Purdue University

Purdue e-Pubs

Charleston Library Conference

Leveraging Usage Data and User-Driven Development to Extend the Use of Collections

Kristen Garlock
ITHAKA/JSTOR, kristen.garlock@ithaka.org

Leveraging Usage Data and User-Driven Development to Extend the Use of Collections

Kristen Garlock, Associate Director, Education & Outreach, ITHAKA

Abstract

In 2014, the JSTOR Labs team used an algorithm to identify more than 9,000 articles on JSTOR that exhibited patterns of use consistent with being used in the classroom or assigned as coursework. Using a low-cost rapid development approach called "flash builds," the team validated and built a prototype browser for this dataset with the direct involvement of teachers at the secondary and introductory college levels. This dataset is now available as "Classroom Readings," a free and open experimental resource. Classroom Readings (http://labs.jstor.org/readings) is designed to help educators find articles on JSTOR that are good candidates for teaching, adding value to the investments libraries have made in JSTOR collections.

Text

As larger proportions of library budgets are invested in electronic journals each year and we have more data about their use, we can leverage that data to help libraries get even more out of these investments. In 2014, JSTOR employed usage data analysis and rapid development techniques to build a new tool called "Classroom Readings." The tool enables educators to find articles that have been used in courses and filter the results by reading level and length.

When we began this project we did not initially set out to create this particular tool. Our original idea was to help support the use of content from JSTOR in teaching by creating expert-curated lists of articles and books that aligned with core first-

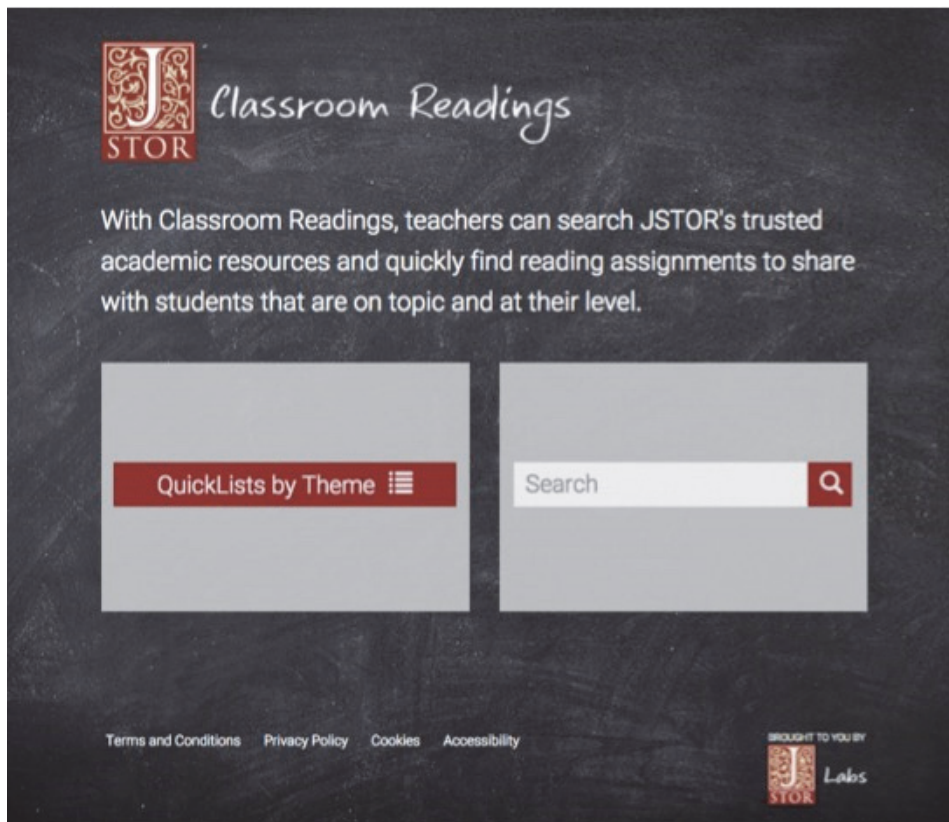| doi	discipline	article title
10.2307/450483	Language & Literature	Guilt and Death: The Predicament of The Ancient Mariner
10.2307/3175999	Language & Literature	Elephants in the Labyrinth of Empire: Modernism and the Menagerie in "The Old Wives' Tale"
10.2307/462329	Language & Literature	The Doll House Backlash: Criticism, Feminism, and Ibsen
10.2307/2933083	Language & Literature	Resistance, Rebellion, and Marriage: The Economics of Jane Eyre
10.2307/462130	Language & Literature	Frankenstein: Creation as Catastrophe
10.2307/2871293	Language & Literature	Is There a Performance in this Text?
10.2307/461329	Language & Literature	Alice's Invasion of Wonderland
10.2307/3844618	Language & Literature	"Jane Eyre", from Governess to Girl Bride
10.2307/3174801	Language & Literature	A Genealogy of Dependency: Tracing a Keyword of the U.S. Welfare State
10.2307/1343054	Language & Literature	The Marginal Gloss
10.2307/25115310	Language & Literature	"The Great Gatsby" and the Obscene Word
10.2307/461001	Language & Literature	Wuthering Heights and the Limits of Vision
10.2307/20027310	Language & Literature	What Is Security?
10.2307/3174406	Language & Literature	Transforming Franz Kafka's "Metamorphosis"
10.2307/2853637	Language & Literature	Mohammed and Charlemagne: A Revision
10.2307/25293689	Language & Literature	Faust's Damnation: The Morality of Knowledge
10.2307/2928384	Language & Literature	"Shaping Fantasies": Figurations of Gender and Power in Elizabethan Culture
10.2307/1343408	Language & Literature	A Few Canonic Variations
10.2307/2851941	Language & Literature	The Carolingian Age: Reflections on Its Place in the History of the Middle Ages
10.2307/3844619	Language & Literature	"Sympathy" in Jane Eyre
10.2307/20024056	Language & Literature	Deep Play: Notes on the Balinese Cockfight
10.2307/3174446	Language & Literature	Looking and Listening: The Construction of Clinical Knowledge in Charcot and Freud
10.2307/1261283	Language & Literature	Adam, Eve, and the Fall in "Paradise Lost"
10.2307/3844532	Language & Literature	Discovery and the Domestic Affections in Coleridge and Shelley
10.2307/30166165	Language & Literature	Kafka's "Metamorphosis": Rebellion and Punishment
10.2307/361359	Language & Literature	The American Civil War and the Problem of Inevitability
10.2307/1350027	Language & Literature	Moving Tropes: New Modernist Travels with Virginia Woolf / مجازات التنقل: ترحلات الحداثة الجديدة عند فرجينيا وولف
10.2307/3174621	Language & Literature	Feminist Critics and Literary Mothers: Daughters Reading Elizabeth Gaskell
10.2307/813939	Language & Literature	Gatsby: False Prophet of the American Dream
10.2307/372763	Language & Literature	Satiric Strategy in the Importance of Being Earnest

Figure 1. Thematic patterns in dataset.

Figure 2. JSTOR Classroom Readings.

Using this technique, we identified about 9,000 articles that had this usage pattern in 2011–2013. In looking across this dataset, we saw some interesting trends: there was very little duplication of articles assigned across institutions, but there were strong thematic patterns. We thought that perhaps we could identify articles in a theme and use these as the basis for our course lists.

To help us explore the data in more detail, we created a rudimentary web browser that enabled us to much more quickly spot these thematic patterns, and we began to use it to build lists of content. This led to our a-ha! moment . . . perhaps this browser itself might have value to educators.

To test the concept of making the browser available to educators, we worked with our small JSTOR Labs team to build and test a prototype with high school and college teachers. Using a rapid development technique called a "flash build," we were able to quickly validate the concepts and features of most value.

The teachers were enthusiastic about the concept and particularly interested in the thematic lists, or QuickLists, as we named them. We refined the tool in July and August of last year, built a number of sample thematic lists, and rolled out the tool as Classroom Readings last fall. The tool is free and open for anyone to use, though links to the full-text content on JSTOR may require an institutional or individual account. Recently, we reran the algorithm against usage on JSTOR for January 2014–June 2015, and expanded the dataset to approximately 50,000 articles.

year courses. We quickly realized that such a manual, top-down approach would be time and resource intensive; engaging already busy subject matter experts to create these lists would be difficult. So, we took a step back and challenged ourselves to find another way to correlate content on JSTOR with course topics.

From years of anecdotal reports, we know that content on JSTOR is used in teaching in many ways, not just for research papers. But what content? How could we discover which articles were most useful in a teaching context? Would teaching use show up in the usage patterns?

We decided to experiment with looking for usage patterns that would distinguish use of an article assigned in a class versus use of an article accessed by an individual for a research paper.

Our data scientist created an algorithm to look for a "teaching use" pattern, which we defined as spikes in usage of an article at a single institution within a two-week period.

Figure 3. Understanding Shakespeare.

Of course, usage date can only tell us what's happening on-site, not about other teaching use via course tools and sharing of PDFs. In the future, it may be possible to enrich the dataset further by looking at other metrics and enabling educators to contribute to, review, or annotate the lists and content. Longer term, the features of Classroom Readings may be useful to incorporate directly into the main JSTOR interface; for right now, it's a beta tool providing a means to test the concept and features more broadly.

While not perfect or entirely comprehensive, the teaching use algorithm and digging into the data first helped us to develop hypotheses that could be tested through quick, low-cost methods. We've since used these techniques to build and test several other tools. One example is Understanding Shakespeare, a collaboration between JSTOR and the Folger Shakespeare Library. Using a matching algorithm developed by JSTOR Labs, we created

links from the lines in the plays to the secondary literature on JSTOR.

We're also thinking about other usage analysis that might indicate opportunities to improve discoverability. Perhaps we could using an analysis of the frequency of search terms to identify topics that could be enhanced with on-site subject guides.

The teaching usage data might also be applied in other ways to help libraries. One librarian has suggested that being able to identify titles used in teaching might help with collection development. For example, would knowing that a title gets relatively smaller use overall but has strong teaching use impact a cancellation decision? These are some of the things we're thinking about to help students and educators, and to add value to the investments libraries have made in JSTOR.