

Simplifying Information Retrieval Studies with Data Collection Software

Presenter: Samantha Lee
Advisor: Dr. Chirag Shah



ABSTRACT

As people become increasingly reliant on the Internet as a source of information, **information retrieval** becomes more complex in nature, and researchers seek to understand patterns between users and how they approach a web search. User studies serve the purpose of shaping to researchers' particular needs and providing crucial data on web search behavior through assigned tasks, yet present challenges while in development and are often tedious to design.

Coagmento is a software that helps researchers by streamlining the process of creating and running a study. The content of the poster pertains to

Coagmento's **custom Chrome extension**, which reads various pieces of data through JavaScript functions. In determining what pieces were necessary to collect and so what the extension should and should not filter, Chrome Developer documentation proved essential to understanding the extension's capabilities and limitations. In its current state, Coagmento records data on passive inputs (clicks, scrolls, keystrokes, etc.), changes in the browser (tab and window actions), and web navigation (pages and queries) and pushes it into a **MySQL database** for easy accessibility. From there, a Python script extracts browsing features from the search session that researchers can utilize as training data for **machine learning**.

Coagmento is a web application that simplifies information retrieval studies with a custom Chrome extension.



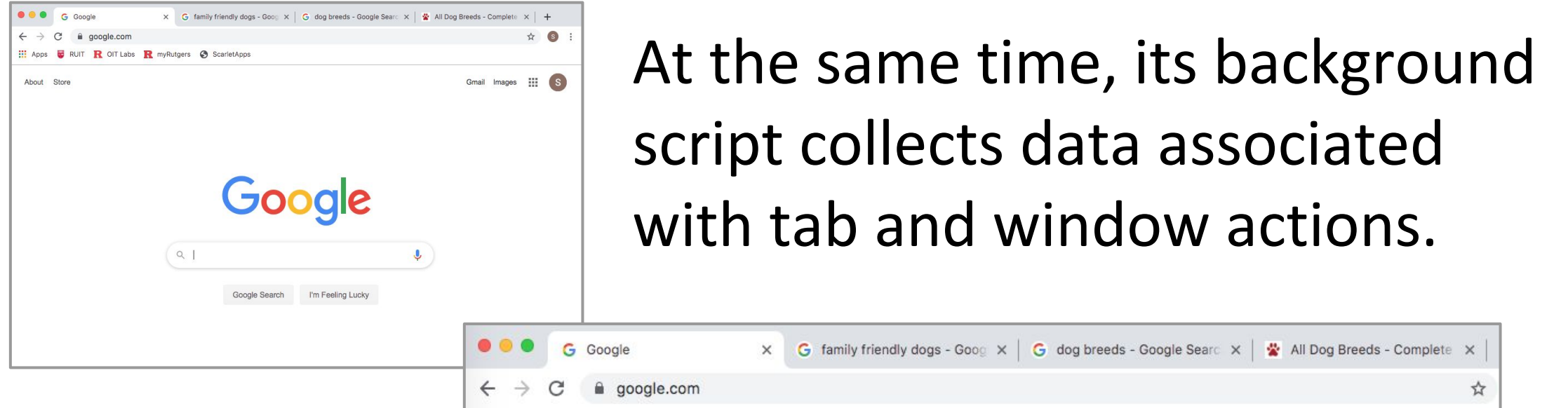
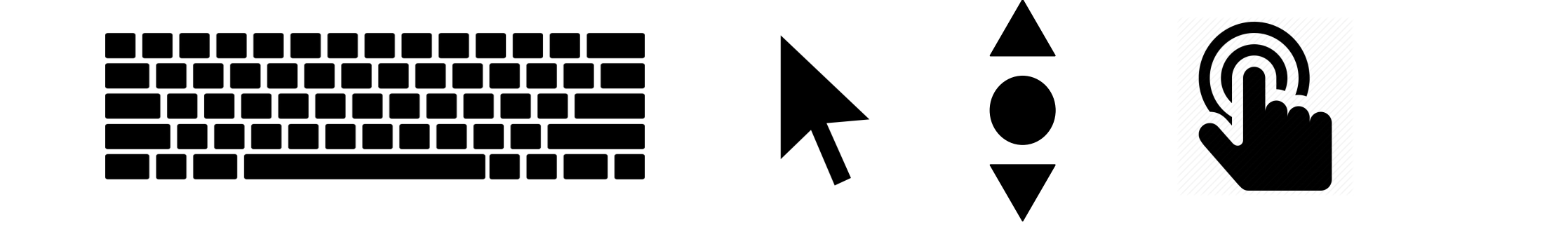
← Check out Coagmento!

METHODS

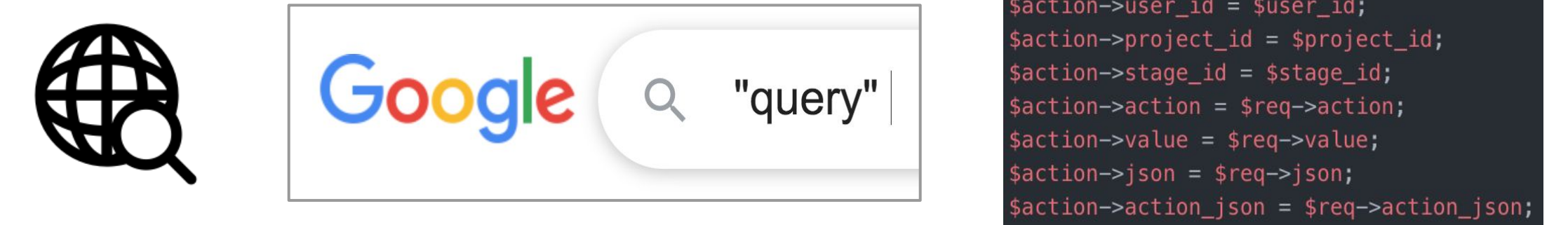


An overview of Coagmento's dataflow.

The extension's content script collects data associated with keystrokes, scrolls, clicks, and mouse movements.



The streams of data are sent to Coagmento, where files called controllers organize them and derive details of the pages visited and queries searched.



Finally, Coagmento sends the data to a database.

RESULTS & DISCUSSION

The database presents all of the user's actions clearly, allowing for easy access during a study.

id	user_id	project_id	stage_id	action	value	json	action_json	created_at_local
32427	4	1	2	webNavigation.onCommitted	1768	["active":true,"audible":false,"auto..."]	["frameId":0,"parentFrameId":-1,"p...	2019-08-28 22:09:30
32428	4	1	2	tabs.onUpdated	1768	["active":true,"audible":false,"auto..."]	["status":"complete"]	2019-08-28 22:09:31
32429	1	1	1	webNavigation.onCommitted	1768	["active":true,"audible":false,"auto..."]	["frameId":0,"parentFrameId":-1,"p...	2019-08-28 22:09:40
32430	1	1	1	tabs.onUpdated	1768	["active":true,"audible":false,"auto..."]	["status":"complete"]	2019-08-28 22:09:40
32431	1	1	3	webNavigation.onCommitted	1768	["active":true,"audible":false,"auto..."]	["frameId":0,"parentFrameId":-1,"p...	2019-08-28 22:09:52
32432	1	1	3	tabs.onUpdated	1768	["active":true,"audible":false,"auto..."]	["status":"complete"]	2019-08-28 22:09:53
32433	1	1	3	webNavigation.onCommitted	1768	["active":true,"audible":false,"auto..."]	["frameId":0,"parentFrameId":-1,"p...	2019-08-28 22:09:55
32434	1	1	3	tabs.onUpdated	1768	["active":true,"audible":false,"auto..."]	["status":"complete"]	2019-08-28 22:09:56
32435	1	1	3	windows.onFocusChanged	-1	NULL	null	2019-08-28 22:10:17
32436	1	1	3	windows.onFocusChanged	28	["alwaysOnTop":false,"focused":tru...	null	2019-08-28 22:10:23
32437	4	1	1	windows.onFocusChanged	-1	NULL	null	2019-08-28 22:23:07
32438	4	1	1	windows.onFocusChanged	28	["alwaysOnTop":false,"focused":tru...	null	2019-08-28 22:23:09
32439	4	1	1	tabs.onHighlighted	1921	["tabId":1921,"windowId":28]	["tabId":1921,"windowId":28]	2019-08-28 22:23:10
32440	4	1	1	tabs.onActivated	1921	["active":true,"audible":false,"auto..."]	["tabId":1921,"windowId":28]	2019-08-28 22:23:10
32441	4	1	1	page	4811	NULL	NULL	2019-08-28 22:23:10
32442	4	1	1	query	547	NULL	NULL	2019-08-28 22:23:10
32443	4	1	1	windows.onFocusChanged	-1	NULL	null	2019-08-28 22:23:13
32444	4	1	1	windows.onFocusChanged	28	["alwaysOnTop":false,"focused":tru...	null	2019-08-28 22:23:31
32445	4	1	1	webNavigation.onCommitted	1921	["active":true,"audible":false,"auto..."]	["frameId":0,"parentFrameId":-1,"p...	2019-08-28 22:23:33
32446	4	1	1	tabs.onUpdated	1921	["active":true,"audible":false,"auto..."]	["status":"complete"]	2019-08-28 22:23:35

A feature extractor script fetches important information about the user's session. Researchers can use it to analyze patterns in information seeking behaviors and to build predictive models via machine learning.

ACKNOWLEDGEMENTS

I would like to thank Dr. Matthew Mitsui and Dr. Chirag Shah for their mentorship and support, as well as the Douglass Project for providing guidance and funding through Project SUPER.