# Proposal to Turn Off Summon Topic Explorer at GVSU

# Executive Summary

We should turn off the right-hand sidebar of Summon, which provides contextual information because:

* Wikipedia and other encyclopedic entries are often inaccurate or misleading, due either to choices made in indexing or because of the way they are written and abstracted automatically by algorithms.
* Nearly 1% of Topic Explorer Reference results are biased against already marginalized groups, like women, blacks, the LGBTQ community, the mentally ill, and Muslims.
* Localized algorithms that work to show GVSU-specific results rely on third-party tools and shoddy assumptions from the engineers, resulting in unreliable and inaccurate results.
* Other institutions are pushing back against these problems by turning off the sidebar.

# Details

For the past 3 years, I have been researching the accuracy and effectiveness of the University Libraries' Summon Discovery Service algorithms, and in particular, the algorithms that make up the "Topic Explorer," the contextual information that makes up the right-hand sidebar of the search results screen. Based on my research, I find that these algorithms often cause more harm than good, and should be turned off in GVSU's instance of Summon. Results that show bias in nearly 1 percent of the Topic Explorer results. What's more, poor infrastructure design of the Topic Explorer compounds the problem, showing biased and inaccurate results more and more frequently.

Wikipedia, the most common reference source in Summon, is useful for libraries to include because users trust Wikipedia to have up-to-date content. However, Wikipedia entries in Summon are not pulled from Wikipedia's updated content. In the Summon of 2019, Ruth Tillman of Penn State University Libraries and I discovered that the Summon team loaded Wikipedia results into the Summon index at some time before February 20, 2013, a full month before the Topic Explorer was announced in a press release. **They have never updated the results.** (Brent Cook, the project manager for Summon, reluctantly confirmed this finding.) Now searches for living individuals, such as Barack Obama and Donald Trump, are wildly inaccurate. (Obama is listed as the 44th and *current* president of the United States. Trump is a reality TV star and real estate developer.) Many more recently deceased individuals are listed as alive, such as Barbara Bush. If the Topic Explorer cannot provide correct information, it is not useful to our users, and will degrade their trust in our other services.

In addition, nearly 1% of all results show bias against blacks, LGBTQ people, women, the mentally ill, Muslims, and more. Searches for information on stress in the workplace returned a result for "women in the workforce," and searches for "rape in United States" showed a a result for "Hearsay Evidence." (Ex Libris has blocked these particular results, but not addressed the underlying issues in the search algorithm.) Any search with the words "mental illness" returns a Topic Explorer result for "The Myth of Mental Illness," despite my reports in January of 2016 that this was unacceptable. Many more examples can be found in my research.[[1]](#footnote-1)

In some instances, both of these problems merge together. Chelsea Manning, a transgender woman who served prison time for violations of the Espionage Act, is still listed in Summon only as "Bradley Manning," her dead name. Not only is this article out of date, but the act of deadnaming a transgender person is to deny their actual identity.

Other reference sources are not designed and written to be excerpted by algorithms. In many cases, Credo Reference articles start with some tangential preamble, rather than being structured like an inverted pyramid (as Wikipedia's articles are). This can lead to entries like one for "alcohol consumption," which shows the Credo entry for alcohol that begins, "Prisoners are not allowed to drink alcohol while they are in prison," implying that alcohol and incarceration are connected. A similar search for "alcoholism" (until recently) began, "The history of women's relationship with alcohol constitutes a profound commentary on U.S. cultural attitudes about gender and power." This implies that alcoholism is a gender-specific issue.

Related topics are another area where the Topic Explorer shows bias, such as a search for "women in prison" shows a related search of "sex in film," as if women in prisons must be related to sexploitation films. (The reference result for this search is also "Women in prison films.") Searching for "murder" or "lying to patients," two unethical practices, recommends searching for Islamic dietary laws. "Schizoaffective disorder" is connected by related searches to both "cocaine addiction" and "pedophilia," despite having no logical connection at all.

Of the other algorithmic results shown in the Topic Explorer, including recommended librarians and guides, the assumptions the engineering team made about how these would work has introduced a number of problems. Based on keyword matching, we have the wrong librarian listed for a number of subjects. For instance, the owner of the modern languages guide "Spanish for Business" is always listed as the business liaison, because the numeric guide "id" in the LibGuides database is lower than the actual Business guide. What's more, in some cases basic word proximity errors lead to strange match-ups, like Debbie Morrow, our engineering, math, and physics liaison being listed as a subject expert for Capital Punishment, because one of her guides uses the phrase "questionnaire *execution*."

While some of these problematic searches have been suppressed since they were discovered, there will continue to be more biased and incorrect results, like a game of software whack-a-mole. We would not be alone in turning off the Topic Explorer. Most recently, Penn State University Libraries turned off the TE after Ruth Tillman of Penn State University Libraries and I uncovered the inaccuracies in Wikipedia article matching. The right-hand sidebar can be turned off with one option in the Summon Administration Console. Usage data is difficult to get, because much of the sidebar is designed to be read, not necessarily acted upon. What data we do have, however, suggests that clicks on recommended searches happen in less than a tenth of one percent of all searches, while the number for clicks on recommended guides and librarians is even lower.

Written by: Matthew Reidsma, Web Services Librarian

Updated: January 21, 2019

1. Reidsma, M. (2016). Algorithmic bias in library discovery systems. Retrieved from <https://matthew.reidsrow.com/articles/173>. See also the forthcoming Reidsma, M. (2019). *Masked by trust: Bias in library discovery.* Sacramento, CA: Library Juice Press. [↑](#footnote-ref-1)