PostScholar: Surfacing Social Signals in Google Scholar Search

Larry Chan*
Shagun Jhaver*
Sandeep Soni*
School of Interactive Computing
Georgia Institute of Technology
larrychan@gatech.edu
jhaver.shagun@gatech.edu
sandeepsoni@gatech.edu
* All authors contributed equally

Abstract

PostScholar is a service that augments the results returned by Google Scholar, a search engine for academic citations. PostScholar detects the social media activity related to an article and displays that information on the search results page returned by Google Scholar. This enables Google Scholar users to interpret the social media impact of an article, in addition to the citation impact provided by Google Scholar. The PostScholar service is implemented as a browser extension, and can be deployed across a variety of devices.

Author Keywords

Twitter; commentary; social media

ACM Classification Keywords

H.3.5 [Information Search and Retrieval]: Online Information Services - Web-based Services

Introduction

An ever increasing count of scholarly articles makes searching for them increasingly difficult. The number produced in 2006 alone is estimated to be 1.5 million [1]. For this reason, finding research articles relevant to one's purpose or interests typically entails the use of tools, particularly databases and search engines.

Google Scholar[1] (GS) is a free, online service that presents a simple interface to users while enabling them to find scholarly work using a variety of search terms. Scholarly work has been made more accessible by the GS service in a number of ways. Firstly, while the use of academic citation databases frequently requires payment of fees, and access is granted to members of those institutions that have the means to pay those fees, GS is free for anyone to use. Secondly, the citations returned as results from this search engine are in some ways more inclusive than those from other popular citation databases, such as Web of Science [2] [2]. This is because GS draws from a wider range of resources. Finding citations made by conference papers and dissertations, that are not necessarily published in traditional peer-reviewed journals, yields a more complete view of the impact of an article.

An even more inclusive impression of the impact made by a scholarly paper can be achieved by incorporating some of the more spontaneous responses to it. The general perception of a research paper might be better summarized by incorporating the amount of discussion that it evokes in social media.

To that end, PostScholar searches Twitter for posts that refer to the article in a GS record and makes them available as part of the GS result for that record. While GS already includes a hyperlink indicating the number of times a record has been cited by other scholarly works, PostScholar adds another link that indicates the number of relevant posts made on Twitter. Additionally, the service gives quick access to the posts themselves, and information about the posters. Exposing the *social life* of a paper in this way bridges the boundary between the slower, selective inter-

actions of academic publication processes and the fast, inclusive dialog of social media platforms.

Related Work

Some systems like Altmetric[3] and Impactstory[4] have been developed to harness social media data to be used as alternative metrics to measure the impact of a paper. But these systems have significant limitations. Altmetric is a commercial service requiring site licence access, and, at the time of this submission, offers data on fewer than 3 million papers. ImpactStory organizes information into author profiles, where the impact of that author's papers can be displayed. A researcher must first identify an author of interest before seeing the social media impact of her work. In contrast, PostScholar allows a researcher to select an author on the basis of that impact. PostScholar is free, and it can retrieve existing social media discussion for a very large number of records because the records are gleaned from the search results returned by GS.

System Design and Implementation

From the perspective of a user, PostScholar is a browser extension that is active only on the Google Scholar page. After installing and enabling the browser extension, a user can perform literature searches on GS in the usual way. No additional procedure is necessary. When GS returns results for a query, an additional hyperlink appears in the results for each record that has Twitter activity associated with it. Figure 1 shows the GS results page with PostScholar enabled. The text of the hyperlink (A) indicates the number of Twitter posts found for that record. Hovering over the hyperlink displays a tooltip (B), which shows the number of unique posters and the date of the most recent post. Clicking on the hyperlink (A) displays all Twitter postings on the article (Figure 2).

---

[1] scholar.google.com
[2] https://webofknowledge.com

[3] www.altmetric.com
[4] https://impactstory.org/
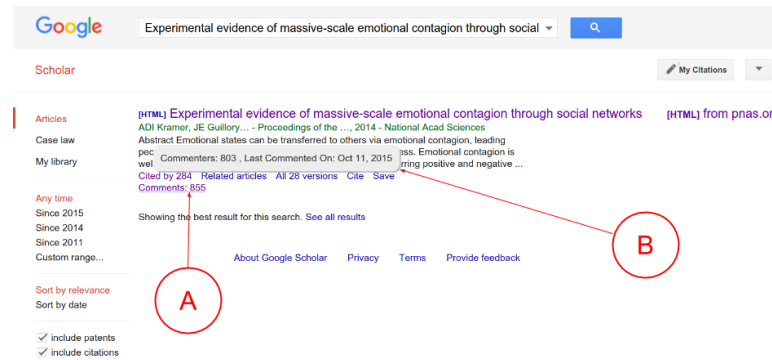
Figure 1: PostScholar results for a query (Google and the Google logo are registered trademarks of Google Inc., used with permission).

The PostScholar browser extension communicates with a server in order to accomplish the features described. When GS returns the results of a query, the PostScholar extension extracts information like title, url, authors, etc from each record and sends this information to the PostScholar server. That server then searches Twitter for posts that match the retrieved information for each record. We calculate the sentiment score for the fetched posts for each record using sentiment data dictionaries. Relevant Twitter posts are stored in a database on the server and sent back to the browser extension, along with summary and sentiment information. That summary information is displayed in the hyperlink text and in the tooltip text, which are HTML elements inserted into the DOM by the browser extension.

If a recently processed record shows up again in the search results, PostScholar bypasses twitter search and retrieves the information directly from the database. If the user clicks on the comments hyperlink for any article, the corresponding posts are retrieved from the database and displayed to the user (Figure 2). If no posts are found, no HTML elements are inserted for that record.

Limitations and Future Work
Currently posts are displayed in reverse chronological order. We plan to give users the option to sort the posts by the prominence of the poster. One measure of prominence for a Twitter poster could be the number of followers. We also plan to fetch and include posts from other social media platforms.

References
[1] Bo-Christer Björk, Annikki Roos, and Mari Lauri. 2008. Global annual volume of peer reviewed scholarly articles and the share available via different Open Access options. In *ELPUB2008*.
[2] Anne-Wil Harzing and Ron Van der Wal. 2007. Google Scholar: the democratization of citation analysis. *Ethics in science and environmental politics* 8, 1 (2007), 61–73.

Figure 2: External webpage showing tweets when hyperlink on posts is clicked