoveries of the English nation, made to the North and Northeast quarters of the World, with the directions, letters, privileges, discourses, and observations incident to the same.

Certeine testimonies concerning K. Arthur and his conquests of the North regions, taken out of the historie of the Kings of Britaine, written by Guifridus Monstersensis, and newly printed at Heidelberg, Anno 1587.

Lib. 9, cap. 10.

ANno Christi, 517. Arthurus, secundo regni sui anno, subiugatis totius Hyberniae partibus, classem suam directit in Islandiam, quaque debellato populo subiugauit. Exin ditigato per caeteras insulas rumore, quod e nullo Provincia resistere poterat, Doldaisius rex Gotlandiae, & Gunfiaucus rex Orcadum, ubi viderunt, promissaque vendebant subiectionem feellit.

Emessa deinde hystene, reuersus est in Britanniam, statimque regni in firmam pacem renouans, moimon duodecim annis ibidem fecit.

The same in English.

IN the yere of Christ, 517. king Arthur in the second yeere of his regne, hauing subdued all parts of Ireland, sailed with his fleet into Island, and brought it and the people thereof under his subjection. The rumour afterwards being spread thoroughall all the other Islands, that no countrey was able to withstand him, Doldaisius the king of Gotland, and Gunfiaucus the king of Orkney, came voluntarily vnto him, and yeelded him their obedience, promising to pay him tribute. The Winter being spent, he returned into Britaine, and establishing his kingdome in perfect peace, he continued there for the space of twelues yeeres.

Lib. 9, cap. 12.

Missis deinde in diversa regna Legatis, insitantur tam ex Gallijis, quam ex collateraliibus Insulis Oceani, qui ad curiam venire debent, &c. Et paulo post: Ex collateraliibus autem Insulis, Guilaumuriius rex Hyberniae, Malusius rex Islandiae, Doldaisius rex Gotlandiae, Gunnasius rex Orkudum, Lot rex Noruegiae, Aschilius rex Danorum.

The same in English.
Which meets your needs? Try it...

In both interfaces...
• Try a keyword/full text search for “cookery”
• Try searching within verse for “love”

To make things more interesting, add these additional variables:
• Try it with and without spelling/form variants allowed in EEBO
• Try changing the limits in EEBO (all items, full text only, etc.)

What do you find?
• Questions?
• Frustrations?
• Inconsistencies?
• Do you find you have a clear preference for one over the other?

Rebecca Welzenbach | tcp-info@umich.edu | textcreationpartnership.org | @TCPStream
Different search results
Which one do I use?

ProQuest’s EEBO Interface

• search more than 120,000 books, whether or not they have full text
• Use fuzzy spelling or variant searching
• Easily download/print sections of a work as a PDF

University of Michigan’s TCP Interface

• Only search works with keyed full text
• Get most up-to-date batch of texts
• Some more precise searching options (e.g. proximity, sorting by frequency)

Both can: move between text and image view, display table of contents, restrict search to certain areas or kinds of text (e.g. letters, drama, verse, etc.)

Big picture: many of the same capabilities, some differences. Use what you’re comfortable with and what meets your needs. If something seems missing from one, try the other. Be aware that both—and any interface—will cause things to look and behave slightly differently, even with the same data underneath
What’s going on behind the scenes?

Two key components:
• Accurately transcribed text
• Structural markup: XML (more precisely, SGML)

That is: tags that explicitly describe the structure of a document, allowing a computer to “read” and understand the relationship of parts of the text to one another, just as we have learned to do when looking at a page.
1 And to

2 Syracusae</p>

REBECCA WELZENBACH

tcp-info@umich.edu

textcreationpartnership.org

TCPStream
Document analysis exercise

• Pick a printout or several
• Work in small groups
• Identify elements of the document’s structure: what can you tell by looking at the page that a computer wouldn’t know? What information would you want to have pulled out?
• Do you agree about what is important to tag, and how to categorize it?

Rebecca Welzenbach | tcp-info@umich.edu | textcreationpartnership.org | @TCPStream
What is the point of all of this?

• An interface/program can only work with what information you give it in the first place

• Our goal is to produce a lot of high-quality data that can more or less be generally agreed upon. This gives a search engine something useful to grab onto, but also leaves room for further enhancements

• We turn wheat into flour, and hope you’ll make bread
Examples of new projects built on TCP data

• Specialized scholarly editions: The Holinshed Project
• Alternative search/display interfaces: PhiloLogic @NU
• Additional analysis: Monk Workbench
• Thematic groupings/websites: Witches in Early Modern England
• For readable, scholarly comments on text analysis: Ted Underwood’s blog
EEBO-TCP in the classroom

• Sample assignments:  
  http://www.lib.umich.edu/tcp/eebo/scholarship.html

• 2012 Spring Term course on history of taste: mining EEBO- and ECCO-TCP for recipes

• Document analysis?
Critical Thinking: Reading the Database

What problems, gaps, shortcomings do you see? (in the interfaces we looked at or in the data underneath)

What would you/do you warn your students about when encouraging them to use EEBO or TCP?
Further questions/discussion

• Please get in touch if we can be of help with your research, or with an assignment in your classroom (happy to give targeted overviews/presentations)
• Thanks and good luck!
• Please fill out the evaluation: http://www.umich.edu/~teachtec/es.html