Code You Can Use: Searching for web automation scripts based on reusability

James Admire, Abbas Al Zawwad, Abdulwahab Almorebah, Sanchit Karve, Christopher Scaffidi
School of Electrical Engineering and Computer Science
Oregon State University
Corvallis, OR 97330
admire.james@gmail.com, abbaszawwad@gmail.com, almorebah@gmail.com, write2sanchit@gmail.com, cscaffid@eecs.oregonstate.edu

Abstract—Web scripting enables users to automate interactions with websites. Online open source repositories provide scripts available for reuse. Yet just because these scripts are open source does not mean they are all reusable: many are specialized and irrelevant to most peoples’ needs, while others are hard to understand or learn from. Repositories offer keyword-based search engines to find scripts relevant to specialized needs, but they lack any means for filtering search results according to reusability.

To address this shortcoming, we present an approach for creating a model to automatically estimate the reusability of web automation scripts. To test this approach, we prototyped a search engine that uses these reusability estimates to sort one particular kind of web automation scripts, CoScripter macros, according to reusability. An empirical evaluation confirmed that the system’s reusability estimates are significantly correlated with user perceptions of macro reusability, thus implying that our approach presents a viable means for helping end-user programmers to find reusable web automation scripts.

Keywords—reuse; end-user programming; scripting

I. INTRODUCTION

The traditional view of the web is one where people manually click on links and fill in forms, but recent years have seen a proliferation of tools enabling people to create scripts that automate user interactions with web sites. An example of such web automation scripts is the CoScripter macro, which automates tasks by clicking links or filling forms, thereby enabling people to finish tasks far faster than they could manually [21]. Another example is the “mashup,” a program that extracts, combines and visualizes data from multiple sites, thereby providing a synthesized view to support decision-making. Because of these benefits, there is a strong and increasing demand for web automation scripts, which are starting to be used in climate science [11], education [23], business [15], disaster response [24], health [1], national defense [12], accessible computing [4], and government [13].

Commercial programming tools for creating web automation scripts are mature and plentiful (e.g., [5][8]), with a potential market of approximately $700 million [33]. Some of these tools include online environments to support sharing and reusing of web automation scripts. For example, repositories exist for scripts that extend the FireFox browser [7] and programs for processing news feeds [10]. Up until recently, IBM operated one of the largest open source script repositories, the CoScripter web macro environment [21]. Like repositories that support sharing of other kinds of code, such as scripts that run on mobile phones [30] or programs that show animations [25], these repositories of web automation scripts provide features to search for scripts based on keyword queries.

Even though the users of these environments are end-user programmers—people who primarily create programs for their own use [19]—it is still valuable to share scripts via these repositories because doing so can provide resources for other people to download, adapt, or run without modification. For example, one study found that some IBM employees benefited from reusing one another’s macros for automating recurring tasks such as getting reimbursed for monthly networking bills, for repeating hard-to-remember tasks such as how to order business cards in the procurement web site, and for teaching other people how to do complex tasks such as filling out necessary paperwork [21]. As another example, the Yahoo! Pipes repository enables people to download and reuse scripts to mash up news feeds [29].

Yet just because a web automation program or other script is open source does not mean that it is reusable—in fact, as few as 10% of CoScripter macros may ever be used by other people [28]. The remainder includes macros that are specialized to the needs of specific users and might be of no use to others, and/or which are so hard to understand and customize that it is difficult for others to modify them to meet their own needs. Less reused macros also include those that no longer work properly and that might be a challenge to understand and tweak so they work properly. These less useful macros create a clutter that, in turn, interferes with finding the truly useful macros.

Existing repositories provide little help with overcoming this challenge because they typically only provide keyword-based search engines for finding scripts relevant to a specific task. For example, within a typical repository of several thousand scripts, users might find 10 macros related to filing monthly reimbursements, but they might not be able to differentiate among these to identify the 2 or 3 that could easily be understood and customized.

This work was sponsored in part by the National Science Foundation.