RELEVANCE RANKING IN THE SUMMON® SERVICE

RELEVANCE RANKING IS ABOUT CONNECTING RESEARCHERS TO THE INFORMATION THEY ARE SEEKING—WHETHER THE USERS KNOW WHAT THEY’RE LOOKING FOR OR NOT. Usability studies and usage data analysis have shown that most researchers use the Summon® service in the same way they use open web search engines—entering natural language or short queries consisting of 1–3 keywords, then clicking on records found in the first page of results or using filtering tools to navigate to the content they need. Relevance, along with speed, ease of use and reliability, is essential and integral to “discoverability.”

The relevance ranking algorithms and architecture of the Summon service ensure that the results that best match a user’s query are returned at or near the top of results lists. Intuitive facets and filters applied to the complete results set make it possible for users to zoom in on precisely the results they need.
THE BOTTOM LINE: WHY SUMMON RELEVANCE RANKING IS SUPERIOR

In order for any discovery service to meet its promise of being easy to use, delivering highly relevant results to users, and increasing usage of library resources across the library’s collection, relevance ranking must work differently than traditional databases and OPACs.

Looking to open web search technologies that set the bar for user expectations, while keeping firmly rooted in the need to provide a service that supports academic research, a combination of key factors play a role in making relevance ranking in the Summon service superior:

Architecture—A single, unified index makes it possible to apply relevance and offer features in ways not possible by other discovery services.

Records Optimized for Discovery—Unique match-and-merge technology ensures users are searching the richest and highest quality of information about an item.

Dynamic Rank + Static Rank—The combination of ranking algorithms working in concert takes into consideration the information about the item itself as well as matching the user’s exact query with metadata and full text in the Summon index.

Recommendations—Beyond the results list, recommendations guide users to additional relevant content and assist them in refining their searches.

Continuous refinement of the relevance ranking algorithms is one of the key strengths of the Summon service. In order to continue to meet changing user expectations as well as accommodate a constantly growing index of content, dedicated expert staff monitor and tune the relevance ranking algorithms guided by these goals:

Superior Recall—Ensure that all possible items from across the library’s collection, no matter the source, with some level of relevance to a user’s query are returned in the results set. This is critical to providing facets and limiters that allow users to zoom in on relevant material and ensure that there is no bias in result sets.

Superior Precision—Ensure that the most relevant items to a user’s query are returned at the top of the results set.

Superior Rights Management—Ensure results returned are items available to the user through their library. Beyond discovery, users want immediate access to the most relevant content.

Ultimately, superior relevance ranking contributes to the Summon service’s ability to deliver an experience that truly meets user’s needs and expectations.

THE BACKGROUND: WHAT CONTRIBUTES TO SUMMON RELEVANCE RANKING

Superior relevance ranking is dependent on the underlying architecture of a discovery service. The Summon service was built to purpose using state-of-the-art technology rather than repurposing older database platforms or next-generation catalog structures.

With Summon’s single, unified index architecture, relevance ranking algorithms are applied equally across the entire body of indexed content, and results are returned in a single, unbiased, relevance-ranked results list. The Summon service does not “blend” results from disparate sources or use separate and multiple indices to include institutional repositories, archives or catalogs. These approaches do not provide relevance across all results, lack content and vendor neutrality, and negatively affect the speed in which results are returned.

Unlike other discovery services, the Summon service is the first and only discovery service based on a unified index of content, leveraging its unique match-and-merge technology to combine rich metadata and full text from multiple sources into a single record optimized for discovery. The Summon service doesn’t simply display the record with the most metadata (thereby biasing results) or group records forcing users to select one record from the grouped set. With match-and-merge technology, metadata and full text from multiple sources are brought together in a single Summon record for an item that contains more metadata than any one provider’s single record.
With a growing index currently consisting of more than 1.1 billion items spanning 90+ content types, relevance ranking in the Summon service is applied across a content corpus 500 times larger than the average library catalog.

Not a "set it and forget it" field-weighting formula typically employed by library databases and catalogs, the Summon relevance algorithm is continuously tuned to accommodate newly added content, insights revealed by usability testing and usage data analysis, and direct feedback from users and librarians.

THE BASICS: HOW SUMMON RELEVANCE WORKS

The Summon service relevance ranking consists of scoring two sets of information about an item according to an algorithm. These two scores are called **Dynamic Rank** and **Static Rank**, and they work in concert to deliver the best results for a user’s query. Scoring is relative to all the other items that match a user’s search, whether a results list has one hundred items or one million.

In addition to relevant results, the Summon service presents users with a variety of **recommendations** based on relevance scoring and other inputs that provide users with valuable contextual guidance at the point of need. Recommendations steer users to additional highly relevant content and provide proactive suggestions for refining searches in ways that lead to better search outcomes.

**Dynamic Rank**

Dynamic rank is the most important part of the Summon relevance score because it focuses on matching up a user’s exact query with all of the metadata and full text in the Summon index. Among other things, the Summon relevance algorithm leverages industry standard concepts such as:

- **Proximity**—How close together are the query terms found in the record? If closer together, then a record is more relevant.
- **Term Frequency**—How often do the query terms appear in the record? The more often a query term appears, the more relevant the record.
- **Inverse Frequency**—If a query term appears less often across the entire corpus of content searched, but is found in a record, that record is more relevant.
- **Field Weighting**—Where is the query term found? Finding query terms in some fields are more important than others. For example finding a query term in the title is more important than finding that term in the full text.

In addition, Dynamic Rank takes advantage of other advanced search concepts that ensure users can be successful searching the way they normally do with natural language and keyword search queries. These concepts include:

- **Term Stemming**—The most accurate form of the word is ranked higher than a slightly different version of the word.
  
  Note: Exception-based stemming helps to avoid problems such as "Robert ≠ Roberts" "aid ≠ aids"

- **Stop-Word Processing**—Stop-words are or aren’t used in the search based on their importance to the search phrase.
  
  Note: Intelligent stop-word handling addresses searches such as "to be or not to be" or "man of the year"

- **Synonyms**—Different variations of query terms are searched as well to ensure users don’t miss out on the most relevant content. For example, a search for "heart attack" will also match on items about "myocardial infarction."
  
  Note: Users are able to turn off this feature to reduce false positives.

- **Language Processing**—Result relevance is influenced by the interface language a user has selected. For example, if a user is using the German-language Summon interface, content in German will be considered more relevant.

- **Free-Form Identifiers**—Identifiers (DOI, ISSN, ISBN, call number, and more) in the query are detected and used to influence results.

- **Cut-and-Paste Excerpts**—Support a common search technique where researchers cut-and-paste a citation or excerpt as a search query in order to retrieve not only that item but also other works that might cite that reference.
Static Rank

Static rank also contributes to the overall “ranking” of a record in the results. While Dynamic Rank focuses on the query, static rank focuses on the item itself. Static ranking helps boost relevance based on attributes of an item that can be critical in a research environment. Static factors include:

- **Content Type**—Some content types are boosted over others, such as books and journal articles over newspaper articles and book reviews. This ensures that a search for the journal *Nature* returns the record for the journal first and articles from the journal lower in the results.
- **Scholarly/Peer-Reviewed**—Scholarly content is boosted over non-scholarly content.
- **Publication Date**—Generally, items with newer publication dates are favored over older items.
- **Citation Counts**—Although most academic journal articles are cited fewer than 10 times, more highly cited items are boosted so they don’t get lost in large sets of search results. *Web of Science* citation counts, as well as other sources, are used to boost relevance regardless of whether libraries subscribe to these sources.
- **Local Collections**—Content from local collections, such as institutional repositories, is boosted.
- **Content Size**— Longer works aren’t necessarily more relevant. Taking content size into consideration ensures that the full text of *War and Peace* doesn’t outrank the full text of a key journal article that is only two pages.

Recommendations

The third component of relevance and key in leading users to relevant results is recommendations. A variety of features in the Summon service offer guidance to users at the point of need. These recommendations guide users to additional relevant content and assist them in refining their searches.

- **Library-Curated Recommendations**—Libraries are able to direct users to additional relevant information via “Best Bets.” Completely controlled by the library, Best Bets promote specific library resources such as research guides, specialized collections, library web pages, course reserves, announcements, current events, important contacts, library hours, help tools and more.
- **Database Recommender**—Blending library control, community-sourced tags and relevance-based recommendations, this features points users to specialized databases for targeted research and discipline specific searching.
- **Related Search Suggestions**—To encourage exploration of related concepts, related search suggestions dynamically display, encouraging users to expand their query to aid their research. Leveraging real-time, global Summon usage data, related searches offer users scholarly and multilingual suggestions for query expansion and refinement.

TO LEARN MORE about benefits of the Summon service, visit us at www.serialssolutions.com/summon or contact your rep @proquest.com.