Relationship-Aware Code Search for JavaScript Frameworks

Xuan Li

Institute of Software, School of Electronics Engineering and Computer Science
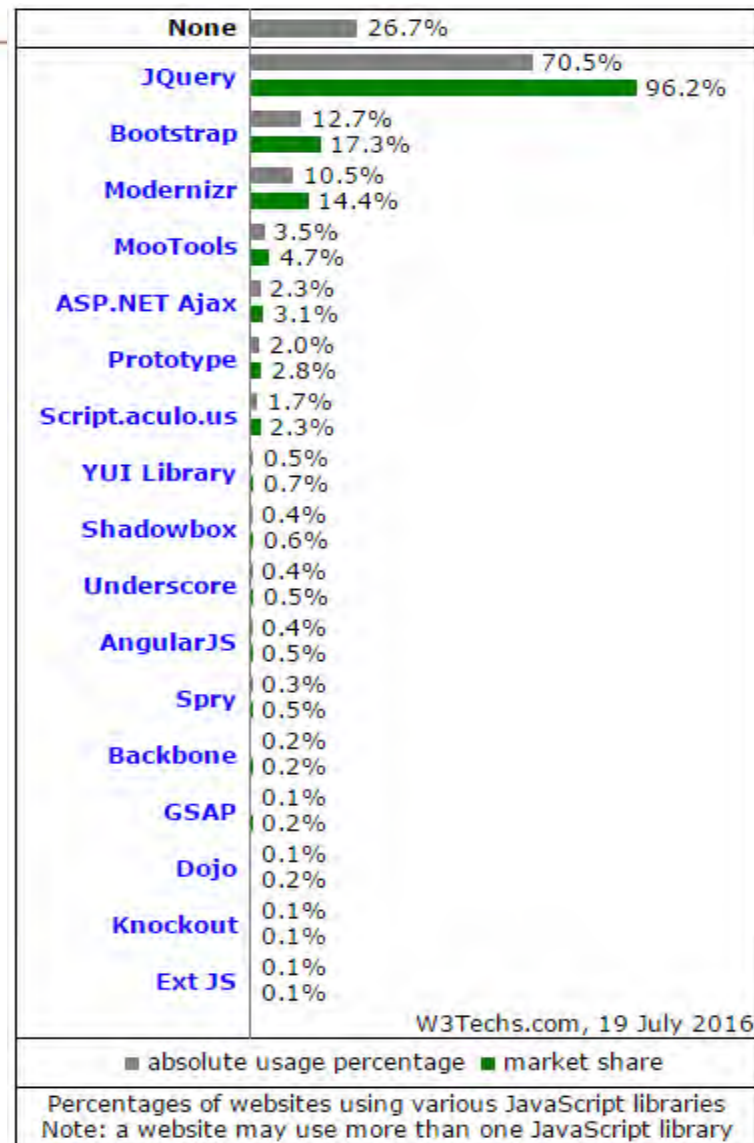
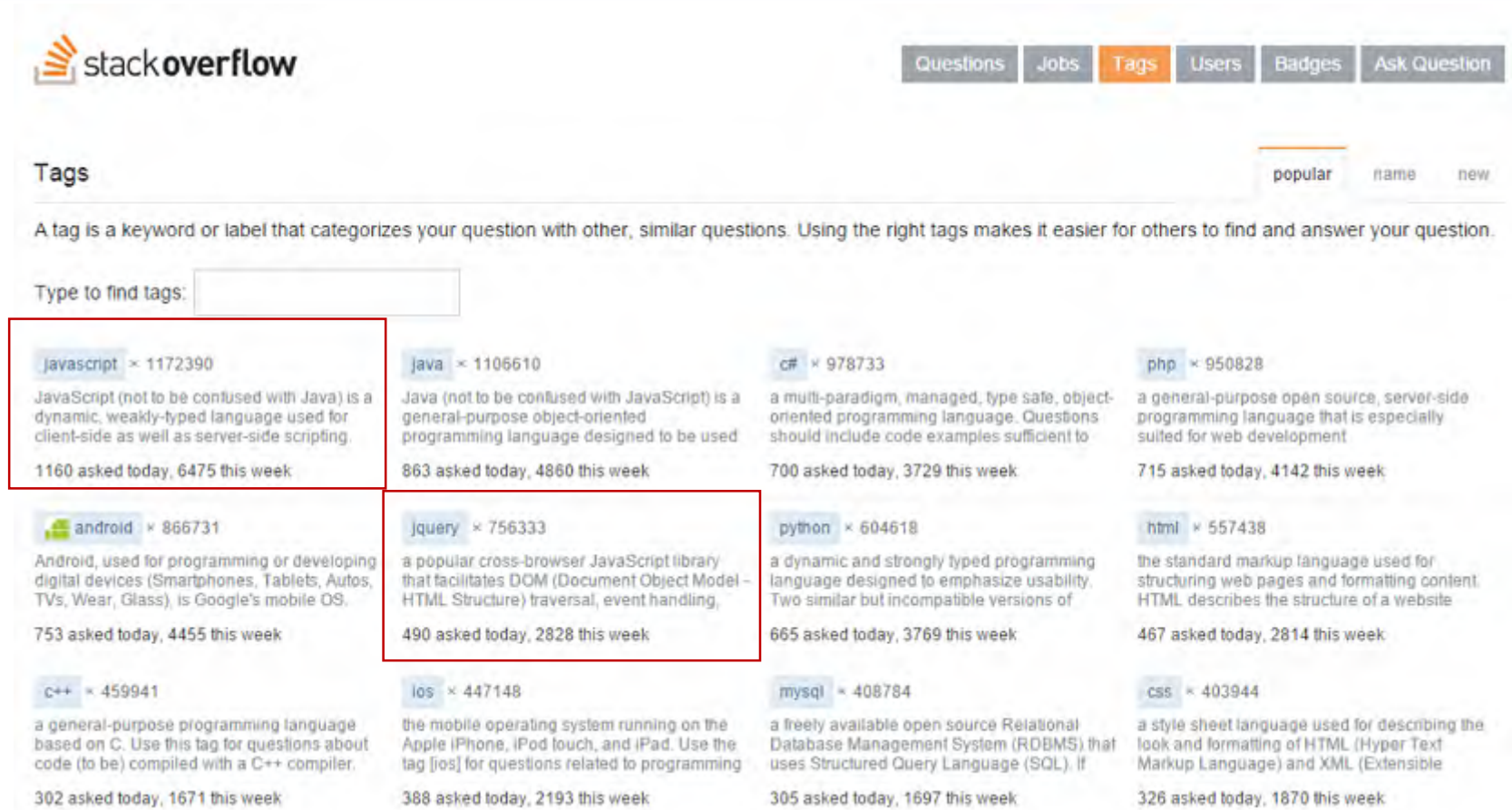Peking University

2016.07.20

Motivation

- JavaScript frameworks are widely used for developing web applications. A recent survey[1] has shown that 73.3% of the top 10 million websites use JavaScript frameworks, such as jQuery, MooTools, Prototype, YUI, and ExtJS. Meanwhile, among all these frameworks, jQuery has a share of 96.2%.

	absolute usage percentage	market share
None	26.7%	
JQuery	70.5%	96.2%
Bootstrap	12.7%	17.3%
Modernizr	10.5%	14.4%
MooTools	3.5%	4.7%
ASP.NET Ajax	2.3%	3.1%
Prototype	2.0%	2.8%
Script.aculo.us	1.7%	2.3%
YUI Library	0.5%	0.7%
Shadowbox	0.4%	0.6%
Underscore	0.4%	0.5%
AngularJS	0.4%	0.5%
Spry	0.3%	0.5%
Backbone	0.2%	0.2%
GSAP	0.1%	0.2%
Dojo	0.1%	0.2%
Knockout	0.1%	0.1%
Ext JS	0.1%	0.1%

W3Techs.com, 19 July 2016

Percentages of websites using various JavaScript libraries
Note: a website may use more than one JavaScript library

北京大学

[1]https://w3techs.com/technologies/overview/JavaScript_library/all

Motivation

When using these JavaScript frameworks, a large number of programmers are often in great need of help. For example, 9.5% of questions in Stack Overflow are tagged with "JavaScript", which is the top 1 tag.

M

- Wh...
  fun...
  imp...

- Ger...
  -
  -

- Source-code-specific search engines (Ohloh, Krugle, Codase)
  - Return code snippets or projects
  - Treat code as text

- Recent research contributes new approaches that leverage code analysis and code mining, e.g., PARSEWeb, MAPO, SNIFF, and SWCE, to take into account code characteristics, such as API usage patterns and encoded code patterns. However, none of these approaches consider characteristics of JavaScript code snippets or search queries related to using a JavaScript framework API.
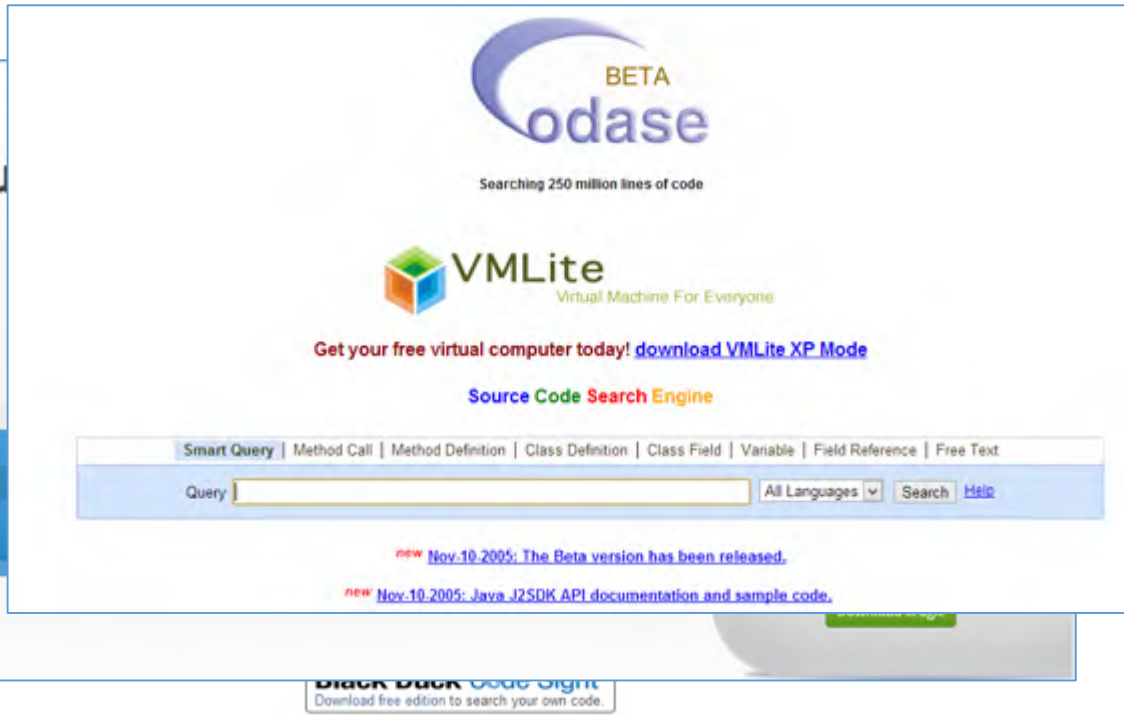
Motivation Example

Stack Overflow Question and Description

How to display loading image while actual image is downloading

Some time images take some time to render in the browser. I want show a busy image while the actual image is downloading, and when image is downloaded, the busy image is removed and actual image is be shown there. How can I do this with JQuery or any javascript?

Accepted Answer

You can do something like this:

```
1| // show loading image
2| $('#loader_img').show();
3| // main image loaded ?
4| $('#main_img').load(function(){
5| // hide/remove the loading image
6| $('#loader_img').hide();
7| });
```

You assign `load` event to the image which fires when image has finished loading. Before that, you can show your loader image.

From http://stackoverflow.com/questions/4635388

2)JavaScript is mainly used for client-side scripting in web browsers. Many typical search queries for JavaScript framework API usage describe the user interaction, browser control, asynchronous communication, and the altering of a displayed document content.

1)In JavaScript, relationships between method calls are complex, beyond sequencing relationships among method calls as commonly captured by existing approaches.

Existing ways of code searching face various respective limitations either in general or in particular for JavaScript.

北京大学

Approach Overview

(1)JavaScript code snippet → method call relationship (MCR) graphs

(2)natural-language (NL) search query → action relationship (AR) graph

(3)RACS reduces code search to the problem of graph search: finding the similar MCR graphs for a given AR graph.

Overview of RACS
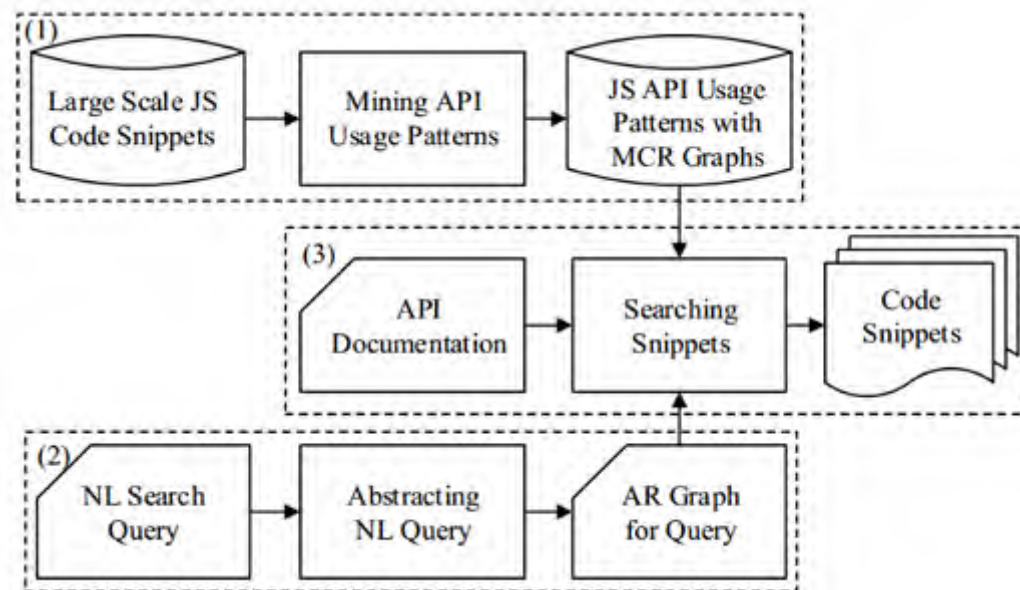
Abstracting Code Snippets

```
$('#loader_img').show();
$('#main_img').load(function(){
          $('#loader_img').hide();
});
if($(window).width() < 960 ){
          $('#warning_img').show();
}
```


relatively conservative clustering: if two code snippets are abstracted to the same MCR graph, they fall into one cluster represented by the MCR graph.

Abstracting NL Query

- Preprocessing

- Action Identification
  - use the Stanford Parser to generate the POS tags and parse tree
  - an action consists of Verb Phrase (VP), Noun Phrase (NP), and optional Prepositional Phrase (PP).

- Relationship Identification
  - **Sequencing relationship:** preposition ("before" and "after"), aconjunction ("and" and "then"), or a punctuation (comma, semicolon, and period)
  - **Condition relationship:** "if"
  - **Callback relationship:** (1) two actions' descriptions are connected by a connecting word being a conjunction or a preposition indicating point-in-time ("when", "after", and "if"), and also (2) the action happening first, should imply an event or describe a completion status;

- Post-processing
  - replaces some specific nouns (e.g., "image", "div" indicating element types) with more general ones (e.g., "element") based on the word list for DOM elements in the domain-specific dictionary

Show a busy image while the actual image is downloading, and when image is downloaded, the busy image is removed and actual image is be shown there

Snippet Searching

- Deriving A-MCR graphs from the AR graph
  - MCS measure of text semantic similarity

$$sim(T_1, T_2) = \frac{1}{2}\left(\frac{\sum_{w \in \{T_1\}}\left(maxSim(w, T_2) * idf(w)\right)}{\sum_{w \in \{T_1\}} idf(w)} + \frac{\sum_{w \in \{T_2\}}\left(maxSim(w, T_1) * idf(w)\right)}{\sum_{w \in \{T_2\}} idf(w)}\right)$$

- Searching MCR graphs similar to A-MCR graphs
  - Similarity of MCR graphs

$$sim(G_1, G_2) = \frac{mrn(G_1, G_2)}{rn(G_1) + rn(G_2) - mrn(G_1, G_2)}$$

- Selecting Code Snippets
  - Two heuristics: matched keywords and code lengths

Evaluation

- Research questions
  - RQ1: How effectively can RACS search JavaScript code snippets for a given NL search query?
  - RQ2: Can RACS outperform a relationship-oblivious code search approach?

- Evaluation Setup
  - JavaScript snippet base

Source	#Project	Date of	#files	#snippets
Ohloh	620	2014.06.28	51,949	226,099
Amazon.com	1	2014.08.16	29,591	82,195
Total	621	--	81,540	308,294

  - Benchmark queries
    - 50 real-world representative jQuery related queries from Stack Overflow
  - The details of our evaluation subjects and results: http://taoxie.cs.illinois.edu/racs/.

- Evaluation Metrics
  - Best hit rank, i.e., the highest rank of the hit snippets for the query
  - Success percentage at k, i.e., the success percentage among the set of queries considering only the top k results returned by a search approach.

北京大学

Effectiveness of RACS

Table 2. Selected Stack Overflow queries, search results of RACS, and characteristic of accepted answers

No.	Question ID	NL Search Query	Query AR Graph		Target Snippet MCR Graph		T1 (ms)	T2 (ms)	Top Rank
			#Meth	#Rela	#Meth	#Rela			
1	1854556	If a field is click into, check if input is empty, display a red background.	3	2	4	3	209	463	1
2	6677035	When the user clicks on that input subject, the page should scroll to the last element of the page with a nice animation to scroll to bottom and not to top.	3	1	3	2	247	471	44
3	554273	When someone clicks on an image, change the image source.	2	1	2	1	188	78	3
4	986120	Get the value of the selected radio button when any of these three are clicked.	2	1	2	1	195	100	1
5	1423561	Hide the container if focus is lost.	2	1	2	1	181	80	1
6	699065	When I press Enter on the form, the form is submitted.	2	1	2	1	194	79	1
7	901712	If the age checkbox is checked, then I need to show a textbox to enter age, else hide the textbox.	3	2	2	1	214	356	1
8	152975	Show HTML menus completely when a user clicks on the head of these menus. Hide these elements when the user clicks outside the menus' area.	4	3	3	2	287	2633	NF
9	169506	When I catch the submit form event with jQuery, get all the input fields of that form in an associative array.	2	1	2	1	211	101	1
10	1594952	When the text field is empty the submit button should be "disabled". When something is typed in the text field to remove the "disabled" attribute. If the text field becomes empty again the submit button should be "disabled" again.	6	5	4	3	302	70155	NF
11	24816	Escaped an arbitrary string and display in an HTML page.	2	1	2	1	191	90	NF
12	303767	Grab the height of the window and the scrolling offset in jQuery.	2	1	2	1	200	88	NF
13	1216114	Make a div stick to the top of the screen once it's been scrolled enough to contact its top boundary	2	0	2	1	236	90	NF
14	253689	Change the background image of a div when it is clicked.	2	1	2	1	193	88	1
15	480735	Select all contents of textbox when it receives focus.	2	1	2	1	189	89	3
16	164085	Execute a callback when an IFRAME has finished loading.	2	1	2	1	196	91	NF
17	376081	Loop though the table, and get the value of the "Customer Id" column for each row.	2	1	2	1	192	83	NF
18	4551175	Before the AJAX request if the previous request is not completed I've to abort that request and make a new request.	2	1	2	1	224	85	NF
19	912711	Load javascript file only if the user clicks on a certain button.	2	1	2	1	202	99	1
20	47824	Remove all the options of a select box, then add one option.	3	2	3	2	288	498	NF

Effectiveness of RACS

No.	Question ID	NL Search Query	Query AR Graph		Target Snippet MCR Graph		T1 (ms)	T2 (ms)	Top Rank
			#Meth	#Rela	#Meth	#Rela			
21	540349	Hide the rollover image when the onmouseout event happen	2	1	2	1	194	90	1
22	3709597	Wait for all Ajax requests to be done before I execute the next	2	1	2	1	167	85	NF
23	34830973	If a field is clicked, display a background image	2	1	2	1	209	463	1
24	3044573	Determine the size of the browser viewport, and to redetect this if the page is resized?	3	2	3	2	245	590	NF
25	8423217	An event to fire client side when a checkbox is checked	2	1	2	1	202	86	NF
26	5797539	When you click inside a textarea, its entire content gets selected	2	1	2	1	231	417	4
27	871063	Check radio option whether no default is set and then set a default.	2	1	2	1	211	376	1
28	4177159	When element clicked, toggle between checked and unchecked.	2	1	3	2	223	98	NF
29	1064089	When someone clicks a link, a word or two to be inserted where the cursor is.	2	1	2	1	320	1687	1
30	437958	When one of these links is clicked, hide the links that are not clicked.	2	1	2	1	230	79	1
31	1212500	Create a CSS class and add it to DOM at runtime with jQuery.	2	1	2	1	184	83	1
32	7717527	JQuery smooth scrolling when clicking an anchor link.	2	1	2	1	179	80	NF
33	9398870	Remove the top and left attribute from the inline style on the div when clicked.	2	1	2	1	199	345	3
34	946534	Insert text into a text area using jquery, upon the click of an anchor tag.	2	1	2	1	189	78	1
35	1925614	Get the value selected from a dropdown menu and change the form action	1	0	2	1	201	79	NF
36	360491	Strip white space when grabbing text with jQuery?	2	1	2	1	178	83	NF
37	2358205	Trigger an event after any other type of iterative callback has completed.	2	1	2	1	234	345	NF
38	4687579	I want just the new "blah" div to fade in after the content gets appended.	2	1	2	1	169	76	1
39	3024391	Get child elements and iterate through each of those elements.	2	1	2	1	197	85	NF
40	2380230	Get the selected option from a dropdown and populate another item with that text.	2	1	2	1	203	98	1
41	316278	Have an element fade in, then in 5000 ms fade back out again	2	1	2	1	187	80	NF
42	2330209	If the "Check Me" checkbox is checked, all the other 3 checkboxes should be enabled.	2	1	2	1	223	485	1
43	4613261	Get the position of layer1 and set the same position to layer2.	2	1	2	1	202	87	1
44	5176803	When the radio button is selected I enable an edit box.	2	1	2	1	198	89	NF
45	4996002	Get the index of the child li relative to it's parent, when clicking that li	2	1	2	1	188	104	1
46	13626517	Disable inputs at first and then enable them when click a link	3	2	3	2	174	84	NF
47	2230704	Get the value of the hidden field when the select is changed.	2	1	2	1	210	93	1
48	6658752	Generate a new tag with class name "test" in h2 by clicking the button	2	1	2	1	186	88	8
49	4076770	When the <select> dropdown is changed, get the value before change.	2	1	2	1	213	80	1
50	1314450	Capture the TAB keypress, cancel the default action.	2	1	2	1	204	479	1

Code Snippet Example

"*When I press Enter on the form, the form is submitted*".

```javascript
$("[id^='relExInput']").keypress(function(A) {
  if (A.which == 13)
  {
    $("#i").val(this.value);
    $("#calculate").submit();
  }
});
```

"*If the age checkbox is checked, then I need to show a textbox
to enter age, else hide the textbox.*"

```javascript
$('.hide-postbox-tog').bind('click.postboxes', function() {
  var box = $(this).val();
  if ($(this).prop('checked'))
  {
    $('#' + box).show();
    if ($.isFunction(postboxes.pbshow))
    self.pbshow(box);
  } else {
    $('#' + box).hide();
    if ($.isFunction(postboxes.pbhide))
    self.pbhide(box);
  }
  self.save_state(page);
  self._mark_area();
})
```

Top 1 ranked code snippet for queries 6 and 7.

Comparison with Relationship-oblivious Approach

- **RACS**: semantic similarity + relationship-aware ranking
  ROCS: keyword matching + relationship-oblivious (i.e., support-based) ranking
  ROCS+: semantic similarity + relationship-oblivious ranking
  RACS−: keyword matching + relationship-aware ranking

Table 3. Comparison results

Metrics	Pattern Searching Method Searching	Relationship-oblivious	Relationship-Aware	Typeless Relationship-Aware
P_1	Keyword Matching	4%	16%	10%
P_5		10%	22%	20%
P_{10}		16%	26%	26%
P_1	Semantic Similarity	14%	46%	36%
P_5		34%	54%	52%
P_{10}		48%	56%	56%

RACS: ▮   ROCS: ▯   ROCS⁺: ▮   RACS⁻: ▯

Discussion

- With some modifications, our RACS approach can be applied to a wider scope. For example, when used for another JavaScript framework, RACS needs to use only the framework's corresponding API documentation. For other programming languages, we can define more relationships that best show these languages' features.

- Our RACS approach attains the NL description for an API method from the API documentation's short description, which may not comprehensively capture the API method's semantics. The user may use a high-level description where one action maps to multiple API methods. Automatic techniques of comment generation may alleviate this problem. We can also attain more knowledge by crowdsourcing beyond API documentation.

Conclusion

- Existing code search approaches are not effective in finding code snippets that use JavaScript frameworks to implement a specific feature reflected by the given natural-language (NL) search query. In this paper, we have presented a novel Relationship-Aware Code Search (RACS) approach.

- RACS first collects a large number of code snippets that use some JavaScript frameworks in advance, mines API usage patterns from the collected code snippets, and represents the mined patterns with MCR graphs. Given an NL search query, RACS conducts NL processing to automatically transform the query to an AR graph. In this way, RACS reduces code search to the problem of graph search: searching the MCR graphs for a graph similar to the given AR graph. During the graph search, RACS includes a technique based on text semantic similarity to bridge the gap between NL actions in an AR graph and framework API methods in an MCR graph.

- We have conducted two evaluations against popular real-world jQuery questions (posted on Stack Overflow), based on 308,294 code snippets collected from over 81,540 files on the Internet. The evaluation results show the effectiveness of RACS: the top 1 snippet produced by RACS matches the target code snippet for 46% questions, compared to only 4% achieved by a relationship-oblivious approach (which shares key ideas with existing state-of-the-art code search approaches).

Q&A