

# Technical Disclosure Commons

---

Defensive Publications Series

---

May 2022

## Semantic Indexing and Retrieval of Web Pages in a Web Browser

Duc-Hieu Tran

Follow this and additional works at: [https://www.tdcommons.org/dpubs\\_series](https://www.tdcommons.org/dpubs_series)

---

### Recommended Citation

Tran, Duc-Hieu, "Semantic Indexing and Retrieval of Web Pages in a Web Browser", Technical Disclosure Commons, (May 23, 2022)

[https://www.tdcommons.org/dpubs\\_series/5152](https://www.tdcommons.org/dpubs_series/5152)



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

## **Semantic Indexing and Retrieval of Web Pages in a Web Browser**

### **ABSTRACT**

This disclosure describes techniques for semantic indexing and retrieval of web pages. Per techniques of this disclosure, a web browser includes a semantic bookmark system that indexes the web page content and enables subsequent retrieval of the web page via a semantic search. A content extraction model is applied to extract the content of a web page that is bookmarked. The extracted content is semantically indexed, linked with the corresponding uniform record locator (URL), and stored. During web page content retrieval, bookmarks corresponding to a user search query are retrieved and provided to the user. Semantic indexing and retrieval enables users to retrieve a web page that includes content semantically linked to the search query, even when the query does not contain words that are explicitly included in the title or content of the web page.

### **KEYWORDS**

- Web browser
- Bookmark
- Semantic search
- Content retrieval
- Favorites

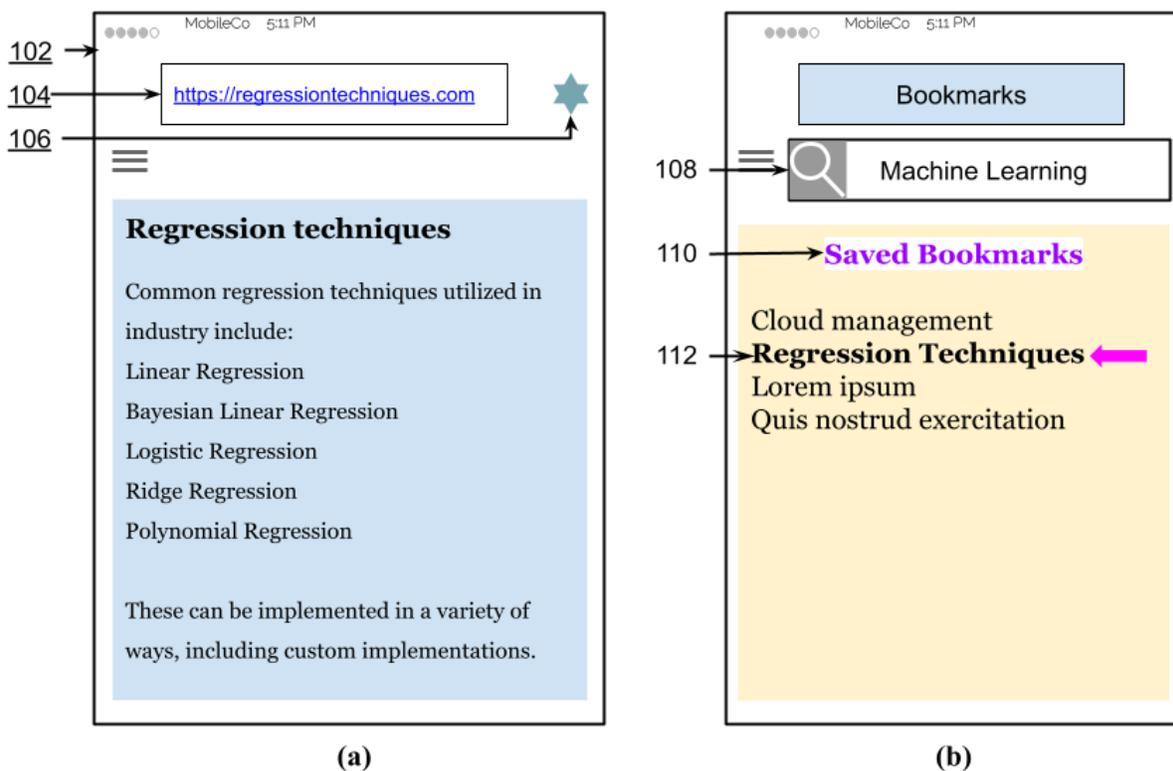
### **BACKGROUND**

Web browsers include a bookmarking feature that lets users bookmark webpages of interest for subsequent access, reference, and/or retrieval. Most implementations of the bookmark feature utilize the page title for indexing and for displaying a list of previously saved bookmarks to the user. Subsequent retrieval of the web page by the user requires the user to

either scroll through their saved bookmarks or remember at least some of the exact words in the page title.

DESCRIPTION

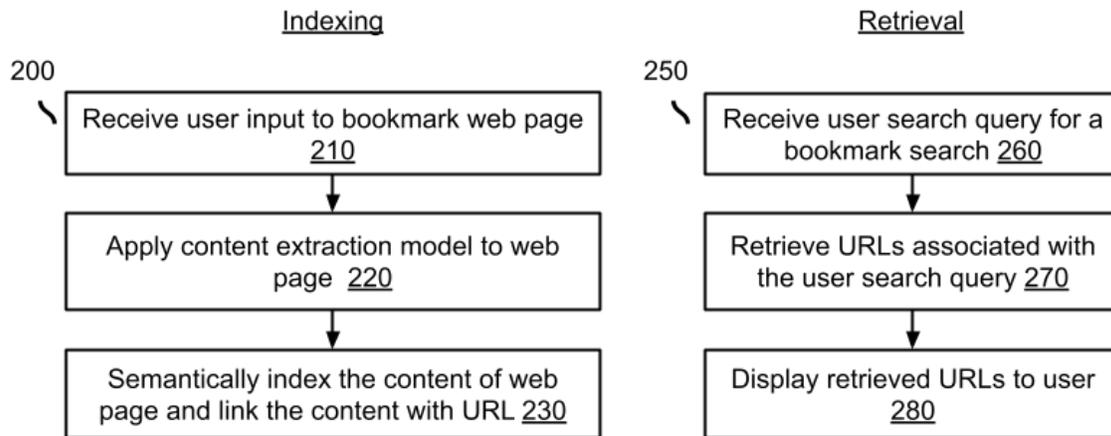
This disclosure describes techniques for semantic indexing and retrieval of web pages. Per techniques of this disclosure, a browser includes a bookmark feature (semantic bookmark system) that indexes the web page content semantically and enables subsequent retrieval of the web page via a semantic search.



**Fig. 1: Semantic indexing and retrieval of web pages**

Fig. 1 depicts an example of semantic indexing and retrieval of web pages, per techniques of this disclosure. Fig. 1(a) depicts a web browser (102) that is currently displaying a web page (104). The user is provided with the ability to bookmark (106) the web page for subsequent retrieval using a bookmark (or favorites) feature built into the browser.

Fig. 1(b) depicts subsequent retrieval of the web page by the user. As depicted in Fig. 1(b), the user can access a list of saved bookmarks (110) and select a web page of interest. The user is also provided with the ability to perform a search (108) of the saved bookmarks. Per techniques of this disclosure, semantic search and retrieval is enabled whereby the user can perform a keyword search and retrieve web page(s) of interest from the saved bookmarks. In this illustrative example, a web page with the title “Regression techniques” is retrieved and highlighted to the user based on the search query “Machine Learning,” which is semantically associated (linked) with the content of the web page, even without a match to the exact words.



**Fig. 2: Workflow for semantic indexing and retrieval of web pages**

Fig. 2 depicts example methods for the semantic indexing (200) and retrieval (250) of web pages, per techniques of this disclosure. User input to bookmark a web page is received (210). A content extraction model is applied (220) to extract the content of the web page. The extracted content is semantically indexed, linked with the corresponding uniform record locator (URL), and stored (230). During retrieval, a user search query is received (260). A search is performed on the saved bookmarks. Using the semantic index, URLs corresponding to the user search query are retrieved (270) and are provided (280) to the user.

Indexing the contents of bookmarked web pages and retrieval based on semantic matching enables retrieval of bookmarks that include content semantically linked to the search query. This eliminates the necessity for the user to enter a query that includes word(s) that are explicitly included in the title or content of the bookmarked web page.

Further to the descriptions above, a user is provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may enable the collection of user information (e.g., information about a user's bookmarks, browser settings, or a user's preferences), and if the user is sent content or communications from a server. In addition, certain data are treated in one or more ways before it is stored or used so that personally identifiable information is removed. For example, a user's identity is treated so that no personally identifiable information can be determined for the user. Thus, the user has control over whether and what information is collected about the user, how that information is used, and what information is provided to the user.

## CONCLUSION

This disclosure describes techniques for semantic indexing and retrieval of web pages. Per techniques of this disclosure, a web browser includes a semantic bookmark system that indexes the web page content and enables subsequent retrieval of the web page via a semantic search. A content extraction model is applied to extract the content of a web page that is bookmarked. The extracted content is semantically indexed, linked with the corresponding uniform record locator (URL), and stored. During web page content retrieval, bookmarks corresponding to a user search query are retrieved and provided to the user. Semantic indexing and retrieval enables users to retrieve a web page that includes content semantically linked to the

search query, even when the query does not contain words that are explicitly included in the title or content of the web page.

## REFERENCES

1. Andreica, Mugurel Ionut, Jonathan Small, and Marcin M. Nowak-Przygodzki.  
"Annotation and retrieval of personal bookmarks." U.S. Patent Application 16/972,250,  
filed August 5, 2021. Assigned to Google LLC.
2. Braun, Simone, Valentin Zacharias, and Hans-Jörg Happel. "Social semantic  
bookmarking." In International Conference on Practical Aspects of Knowledge  
Management, pp. 62-73. Springer, Berlin, Heidelberg, 2008.