

2018

Beyond Bibliographic Discovery: Bringing Concepts and Findings into the Mix

Athena Hoepfner

University of Central Florida, athena@ucf.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Hoepfner, Athena (2018) "Beyond Bibliographic Discovery: Bringing Concepts and Findings into the Mix," *Against the Grain*: Vol. 30: Iss. 1, Article 12.

DOI: <https://doi.org/10.7771/2380-176X.8001>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Beyond Bibliographic Discovery: Bringing Concepts and Findings into the Mix

by **Athena Hoepfner** (Discovery Services Librarian, University of Central Florida) <athena@ucf.edu>

Introduction

Since their introduction in 2007,¹ Web Scale Discovery (WSD), also known as Resource Discovery Services, have become the predominant tools for searching library collections and content in the broadest possible sense. Pre-harvesting metadata and content into one index lets discovery vendors normalize the indexing for efficient searching, apply consistent relevancy algorithms to create a ranked list of resource titles, design user interface features suited to the indexed metadata with fields and filters for refining results, and streamline linking to full text. Their current capabilities and content work well for bibliographic discovery but may not be the best solution for every information need.

Despite encompassing a significant percentage of library content and richly featured interfaces with sophisticated relevance ranking algorithms and an ever-growing suite of function, WSD do not meet every type of user need, as evidenced by the continued existence and use of subject indexes and other databases. Many of the databases adhere to the same basic approach that WSD uses: enter a search term, match the terms in the index, which is mostly document metadata, and present a list of results (document citations) for the user to interact. Two relatively new discovery tools depart from this approach and produce fresh takes on discovery.

Yewno for Concept Exploration

The first thing most anyone familiar with WSD services will notice about **Yewno** is that a search produces a graphical representation of concepts and connections instead of the familiar document title list. The differences do not stop there. WSD services collect metadata and text to create an index. **Yewno** collects full text content and uses artificial intelligence (AI) to identify concepts and recognize relationships. While the discovery interface does have a search entry form, librarian-centric approaches such as Boolean searching, fielded searches, known item, and even natural language searching do not work well for initial searches. Instead, **Yewno** works best when users enter one broad topic to start exploring.

Yewno's results display is a concept map. Circular nodes represent concepts and lines show semantic relationships between nodes. To the side of the concept map, **Yewno** shows an overview of the selected concept, similar to the **Wikipedia** summary in **Google** results. The user can click a tab to see a list of concepts, or a tab to see the list of documents in which the AI found the concept. The initial nodes start with very broad concepts. Users can expand the broad concept into subtopics that are more detailed and their connections. They can also add additional concepts and control how many concepts and connections display in the map. I am sure there are many features that I have not had the chance to try out, and the interface and content are still evolving.

Yewno's method of processing content is AI driven, applying computational semantics, graph theory, and machine learning to the full text of ingested documents. The AI is essentially reading the documents, similar to how humans read, infer meaning, and learn. The process results in a neural network model creating inferences. Using full text instead of metadata avoids problems that stem from inconsistent metadata. In WSD, variable accuracy and quality of metadata have a big impact on discoverability and relevancy ranking. In **Yewno** metadata quality does not affect the ability of the AI to process the paper nor the surfacing of a paper in the results.

While **Yewno** started with scholarly journals, the content included in **Yewno** Discover continues to grow and diversify. It contains over 125 million documents and is currently ingesting U.S. federal documents. Government document discovery has been very fragmented with huge metadata differences and decentralized sources for the full content. In addition to adding the government documents into their discovery system, **Yewno** is hosting the content to provide stable, consolidated access. **Yewno** has processed approximately 17 million government documents as of late January 2018.

Knowtro for Discovering Research Findings

Knowtro takes discovery in a very different direction than either WSD or **Yewno**. Where WSD indexes metadata and content for searching and display in a relevancy ranked list, and **Yewno** analyzes the content and illustrates relationships between concepts, **Knowtro** is a discovery tool for research findings. **Knowtro** analyzes scholarly articles to uncover key findings from published articles. A search on **Knowtro** brings up a list of cards, each with a single findings statement. The findings are stated as relationships between a predictor variable and an outcome variable with the size of the relationship between the two variables expressed as a percentage. The resulting statements on the card take the form of: *m* {predictor variable} increases/decreases/no effect on {outcome variable} by percentage. For example:

Research Suggests: **Job Level** increase **Career Satisfaction** by 58%

Knowtro targets learners, with the intent of making research more accessible by making it more broadly understood. Reading scientific articles is daunting, especially for novices who may not understand the terminology. **Knowtro** distills the research into easily parsed statements, making research findings exceptionally accessible.

Knowtro's vision of making research findings understandable for undergraduates and non-academics guides their content selection. To ensure quality, **Knowtro** selects papers from the top third of academic journals, based on impact factor. As a starting point, **Knowtro** looked at the most common undergraduate degrees. Within those disciplines,

Knowtro chooses studies that will resonate with their target audience, with topics that are likely to be salient. The resulting source set is much smaller than WSD, but perhaps with particular relevance to undergraduates and a higher rate of very-reliable findings presented in the sources. **Knowtro** includes over 70,000 findings.

Knowtro's interface offers a simple one-line search entry form. Like **Yewno**, **Knowtro** does not work with complex approaches to searching. One or two word concepts work best. **Knowtro** also provides a list of clickable concepts, which pull up results more reliably than searches. The search appears to run in the category and subcategories assigned to the findings. Neither article metadata nor full text of articles are indexed for searching. Likewise, the results page offers no filtering or refinement features.

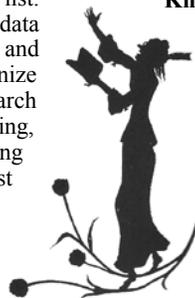
The Detailed View of the results is information dense and requires greater understanding of the **Knowtro** interface and icons set, and a good grasp of how research findings are expressed. The view shows a table with a row for each finding and columns for aspects of the research and finding. The table is sortable by clicking the column headers. The headers include: Unit; Report Outcome Variable; Time; Sample; Theory/Theme; r2; Author(s); and Year. I would love to be able to sort by more than one header and to see these elements available as filters in the Basic results display.

Clicking a result in the table opens up even more details about the research and the finding, with fuller descriptions of the predicting and outcome variables, a description of the methodology used (in some cases), and the full citation to the original paper. In addition there are buttons to see all the findings from the same paper, to show a formatted citation, and to link to the paper via DOI.

Potential Interactions for WSD, Yewno, and Knowtro

The products have potential for interesting interactions, pulling content from one system into others via widgets and APIs. I imagine a use case where a student starts in the familiar discovery search form on a library home page. They would see the typical relevancy ranked documents list, but with a call out that shows the concept mapping for the Subjects with the highest number of hits. The user could click

continued on page 29



into to widget to open the full **Yewno** interface and continue exploring there. In both the WSD and **Yewno** documents list, documents with **Knowtro** findings could show the top Findings card with an icon to see more findings. Ideally, there would be a filter to show only results that have **Knowtro** Findings. I am sure there are many other ways to make these products interact and enhance the knowledge discovery experience for users.

Conclusion

Yewno and **Knowtro** are exciting developments on the discovery. They bring the emphasis away from metadata and document discovery and rely on the full content of documents to enable concept exploration and bite-sized research findings. Their approaches to content handling are truly distinct and their discovery interfaces are also innovative, if not without some predecessors and contemporaries. I admit my first impression of **Yewno** brought to mind various early approaches to using concept maps in a discovery. While many concept map tools of the past were confusing and slow, **Yewno** may have created version that works well. They are not alone in giving the concept map another go. **IEEE's** InnovationQ has a very colorful concept map as a component in its interface, and I suspect there are many examples that I'm not aware of.

Knowtro also reminded me of products from the past, especially FirstSearch's FactSearch. Extracting snippets of content from articles and other sources is not a new idea, and several products still do just that, including RDS TableBase, **ProQuest's** Statistical Insight, and other databases. However, **Knowtro** takes the extraction a step further.

Rather than simply reproducing a snippet unchanged, they transform the content from difficult to understand and nearly impossible to search, to easily grasped and discoverable.

WSD, **Yewno**, and **Knowtro** each have their strengths and advantages. I cannot see **Yewno** or **Knowtro** supplanting the dominance of WSD for document searching, nor do I expect them to pull library spending from WSD. **Yewno** Discover is a subscription service with FTE based pricing. At the time of writing this article **Knowtro** is 100% free with the full version accessible on their website. All three of these tools help users address information needs and connect to the content in valuable ways. I hope they continue to develop, find their markets, and improve discovery for our students and researchers.

Links to the Products

Yewno: <https://about.yewno.com>

Knowtro: <https://www.knowtro.com>

Special thanks to **Michelle Valiani**, Channel Partner Manager and **Ruth Pikerling**, Co-Founder & Chief Business Development and Strategy Officer of **Yewno**, and **Sean R. McMahon**, Founder and President of **Knowtro**, for the long informative conversations. Without their patient explanations I would not have been able to write this article. 🐼

Endnotes

1. **Breeding, Marshall**. 2014. "Chapter 1: Discovery product functionality." *Library Technology Reports* no. 1: 5. General Reference Center Gold, EBSCOhost (accessed January 26, 2018).

Vendor Platforms — Tools for Efficient Library Acquisitions

by **Justin Clarke** (Director of Sales and Marketing North America, HARRASSOWITZ) <jclarke@harrassowitz.de>

Vendors supplying materials to libraries began developing online platforms or systems years ago in order to address the need to interface with libraries in an online environment. By allowing customers to work online, vendors can offer better presentation of material availability, improve acquisitions efficiencies, maintain and store historical purchasing activity, and more. These platforms might vary greatly in usability and performance but in general are meant to enhance the experience of working with vendors by allowing customers to research, tag, order, and track acquisitions all from the comfort of their online workspace.

As a global subscription agency and book dealer to academic and research libraries, **HARRASSOWITZ** maintains vendor systems to provide a friendly user experience to customers for managing acquisitions of all types of materials. Having been in continuous operation since 1872, we carefully evolved from traditional snail mail relationships with customers to embracing the word-wide web for its speed and efficiency. The days of tracking down damaged or missing paper orders, claims, and invoices are (mostly) behind us. Vendor online platforms are central to providing robust information and workflows to library

customers working in all areas of collection development, acquisitions, technical services, assessment, and beyond.

HARRASSOWITZ, for example, launched its first online system for managing subscriptions and standing orders in 1994 at the request of customers to have an easy-to-use platform that provided access to the full **HARRASSOWITZ** serials database. Then later in 2000, **HARRASSOWITZ** released the first version of OttoEditions, the online database for monographs and music scores, which was launched at the time with web based searching, library ordering, and claiming functions. These early systems met the needs of customers who no longer wished to peruse paper catalogs and generate orders dispatched via mail.

In 2005 OttoSerials, the **HARRASSOWITZ** online database for serials and continuations was further upgraded. OttoSerials moved to an entirely web based environment and additional functionality and profiles were added to maintain information about an institution and its descriptors used for providing quotes and verifying pricing. As e-journals gained quick popularity throughout the academic research community, subscription vendor systems were

again enhanced to accommodate additional e-resources related data such as format preferences, IP ranges and proxy information.

As technology rapidly advances, vendors respond to market needs by improving and refining their platforms to meet library workflow and data demand. In 2016 **HARRASSOWITZ** launched its Fokus system which replaced the OttoSerials interface for maintaining subscriptions and standing orders, and development is currently taking place to migrate OttoEditions into the Fokus system for management of approvals, firm orders, and music scores. Again, the latest technologies for web development were utilized, search engine speed and device optimization improved, and graphical interfaces and best practices were employed to provide a modern and enhanced user experience. Library supply vendors frequently consult with users when developing or enhancing online platforms via focus groups, library advisory boards, in-house visits, and surveys. **HARRASSOWITZ** also solicited customer feedback throughout the entire process of conceptualization, mockups, prototypes, early adopters, migration, and post-migration launch. It's important to offer feedback to your vendors

continued on page 30