Reading as Gathering

In the Transliteracies' research clearing house we've been seeing many research reports that focus on reading as a collective activity that brings together a group of readers to one reading locus in the online or new media environment. In such spaces multiple readers have access to the same text—they read it, they respond to it, and they help write it, either explicitly in the case of wikipedia or indirectly (or, at least, differently!) in the case of blog platforms. This is extremely interesting, but I am equally intrigued by the way that reading is not just a collective activity, but a *collecting* one, a process that not only allows people to come together and gather as a group of readers, but one that also involves the act of gathering, sorting, sifting, organizing, processing, and editing information.

I would like to focus here on precisely this: reading as a collective and collecting activity, reading as gathering. By looking at a few case studies and excerpts from my own research reports, I wish to explore, however briefly, the way collective and collecting reading shift in the online environment and to consider what happens when the act of collecting or selecting becomes an automated or partially automated part of the reading process.

The insight into reading as a collecting activity comes from a seminar that I've been sitting in on with Wolf Kittler in the German department called "Literally: Derrida Reads Plato, Rousseau, and Artaud." We've been looking at several etymologies in our reading, and, not surprisingly for a course on this topic, etymologies of the word "reading" itself, in multiple languages.
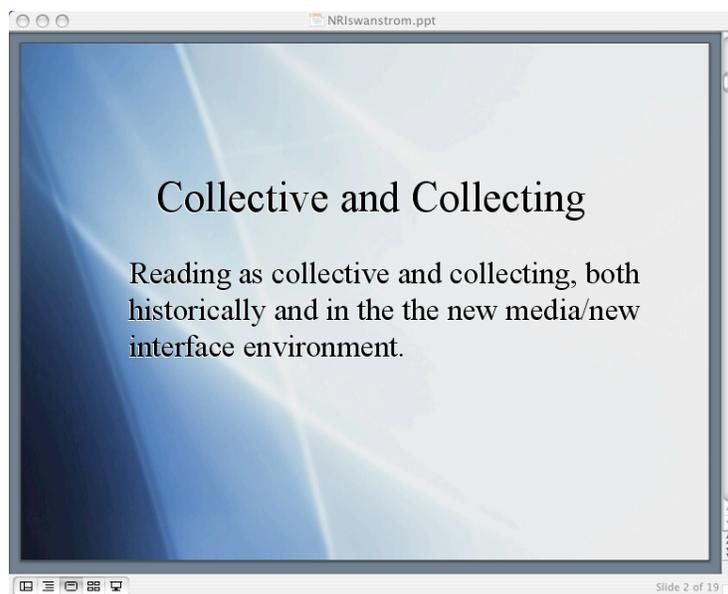
These word origins indicate that the words for reading contain within them a complex history, one that includes not only a coming together to experience a text in a collective manner, but one that always involves a process of selection and collec*ting* on the part of the reader.
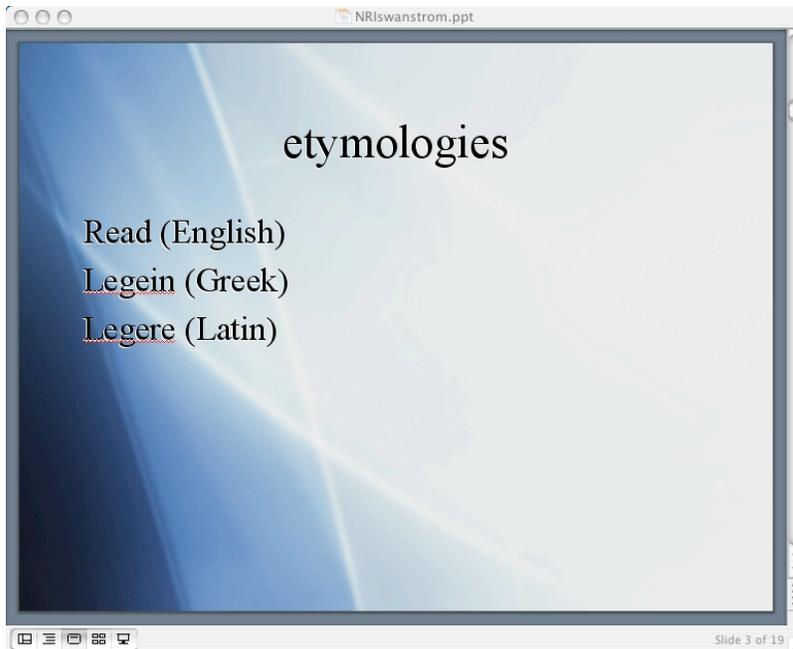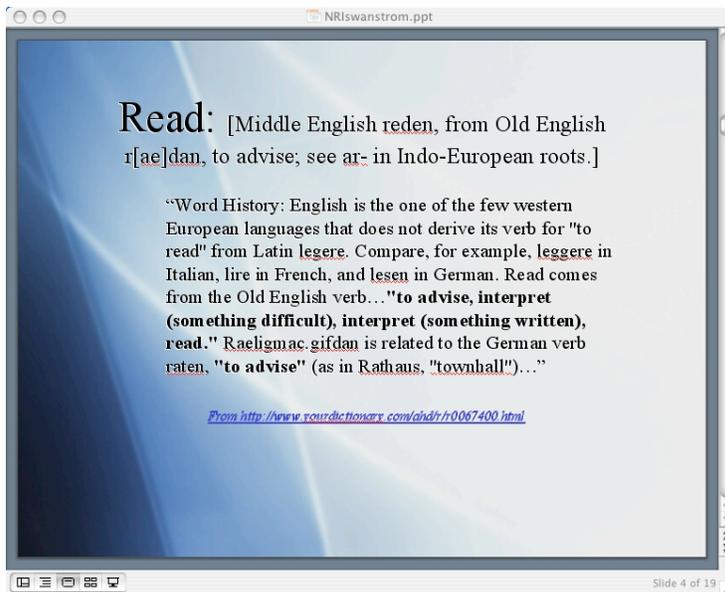
Etymologies
I'd like to focus briefly on three etymologies:
Read (English)
Legein (Greek)
Legere (Latin)

"Read"

The etymology of the word reading in English reveals an emphasis is on advising and interpretation. It's interesting to note that "English is the one of the few western European languages that does not derive its verb for "to read" from Latin legere." Read

instead comes from the Old English verb…"to advise, interpret (something difficult), interpret (something written), read." The word is related to the German verb raten, "to advise." (*http://www.yourdictionary.com/ahd/r/r0067400.html*)

In looking at the definitions of the Latin *legere* and the Greek *legein*, however, we can see this notion of collecting or gathering as something that is inherent to reading in Western tradition. In addition to their more common meanings from which we drive words like "legal," "lexicon," "logos," etc., we see the words defined here as meaning "to collect, gather together, pick"; "to survey, scan, read, peruse"; "to choose for oneself, to pick out, to count." Most interesting for this reader is the strangely evocative usage of the word "legein" outlined in Liddell and Scott's Greek-English lexicon: "picking out

stones for building walls," which occurs in the *Odyssey*. I'll return to this usage later on in the paper.

This is not to say of course that the etymologies reveal some sort of final word or "Ur" meaning for our words for reading, but it seems worthwhile to look at these early versions of the concepts related to reading in order to argue that the act of collecting, something that seems so fraught in terms of new media advancements in the fields of data mining, searching, filtering, and information architecture, is something that has always been a part of what it means to read.
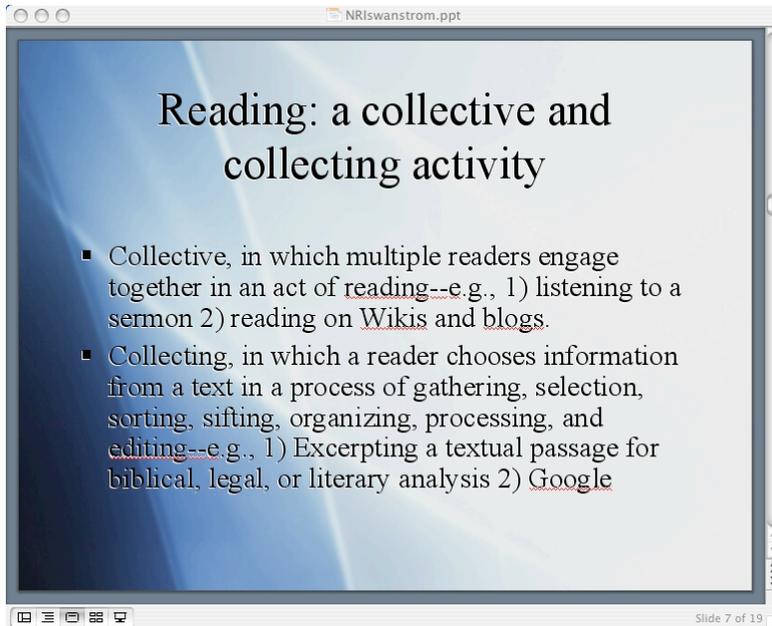
Reading: collective and collecting

We seem fairly comfortable talking about the reading as a collective activity, both historically and into the new media environment. We can talk about how the new media environment is adding to the history of collective reading—and writing—in many interesting ways.

Consider two examples of collective reading, in which multiple readers engage together in relation to a text--e.g., 1) listening to a sermon 2) reading on Wikis and blogs. You could argue that there are many similarities between the two acts. One difference between them might be the way automated processes enable the latter.

We can similarly consider two examples of reading as collecting, in which a reader chooses information from a text in a process I take to include gathering, selection, sorting, sifting, organizing, filtering, and editing--e.g., 1) Excerpting a textual passage for biblical, legal, or literary analysis 2) typing a word into Google's search form and clicking "I'm feeling lucky."

Again, we might see many similarities between these two examples—then again, we might not—but one difference between them might be the manner in which invisible automated processes enable the latter.
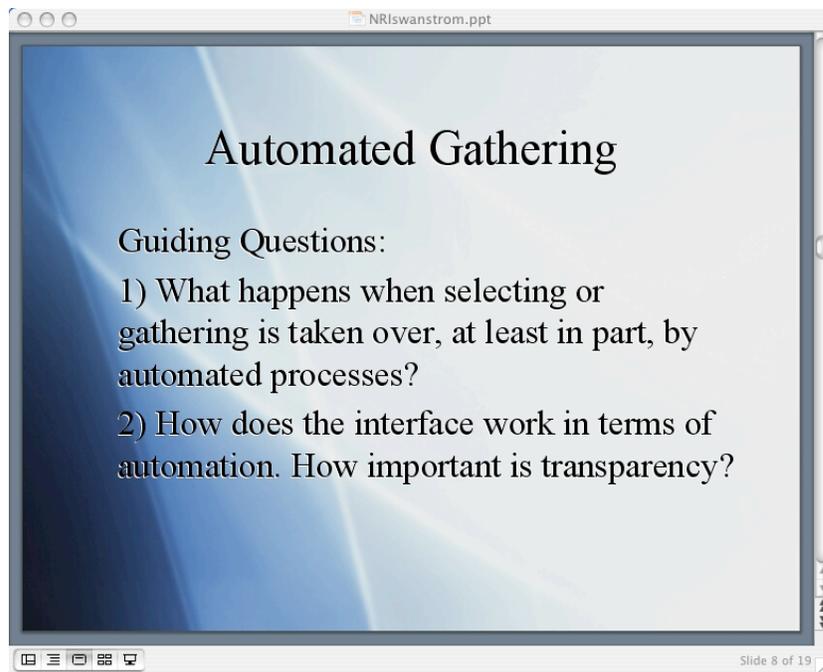
Automated Gathering

My emphasis now will be to look at how this mode of "collecting" becomes automated. I would like eventually to come up with a more precise definition of the word automated, by the way, so I hope you will forgive the vagueness of the term for now.

Two questions guide my investigation here:
1) What happens when selecting or gathering is taken over, at least in part, by automated processes?

2) How does the interface work in terms of automation? How important is transparency? How does the interface make the process of automated collecting and searching (in)visible?
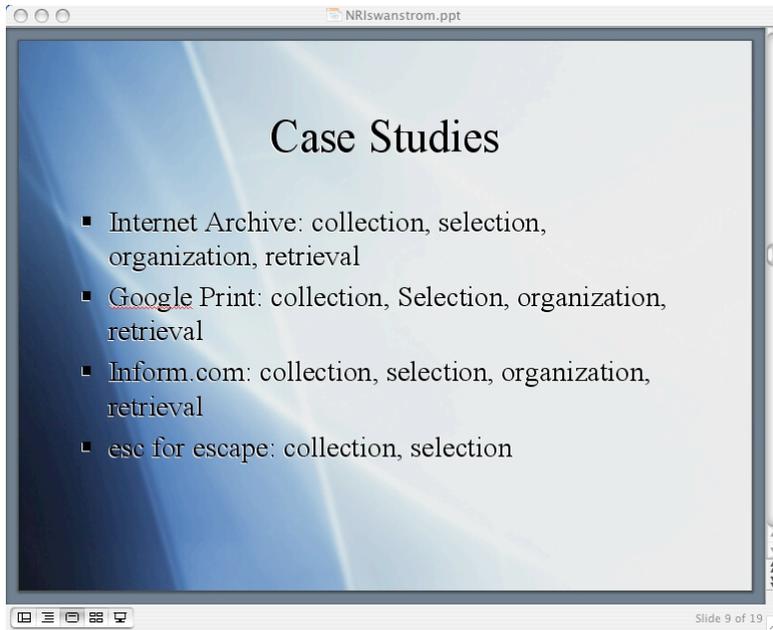
Case Studies

To address these questions, I have selected examples of automated searching—what I am calling collecting—from four of my own research reports that focus on some aspect of gathering or another.   My case studies include:

The Internet Archive's "Way Back Machine"
Google Print's "Man on the Moon" initiative
Inform.com's polytope
Esc for Escape's archiving

Of these reports, the first three involve a high level of automation, while the fourth, an art installation, catalogs reactions to automatically-generated error messages. Rather than just re-stating these reports in their entirety, I'd like to provide you with the following:

1) Brief summary of the "object."
2) An explanation of how the information is collected—as far as my lack of technical expertise allow.
3) The way the interface works in relation to the act of collecting and gathering.
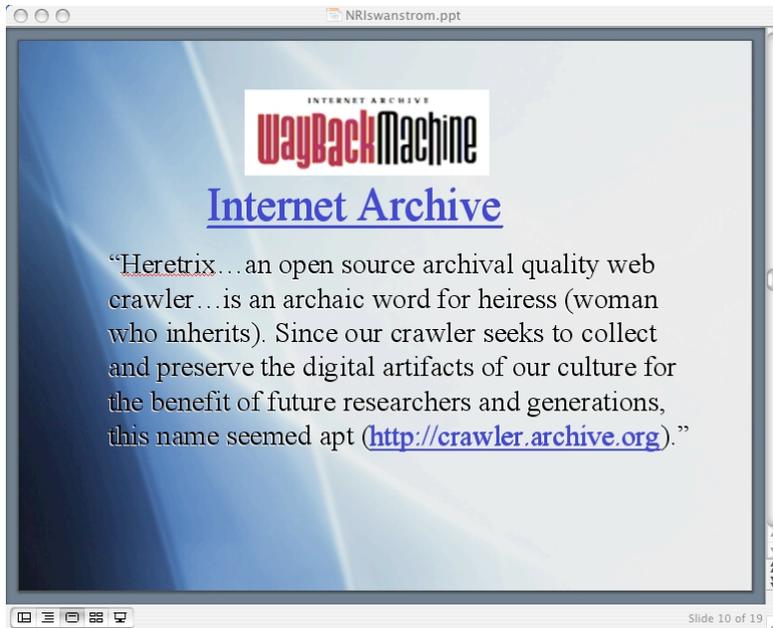
My first case study is the Internet Archive's Way Back Machine. First launched in 1996, the Internet Archive is a non-profit organization based in San Francisco that curates and maintains an accessible, online archive of a vast amount of multimedia objects that have been published on the web since the popularization of the Internet in the mid-nineties.

In terms of their automated gathering processes, the Archive make use of "crawling" technology—i.e., data mining technology used to search, index, and catalogue current information on the web. Archive developers have made sustained efforts to explain and make available the technical specifications of their searching technology, offering on their web site introductory comments as well as more detailed information and schematics for programmers. Additionally, they maintain several forums devoted to discussions and questions about their specific technologies. One such technology is the the "Heretrix," "an open source archival quality web crawler" that searches the web in a variety of ways in its cataloguing efforts:

> Heretrix…is an archaic word for heiress (woman who inherits). Since our crawler seeks to collect and preserve the digital artifacts of our culture for the benefit of future researchers and generations, this name seemed apt. (http://crawler.archive.org).
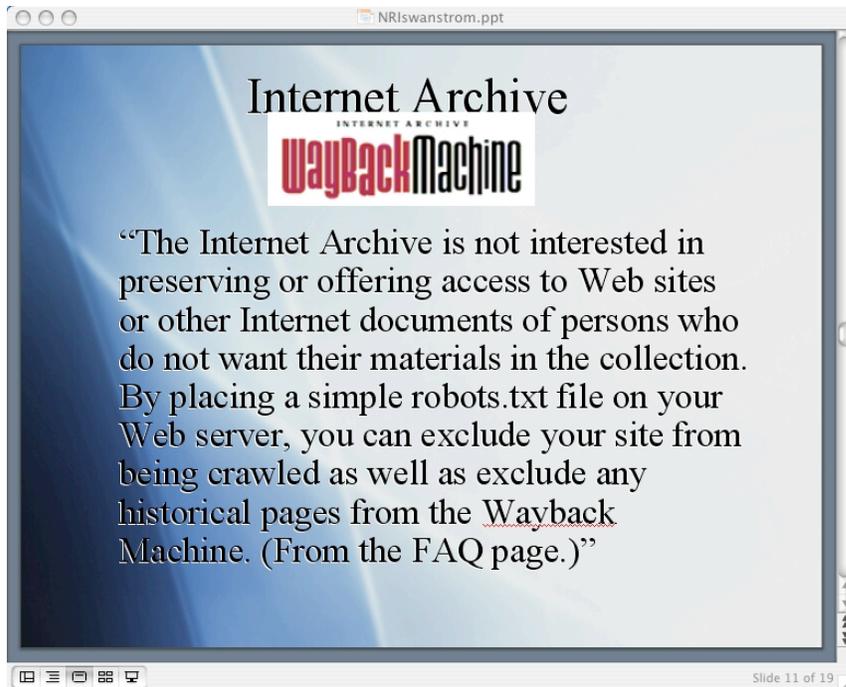
The Interface and front-end design of the page are visually straightforward, with information parsed according to familiar categories: moving images, live music archive, audio, and texts. The site makes many reference to the code that underpins its functioning and invites programmers to aid in the cataloging process.

The entire site, in fact, makes an effort to foreground its technology and, at least in its documentation, encourages user participation in the searching and collecting process. The name of the "Wayback Machine," featured so prominently on the interface, is a reference to Rocky & Bullwinkle's *Peabody's Improbable Adventure;* its presence indicates an eagerness on the part of the developers to offer a user-friendly gateway into the archives.
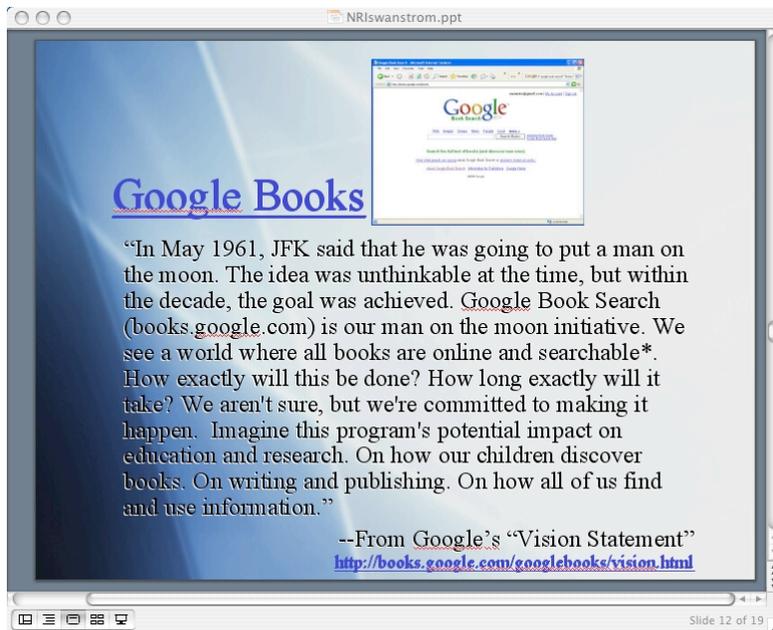
My second case study is Google Book Search, a controversial initiative to dramatically increase the amount of print literature available for on-line consumption. Google's strategy is double-pronged, involving negotiations with both publishers *and* libraries to scan print works and convert them into searchable digital formats. In their own words:

> "In May 1961, JFK said that he was going to put a man on the moon. The idea was unthinkable at the time, but within the decade, the goal was achieved. Google Book Search (books.google.com) is our man on the moon initiative.

The fact that the technology behind Google Book Search is not entirely transparent is not very surprising. Google, after all, has been notably silent regarding the details of its 2003 patent for "interconnectivity-based" page-ranking search capabilities. Regardless of this opacity, however, it is clear that the Book Search project makes use of computer technology to fulfill at least two broad functions: conversion & hosting and searching & retrieval.

In terms of the interface, the front-end design almost exactly like the Google homepage. The most successful search engine is also the most Spartan in terms of its visual aesthetic. The design is Completely bare bones, except on holidays, when Google trots out some holiday clip art to adorn the double Os of its corporate logo. Overall, the interface does not offer very much in the way of a participatory space in terms of giving a reader an active role in searching and collecting. Rather, it channels the reader into one narrow search corridor.
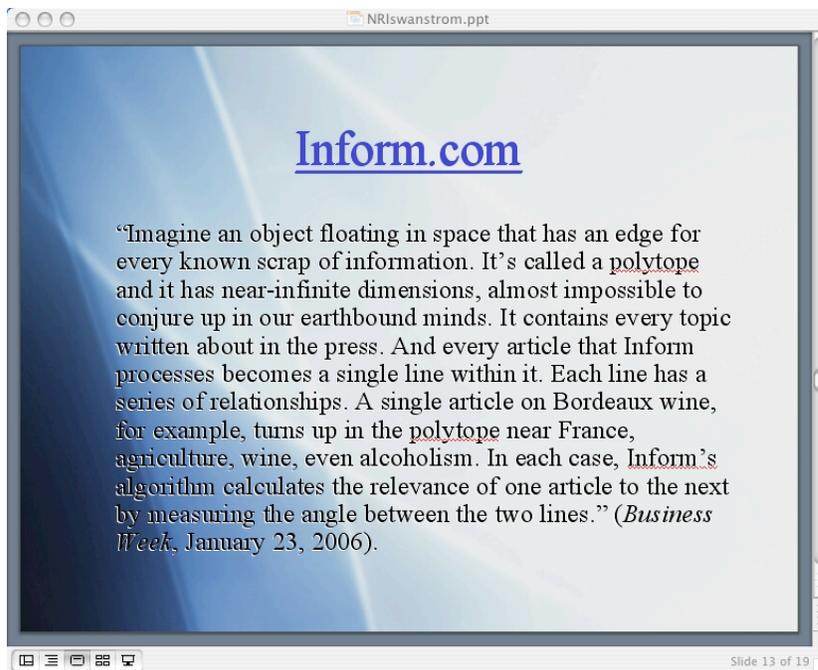
My third case study is Inform.com. Founded in late 2005, Inform.com is an online news synthesizer that allows its users to design and customize "news channels" according to their individual interests. Inform.com is distinct because it actively crawls each news item it hosts in order to more effectively categorize and sort its contents.

In terms of its gathering processes, Inform.com works in the following manner: a user visits the site and registers. After registering, the user has the opportunity to either read pre-sorted "lead" news items that have been collated by Inform.com, or to create individual "news channels" that will scan, sort, and display news items that reflect the user's interests:

> Imagine an object floating in space that has an edge for every known scrap of information. It's called a polytope and it has near-infinite dimensions, almost impossible to conjure up in our earthbound minds. It contains every topic written about in the press. And every article that Inform processes becomes a single line within it. Each line has a series of relationships. A single article on Bordeaux wine, for example, turns up in the polytope near France, agriculture, wine, even alcoholism. In each case, Inform's algorithm calculates the relevance of one article to the next by measuring the angle between the two lines. (*Business Week*, January 23, 2006).

Perhaps as a result of polytope technology, Inform.com has created a new reading interface that offers a significantly different way to access daily headlines than that offered by traditional, tabloid-sized printed matter.

This fairly high-level of customization helps shape the interface in response to the newsreader's preferences. The interface thus has the potential to become dynamic and

interactive, and personalized.   At the same time, however, news items are still parsed into standard categories on the home page. And while Inform boasts of a unique crawling technology that helps increase the relevance of the content it delivers, this crawling technology remains largely invisible to the user. What might the issue of transparency add or detract to the site's service? Along these same lines, how is the user to determine which news sources Inform.com crawls? Are independent, non-commercial news sites included as targets, or are these searches primarily limited to mainstream media providers?

NRIswanstrom.ppt

Giselle Beiguelman's "esc for escape"
Research Report

http://www.desvirtual.com/escape/english/fatal.htm

Slide 14 of 19
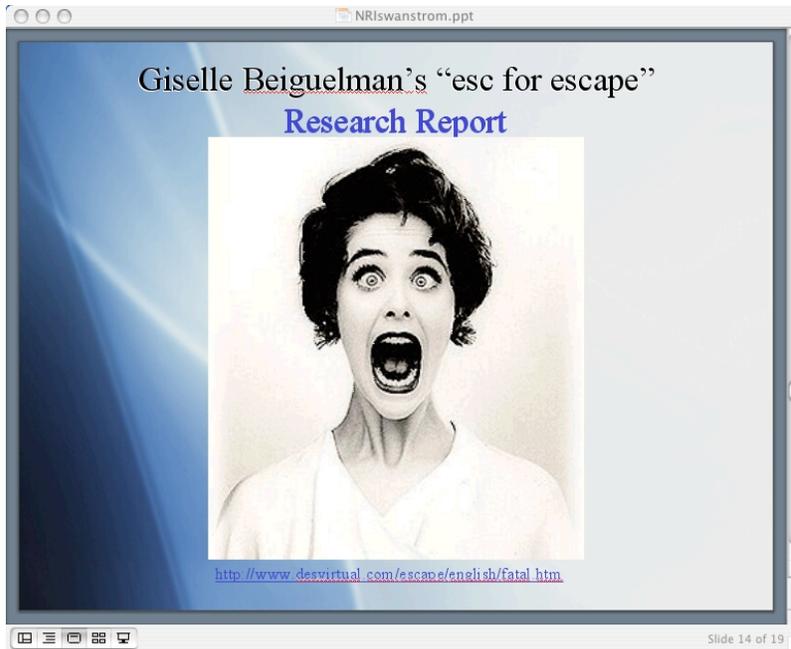
The final case study I'll be looking at is Giselle Beiguelman's "esc for escape," a fascinating collection of computer error messages that the artist has collected from users around the world. In contrast to the other sites I've looked at that have employed a high level of automation in the act of gathering or collecting, Beiguelman's work here is rather a collection of automated responses.

"Esc for escape" (2004) is a multifaceted art project that archives error messages from computer users around the globe and re-expresses them in a variety of contexts and media, including electronic billboards in São Paulo, Brazil; a repository of selected error messages published on the web, entitled "The Book of Errors"; "The Monastery," an archive of all error messages related to the project; a dvd of the project; a project blog; as well as several "trailers," which offer ironic visualizations of various error messages by the artist, of which this image is one.

In this animated .gif the exaggerated expression of fear, even horror, on this woman's face is further exacerbated by violent shaking. That this effect is enabled by a browser's capacity to loop rapidly between multiple images to create a trompe l'oeil, such that the reader is tricked into seeing persistent motion rather than sequential stills is somehow appropriate, in that one's senses are, from the very beginning, manipulated and

conditioned by the automatic processes of the computer. The result is a (darkly) comedic and hyperbolic portrayal of what it might feel like to receive an error message such as the ones found in the "esc for escape" archives.

Beiguelman originally began archiving error messages as a part of an earlier project with which she examined the relationship between anxiety and technological error by posing a pointed and provocative set of questions to computer users: "Have you ever read something scary on your screen? Do you understand why programmers suppose they are programming for programmers? Do you fear error messages?"

The archive of error messages makes for a good read—one of my favorite errors reads as follows:

> First snow, then silence.
> This thousand dollar screen dies
> So beautifully.

The collective and collecting aspects are particularly interesting in this project. Several of our objects for study and research reports (e.g., Wikipedia, youtube.com, delicious, etc) reveal a keen interest in tracing the emerging practices and rhetorical strategies in the field of collective/collaborative authorship; in some sense, what differentiates this piece from other works is precisely this collaborative nature. While Beiguelman is certainly the master architect of the overall project, the core of "esc for escape," i.e., the error messages themselves, represent a variety of voices and social contexts from around the world. And while the piece has no narrative arc in the traditional sense of the word, the confessional nature of many of the posts lends the work an aesthetic cohesion that is worth consideration.

The object of this collective collection is also significant—by gathering examples of text generated when automation gone awry, Beiguelman's work manages to do what more highly automated collection tools and search engines do not: reveal, at least to some degree, the largely invisible processes that drive acts of automated collecting and gathering in online reading environment.
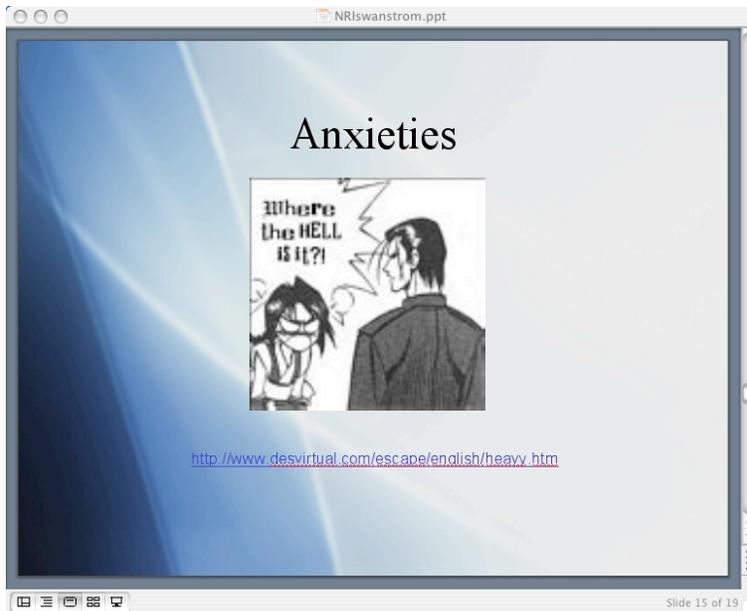
Anxieties

"Esc for Escape" expresses very well anxieties about automated processes in the online reading environment, and Beiguelman's representations are perhaps indicative of a greater trend of suspicion and anxiety about automated collecting, searching, gathering, and reading, as well as about what happens when such processes go awry.
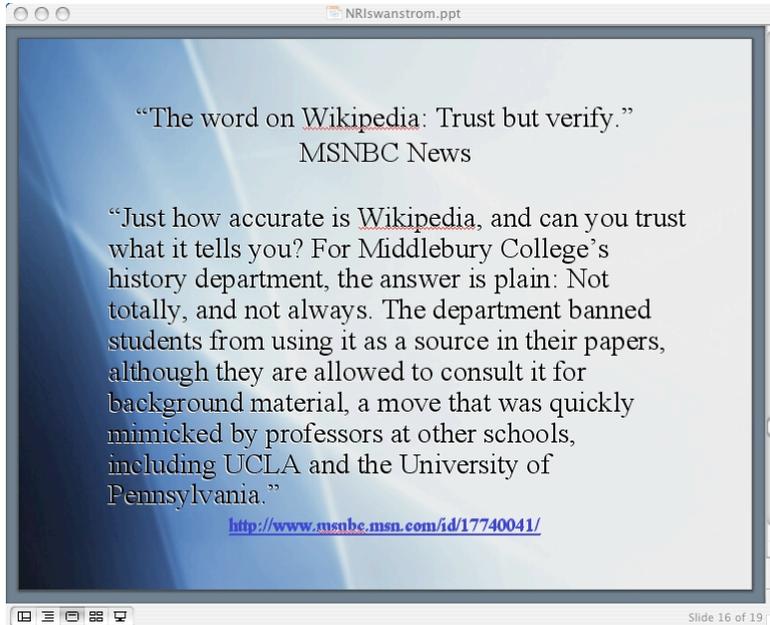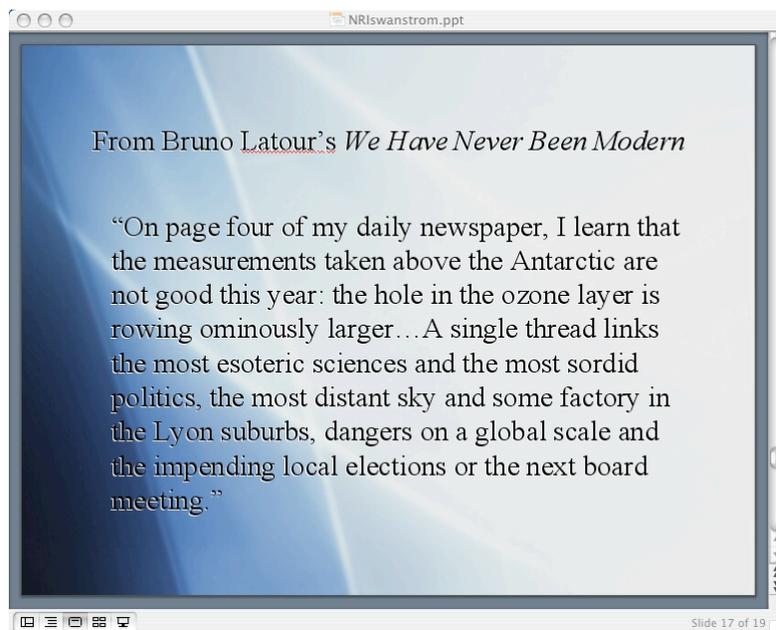

As a recent article from MSNBC relates, in early 2007, the history department of Middlebury College in Vermont instituted a ban on using Wikipedia articles as a source in any research assignment.
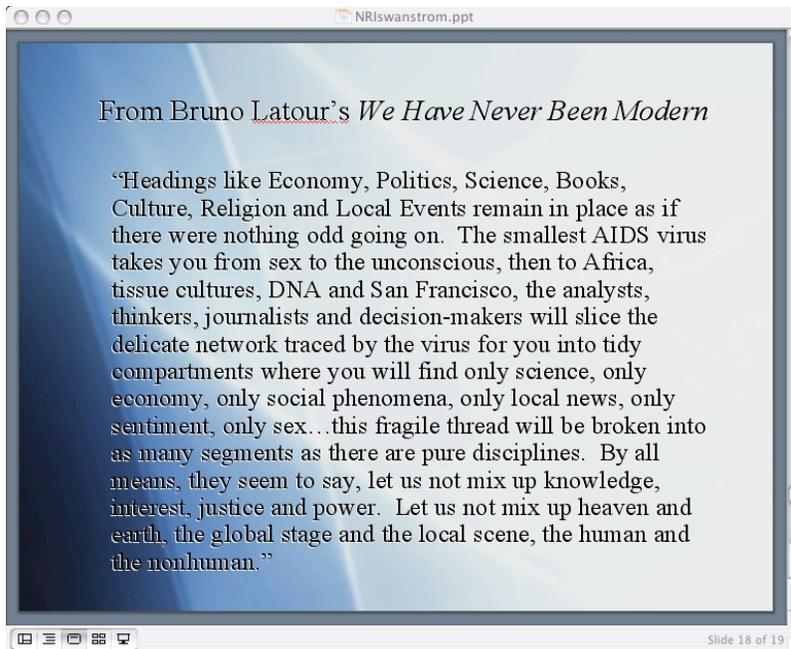
This is interesting in the sense that Wikipedia is a great example of both collective and collecting reading in the new media environment. Automation drives both its collective and collecting aspects by enabling a group reading experience and allowing people to easily search its contents and gather results.

Anxieties about reliability of such information sources are not limited to Vermont. UCLA has considered the ban, and the issue of authority of wiki authors is hotly contested on a variety of forums—including, it is fair to add, on Wikipedia.

One of the fears or anxieties that many of us have as educators is that our students' ease of access can only impair their collecting skills. The argument goes something like this: If they are not in the practice of validating and investigating their information sources, that is, if they no longer have to do the hard work of "picking up the stones for building walls," to finally come back to the etymology of the word "legein" that I introduced earlier on, this can only hamper their critical capabilities, and the "collecting" and critical aspect of their reading skills will atrophy.

But we need to ask ourselves if we buy into this claim. I'm not sure that I do, at least not entirely. I feel as though I might be going out on a limb here, but I would suggest that we might want to think about the problems inherent in building walls in the first place.

From Bruno Latour's *We Have Never Been Modern*

"Headings like Economy, Politics, Science, Books, Culture, Religion and Local Events remain in place as if there were nothing odd going on. The smallest AIDS virus takes you from sex to the unconscious, then to Africa, tissue cultures, DNA and San Francisco, the analysts, thinkers, journalists and decision-makers will slice the delicate network traced by the virus for you into tidy compartments where you will find only science, only economy, only social phenomena, only local news, only sentiment, only sex…this fragile thread will be broken into as many segments as there are pure disciplines. By all means, they seem to say, let us not mix up knowledge, interest, justice and power. Let us not mix up heaven and earth, the global stage and the local scene, the human and the nonhuman."

From Bruno Latour

The entire discussion about reading and gathering reminds me very much of Bruno Latour's introduction to *We Have Never Been Modern.* In response to reading a newspaper in which news items about global warming and the AIDS virus are compartmentalized into overly tidy and unrelated news bites, he writes:

> "…the analysts, thinkers, journalists and decision-makers will slice the delicate network traced by the virus for you into tidy compartments where you will find only science, only economy, only social phenomena, only local news, only sentiment, only sex…By all means, they seem to say, let us not mix up knowledge, interest, justice and power. Let us not mix up heaven and earth, the global stage and the local scene, the human and the nonhuman."

Latour calls attention to the way information can be bracketed and disconnected to the network of associations to which it belongs. He calls for a re-tying of the Gordian knot that binds such things together and conceptualizes a way for us to think about networks and hybrids that are not contained solely within rigid categorical boundaries. While we clearly want to be cautious about how automation is working with our collective and collecting reading skills, such processes may just offer us a way to think about and navigate the information networks that Latour envisions.