

JSTOR Sustainability: Serendipity in an Interdisciplinary Field

Florence Lee

Ithaka

301 E. Liberty St #400

Ann Arbor, MI 48104

Florence.Lee@ithaka.org

Jessica Keup

Ithaka

301 E. Liberty St #400

Ann Arbor, MI 48104

1 (734) 780-2499

Jessica.Keup@ithaka.org

ABSTRACT

Serendipity and the unexpected discoveries of hidden connections has long been a challenge within information retrieval platforms. Supporting both accidental discoveries and purposeful search successfully is a difficult balance to strike, particularly when the topic on hand spans across multiple disciplines and the users themselves come from varying backgrounds and experiences. JSTOR Sustainability is an attempt to facilitate serendipitous discovery within the sustainability field.

CCS Concepts

- *Information systems~Document topic models*
- *Information systems~Query suggestion*
- *Information systems~Search interfaces*
- *Information systems~Similarity measures*
- *Information systems~Information retrieval diversity*

Keywords

Serendipity; Interdisciplinary; Digital library; Linked data; Topic pages; Discovery; Sustainability

1. INTRODUCTION

Serendipity within information retrieval and information seeking platforms has long been a challenge. Serendipity plays the role of revealing hidden connections and analogies through links between information sources [1]. Researchers need the flexibility to find and explore multiple pathways and make connections between topics, often facilitated by browsing, but information retrieval systems also need to allow users to quickly and efficiently find specific information they need [2]. Striking the balance is even harder when one has to consider users from multiple disciplines and varying levels of expertise and experience.

JSTOR is a digital library of academic journals, books, and primary sources. With millions of documents and bibliographic metadata describing the documents, our goal is to help people discover, use, and build upon a wide range of research and content. The goal of the JSTOR Sustainability project was to discover and build ways in which researchers can encounter serendipity in their day-to-day work in an interdisciplinary field. We wanted to create a platform that could provide additional information beyond basic metadata about documents and draw connections between research that may have otherwise gone unseen.

Academics struggle to stay abreast of emerging and transdisciplinary subjects. A digital environment built to

facilitate serendipity can help researchers make connections across disciplines and between topics.

We chose sustainability as a topic because it reaches across academic disciplines and professions. Issues of environmental stress, impacts on human society, energy studies and others are of interest to scholars from a wide array of disciplines, including environmental studies, economics, management, law, policy and urban studies. Librarians, scholars, graduate students, and policy researchers are all working to stay up to date in this field. Information silos make it difficult for researchers to do their work in sustainability—it can be difficult to navigate multiple content sources and stay abreast of the research in multiple fields. We believe having one central convening space built to facilitate cross discipline discoveries will create positive research outcomes.

2. USER INTERVIEWS

To understand what elements of a platform might facilitate serendipity, we conducted exploratory interviews and user research with approximately 40 scholars, policy researchers, and academic library subject specialists working in fields related to sustainability to understand how they conduct research in sustainability now and the impact of working within discipline silos.

Some key takeaways include:

- Researchers have difficulty “discovering” literature across disciplines and staying on top of a discipline outside their “home” field. Not having a central cross-discipline location to work from, inconsistent terminology and vocabulary across disciplines, and lack of expertise within highly specialized fields make it hard for researchers to stay current.
- Both historical and current academic materials are important as ties made to early materials can facilitate, support, and inspire new work.
- Grey literature reports and case studies are important. They often include policy outcomes and implications that are critical to understanding impact and further work in academia and outside of it.

These insights drove the initial prototypes of the Sustainability collection. A serendipitous environment must have aspects and tools that facilitate serendipity. Within a document, tagged keywords, abstracts, editorial reviews, highlights, and tweets about this article were elements we wanted to test as ways to facilitate serendipity. Across documents, topic pages, lists, and influential articles were ways to support further discovery.

3. ENABLING SERENDIPITOUS DISCOVERY

Topic Pages are the hub of serendipity facilitation in our prototype. We identified approximately 800 major topics in sustainability, such as agroforestry, urban ecology, biomass, and erosion control. For each, we programmatically created a topic page about it that includes linked data by way of entity extraction throughout. Each topic page contains sections for:

- introductory text and image from Wikipedia
- list of closely related topics (identified by topic modeling)
- links to the journals and authors who have written most about the topic
- visualization of, and links to, the most influential articles on the topic through time (identified by Eigenfactor)

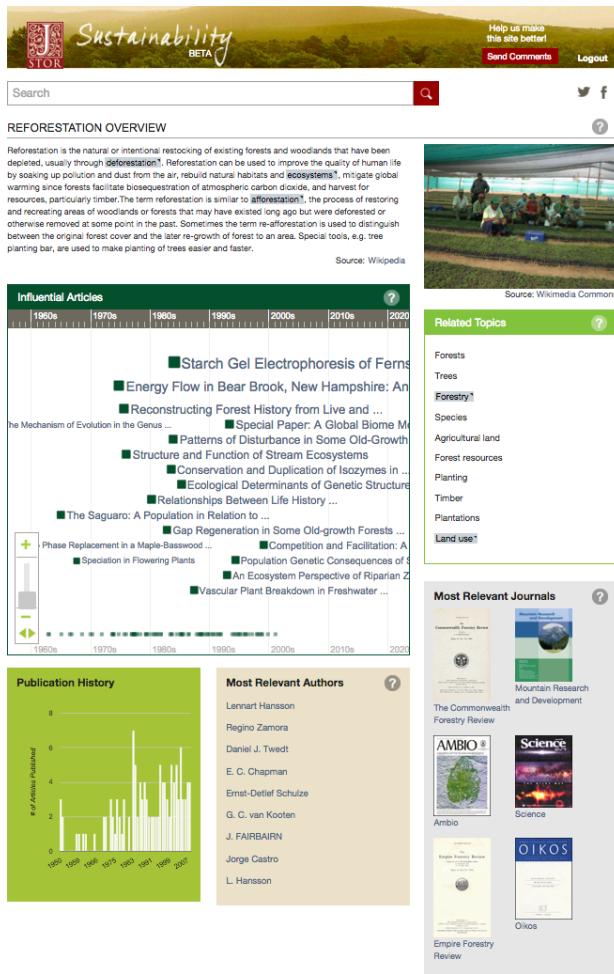


Figure 1. Screen capture of a topic page

These pieces of information allow someone who is less familiar with a particular research topic to get a quick grounding in the discipline and know where to start reading, which keywords to search for, who has been publishing on the topic, and which articles are historically and currently most impactful.

Throughout the prototype, there are numerous ties back to topic pages to facilitate serendipitous discovery. On the homepage,

links to the top six trending topic pages are prominently displayed. When a user searches for a term for which a topic page is available, we display an Info Card similar to what Google's Knowledge Graph shows within its search results. Within the search results, topics that appear in article keywords or article abstracts are linked to their respective topic pages.

The final aspect of the site that enables serendipitous discovery is an article recommender. When a user is viewing a particular article, we show a sidebar with "Background Reading". The two articles that are most influential to the specific topic of the original article are presented here, along with a link to a larger list of articles that influenced the same research area as the article being viewed.

4. FUTURE DIRECTIONS

After completion, the JSTOR Sustainability prototype was subjected to preliminary qualitative testing. From this testing, we've learned of a number of improvements that should be made to the prototype, both in terms of content and functionality.

We know one of the biggest shortcomings is a lack of up-to-the-minute content, which is especially important in sustainability. JSTOR has a wealth of resources to offer, particularly in the social sciences and humanities, but peer-reviewed content from sources other than those on JSTOR, particularly in the hard sciences, would increase the prototype's utility. Some grey literature has been added to the prototype—sustainability practitioners, particularly those outside of academia, rely on it more heavily than journal articles. Therefore we are negotiating with additional research institutes and other NGOs for permission to include their resources in the prototype.

When we list one topic related to another, we would like to include a co-occurrence number to show the strength of the relationship between the topics. We have also had requests for some standard tools that are included on sites like our main platform, jstor.org, such as advanced search and citation tools. Finally, we thought it would be a good idea to include user-generated content to surface new connections and discoveries, such as article summaries and comments, but we heard mixed reactions in user interviews and tests. Interviewees were concerned about the quality of user-generated content that would appear on the site, if the users were not vetted and their qualifications verified.

This prototype is currently available for beta testing, with a plan for full release in 2016.

5. ACKNOWLEDGMENTS

We thank the University of Washington Data Lab for assistance with their Eigenfactor metric.

6. REFERENCES

- Allen Foster, Nigel Ford, (2003) "Serendipity and information seeking: an empirical study", Journal of Documentation, Vol. 59 Iss: 3, pp. 321 – 340
- Cooper, J.W. and Prager, J.M. (2000), "Anti-serendipity: finding useless documents and similar documents", in Spragu, R.H. (Ed.), Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Maui, Hawaii, IEEE Computer Society, Piscataway, NJ.