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?
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
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

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

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

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*, *murdre*, *murdre* and *murdver*.

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 *jealous*, *green-eyed* and *green eyd*).

When typing a search expression that includes Truncation and wildcard operators (e.g. Keyword(s): *je?lo?y*), 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

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

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

Right Honourable, and my very good Lord, after I had long since published in print many Narratives and Discoveries of Stranglers in divers languages, as well here at London, as in the city of Paris, during my late sojourn abroad in France, with the worthie Knight Sir Edward Stafford your brother in law, her Maiesties most prouud and carefull Ambassador with the French King; and had also, in sundry other letters, given some true and exacte knowledge of the secrets of the Garter, and many other pouerfull and excellent Societie, and soe great and most proue Counsell, as I am forced to doe, that may be garded which might commend our nation for their high courage and singular actibitie in the Search and Discoverie of the most unkonowne quarters of the world, therefore, seeing no man to stopp forth to undertake the recording of so many memorable actions, and every man to follow his private affairs, the honest use of my country dissemblies all difficulties, and as it were with a more good promisued at length, and threatned me forward into this most troublesome and painful action. And after great charges and infinite cares, after many waterings, fasts, and prayers; and warning out of my owne body, at length I have collected three several Volumes of the English Narratives, Travellers, and Discoveries, to strange, remote, and farre distant countreyes. Which works of mine have not included within the compass of things only done in these latter days, as though like or nothing worthie of memory had beene performed in former ages, but recording stilke by the space of many hundred yeares, have brought to light many very rare and worthy monuments, which long have lain meassevered in mustie corners, & ratchetly hidden in miste, deseasement, and very like for the greatest part to have beene buried in perpituall oblivion. The first Volume of this workes I haue as yet for the present brought to light, reserving the other two till the next Spring, when by Gods grace they shall come to the Press. In the meantime I humbly offer my service & best hoppfull Patience. I called to mind your honorabull Lordship, who both in regard of my particular obligacion, and also in respect of the subject and matter, might kindly determine the Patrons presence. For first I considered how much it was bound, and how deeply indebted for my youngest brother Edmund Hackly, to whom for the space of foure whole yeares your Lordship committed the government and instruction of that honorable yong noble man, your sonne & here apparent, the lord William

View document image (2)

Howard, of whose high spirit and wonderful towendisse full many a time hath he boasted unto me. Secondly, the bountious grace which I owe to my most dear love the lady Stafford, my singular good lady & honourable mistress, adornned me to be mindful of the renowned family of the Howardes. Thirdly, when I found in the first Patent granted by Queen Marie to the Moscoue company, that the lesse your utter being then late high Admiral of England, was one of the first navigators and furthers, with his purdue and countenance, of the strange and wonderfull Discoverie of Russia, the chiefest centers of this present Volume, then I remembred the sage saying of sweet Isocrates. That seeming ought not only to be inhereters of their fathers substance, but also of their commendable vertues and honours. But what speeche of your ancestors herowys (which to say the truemost, are very great, and such as our Chronicles haue notably blesse) when as your owne Heroicall actions from time to time have shewed themselves so admirable, as no antiquite hath afforded greater, and the future times will not in haste (I thinke) performe the like. To come to some particulars, when the Emperors sister, the spouse of Spaine, with a Fieete of 130. sollys, stoutly and prouudely passeth the narrow seas, your Lordship accompanied with ten ships anlye of her Maiesties Isaac Roalti, enuironed their Fieete in most strange and warlike sort, enforced them to sail, and to vaile their choises for the Queene of England, and made them perfectly to understand that side speach of the prince of Poets.


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

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 Monaometensis, and newly printed at Heidelberg, Anno 1587.

Lib. 9. cap. 10.

ANno Christi, 517. Arthurus, secundo reigi sui anno, subiguatis totius Hyberniarum partibus, classem suam directum in Islandiam, undeque debellato populo subiguatis. Exin diluigato per caeteras insulas rumore, quod e nullo Provinciae resistere poterat, Doldaunius rex Godlandiae, & Gunfacius rex Orkaduum, vir baptizati, promissaque vertitvi subjectionem facienti.

Emensa deinde hyme, reuersus est in Britanniiam, statimque regni in firmam pacem renovans, mox in duodecim annis ibidem fecit.

The same in English.

IN the yere of Christ, 517. king Arthur in the second yeere of his reign, 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 throrowout all the other Islands, that no countrey was able to withstand him, Doldaunius the king of Godland, and Gunfacius 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 kingdom in perfect peace, he continued there for the space of twelve yeares.

Lib. 9 cap. 12.

Missis deinde in diversa regna Legatis, insituantur tam ex Galliis, quam ex collaterabilibus Insulis Oceani, qui ad curiam venire deberent, &c. Et paulo post: Ex collaterabilibus austem Insulis, Guillaumurius rex Hyberniae, Maluosius rex Islandiae, Doldaunius rex Godlandiae, Gunniasius rex Orkhadum, 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?
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.

Rebecca Welzenbach | tcp-info@umich.edu | textcreationpartnership.org | @TCPStream
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.
Paul's Epistle to the Hebrews

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?

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