# Primo Discovery:
# Search, Ranking, and Beyond

March 2015

# Contents

## Overview

Ex Libris Primo® with its Primo Central index enables library users to discover and access more than a billion global and local scholarly materials. The constantly growing collections available through Primo Central consist of traditional scholarly materials, such as journal articles and books; less traditional materials, such as videos and images; and related materials, including reviews and data sets.

The needs and expectations of users of a library discovery system vary. For example, a user may want a quick overview of a specific topic, the latest articles in her research area, material for an undergraduate course assignment or a research project, or simply to gain quick access to a specific book or article. Search and ranking technology relies on a multitude of factors and parameters to understand the user's intent and provide results that meet the user's expectations. The technology works in two stages: First it analyzes and optimizes the search to locate all relevant items; secondly it sorts the result list by relevance to provide the best material at the top. This document describes some aspects of the sophisticated technology that enables Primo to address users' diverse needs.

## Searching with Primo

An examination of search logs shows that users formulate their searches in many different ways. Primo analyzes a user's query and optimizes it so that the system can retrieve all relevant results, regardless of whether the user is searching for a known item or seeking items related to a certain topic. To optimize the query, Primo applies a number of techniques, such as the following:

- Identification of variations in terms and correction of spelling mistakes
- Expansion of searches based on word stems

- Expansion of search to full text if the original search yielded only few results

- Recognition of citation formats (for example, APA, MLA, or Chicago) and year of publication

- Recognition of compound words

First and foremost, Primo searches for the user's search terms in the metadata fields. To improve the user's results, search expansions are applied by Primo under certain circumstances. For example, if the initial search yields only few results, Primo extends it automatically to look for the query terms also in the full text of items; therefore, the search is likely to bring back additional results. Another example is the use of stemming. When Primo finds only few results when searching for the query terms as is, it will expand the search to include other variations of the terms. Primo gives more weight to the original version of the query than to term variations, and therefore it is likely to rank results based on such expansions lower than those based on exact matches of the search terms with the metadata or abstract of an item.

The Primo auto-complete function supports users further, by offering suggestions as a user enters search terms. The suggestions are derived from popular queries in the Primo search logs; from journal titles and subjects contained in the Primo Central index and from authors, titles, and subjects from the institution's local collections. By taking into account the institution's local collection, Primo ensures that auto-complete suggestions are drawn from items that are available to the user. [1]

---

[1] Auto-complete is a cloud service available for Primo cloud customers

**Primo auto-complete: popular searches and specific local titles**

In addition to supporting simple searches, Primo also enables users to submit more sophisticated searches, such as searches including Boolean operators and quotation marks.
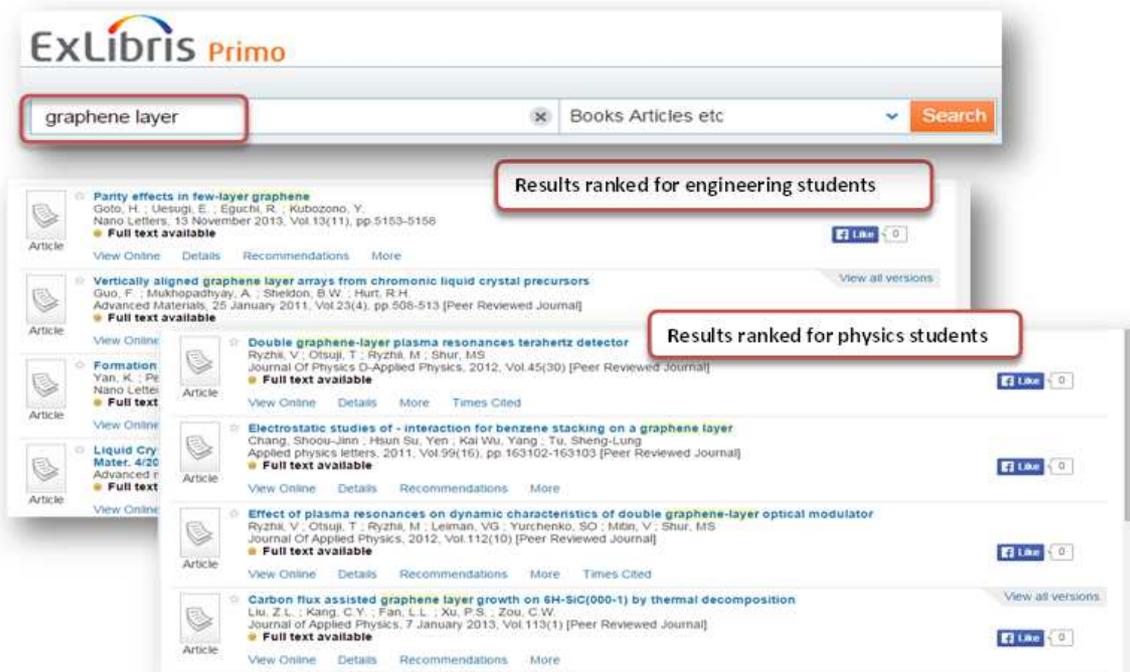
## Primo Relevance Ranking

The Primo ScholarRank technology sorts the search results by relevance on the basis of several criteria.

- **The degree to which an item matches the query**. For example, Primo accords greater weight to an item if the query terms occur in specific metadata fields of the item's record—primarily the author, title, and subject fields—and if the order of the query terms or phrases is the same in the query and the record.

- **A value score representing an item's academic significance.** The item's academic significance is calculated from factors unrelated to the query, such as whether the item was published in a peer-reviewed journal, how many times it has been cited, and what type of material it is, for example a journal article is considered more significant than a newspaper article.

- **An item's relevance to the type of search**. Primo infers the type of search that the user is conducting, such as a known-item search or a broad-topic search, and takes the type of material into account. In broad-topic searches, for example *social intelligence, mining engineering*, or *operator theory*, in which a user usually has no specific item in mind, Primo boosts overview material, such as reference articles. In known-item searches Primo takes into account authors, exact titles or other variations including entire citations to boost the appropriate item to the top of the result list. Large scale log analysis and usability studies are the core source to identify and define such search types, and to understand the variations and their significance.

- **The publication date (recentness) of an item.** Primo takes into account that users usually prefer newer materials.

## Customized Result Lists with Personalized Ranking

Primo enables users to tailor their result lists to their field of interest. This feature is especially useful when queries are relevant to various disciplines. Primo can apply information about a user's area of study to boost materials related to that discipline.

**Personalized ranking: same search, different ranking of results**
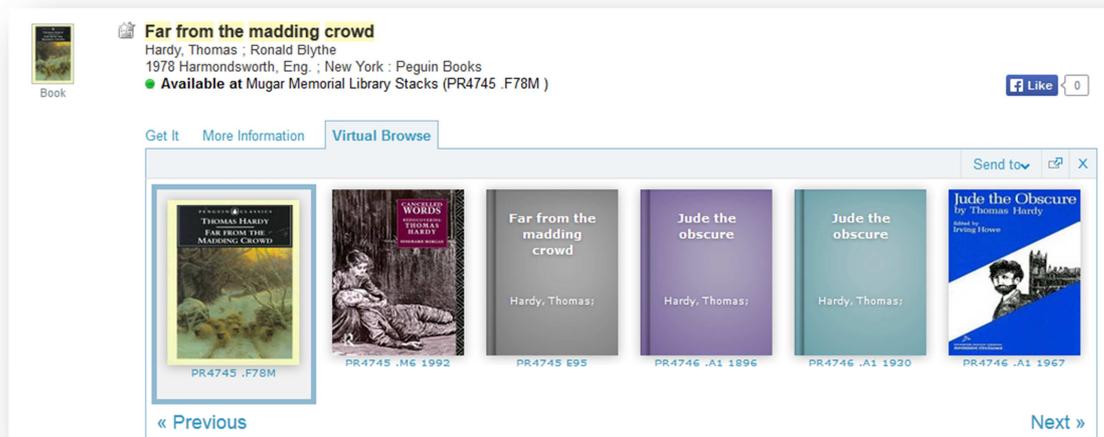
## Ranking of Local Collections

To meet their specific needs, libraries can adjust settings for the relevance ranking of their local materials. First, a library can boost certain metadata fields to give them more weight in the equation. Secondly, a library that chooses to blend results from different sources in a single result list—for example, to blend Primo Central results with results from local collections—can boost local items so that they appear higher on the result list than items coming from Primo Central, all other factors being equal.

## Exploration and Serendipitous Discovery with Primo

Discovery is not just about searching. Users do not always know exactly what they are looking for or how to find material for their needs.

Primo facilitates exploratory and serendipitous discovery in several ways:

- **Alphabetic browsing** for subjects, authors, titles and call numbers.
- **Virtual browsing**, enabling users who have located a book of interest to find similar items on a virtual bookshelf



**Virtual bookshelf**

- **Article recommendations** from the bX Recommender service[2], which enable users to view articles that are related to an article of interest and to navigate onward to additional sets of related articles. For example, the search query *sugar revolution* can relate to a number of topics and therefore yields a mixed result list, with items pertaining to history and economics as well as to diet and health. A user whose interest is history might notice *The Sugar Revolution* by B. W. Higman. When the user clicks the Recommendations tab, bX displays a list of articles related to the historic event, such as articles on slavery. This list can serve as a gateway for exploring specific aspects associated with the sugar revolution in history. For example, the user might choose the article "The Sugar Masters: Planters and Slaves in Louisiana's Cane

---

[2] bX recommendations are based on usage data collected from millions of researchers from institutions around the world. The bX Recommender is an optional service.

World, 1820–1860" to focus on the relationship between sugar and slavery in North
America.



After selecting this article, the user can view yet another list of related articles, which
deal with various aspects of slavery and the slave trade in North America.



## Quality Assurance and Enhancement Processes

The Primo search and ranking technology is subject to continuous monitoring and
improvements. Enhancements are triggered by a variety of factors: new insights gathered
from log analyses on how users search; customer feedback; and changes to the mix of

materials available through the Primo Central Index. Key performance indicators that are regularly monitored include the speed at which users make a selection after initiating a search and the position of selected items in the result list. Search data from the log files are used for large-scale quality assurance tests on an ongoing basis.

The expertise of library staff and the needs of users play an important role in the development of the Primo search and ranking technology. In 2014, Ex Libris formed a ranking technology advisory group, consisting of representatives from a number of Primo institutions. This group has been working closely with Primo product management to provide feedback that is integral to the enhancement-selection process. In addition, Ex Libris conducted user studies that yielded direct feedback from students, researchers, and library staff. Through these studies, Ex Libris has identified use cases that help the Development team optimize current and future enhancements and develop new search tools.[3]

---

[3] The user studies were presented at the Charleston Conference 2014 in a joint Harvard University-Ex Libris session. See http://sched.co/1p0HCQj.