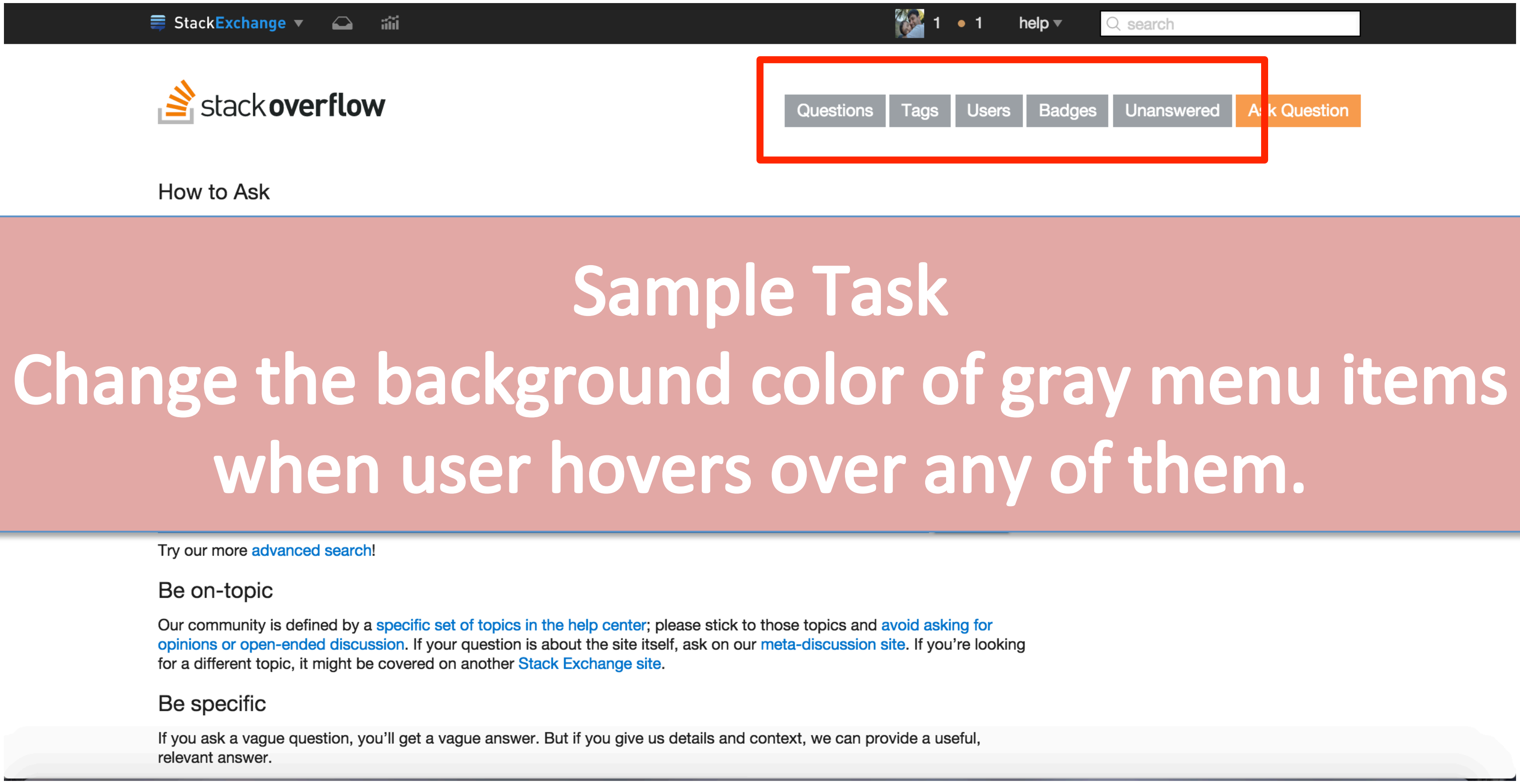
Synthesizing Web Element Locators

Kartik Bajaj, Karthik Pattabiraman, Ali Mesbah
{kbajaj, karthikp, amesbah}@ece.ubc.ca

Running Example



How to Ask

Try our more advanced search!

Be on-topic

Our community is defined by a specific set of topics in the help center; please stick to those topics and avoid asking for opinions or open-ended discussion. If your question is about the site itself, ask on our meta-discussion site. If you're looking for a different topic, it might be covered on another Stack Exchange site.

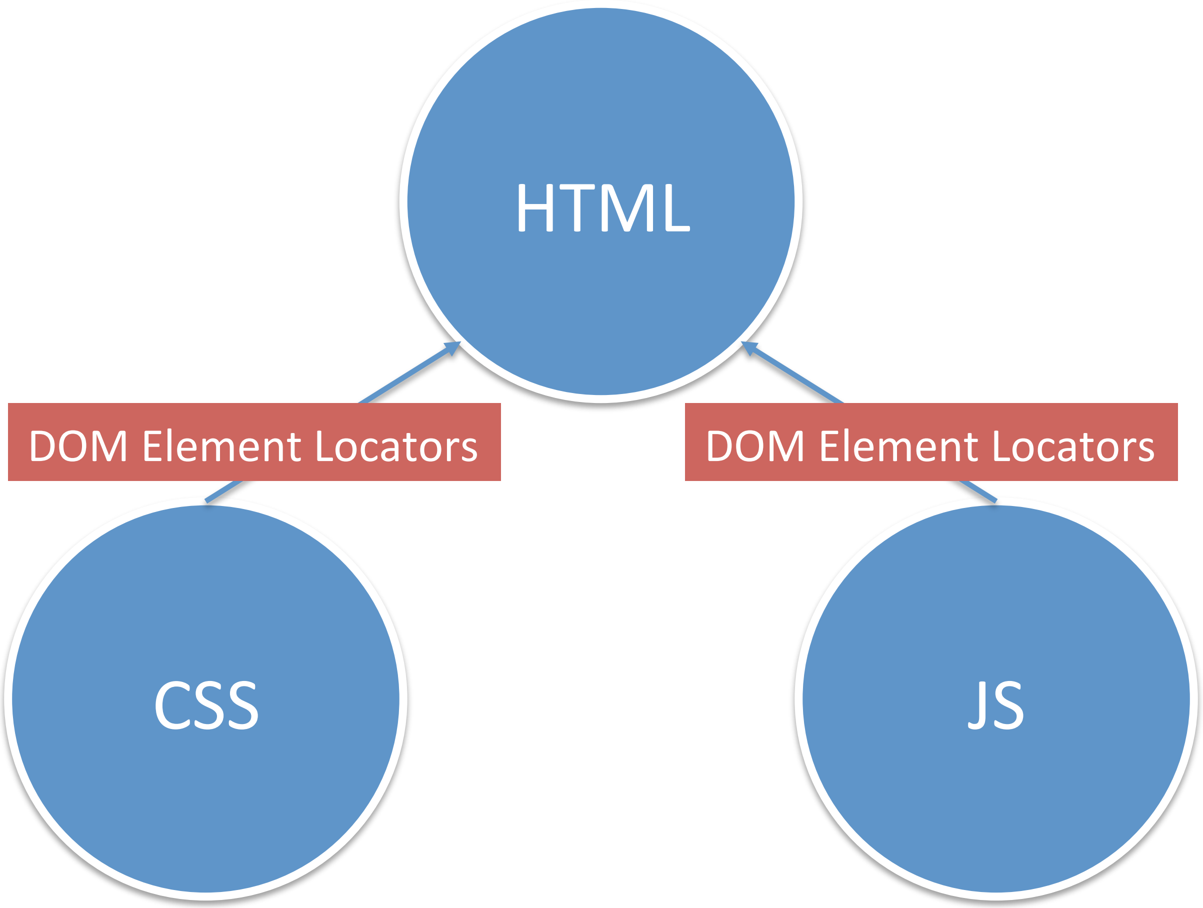
Be specific

If you ask a vague question, you'll get a vague answer. But if you give us details and context, we can provide a useful, relevant answer.

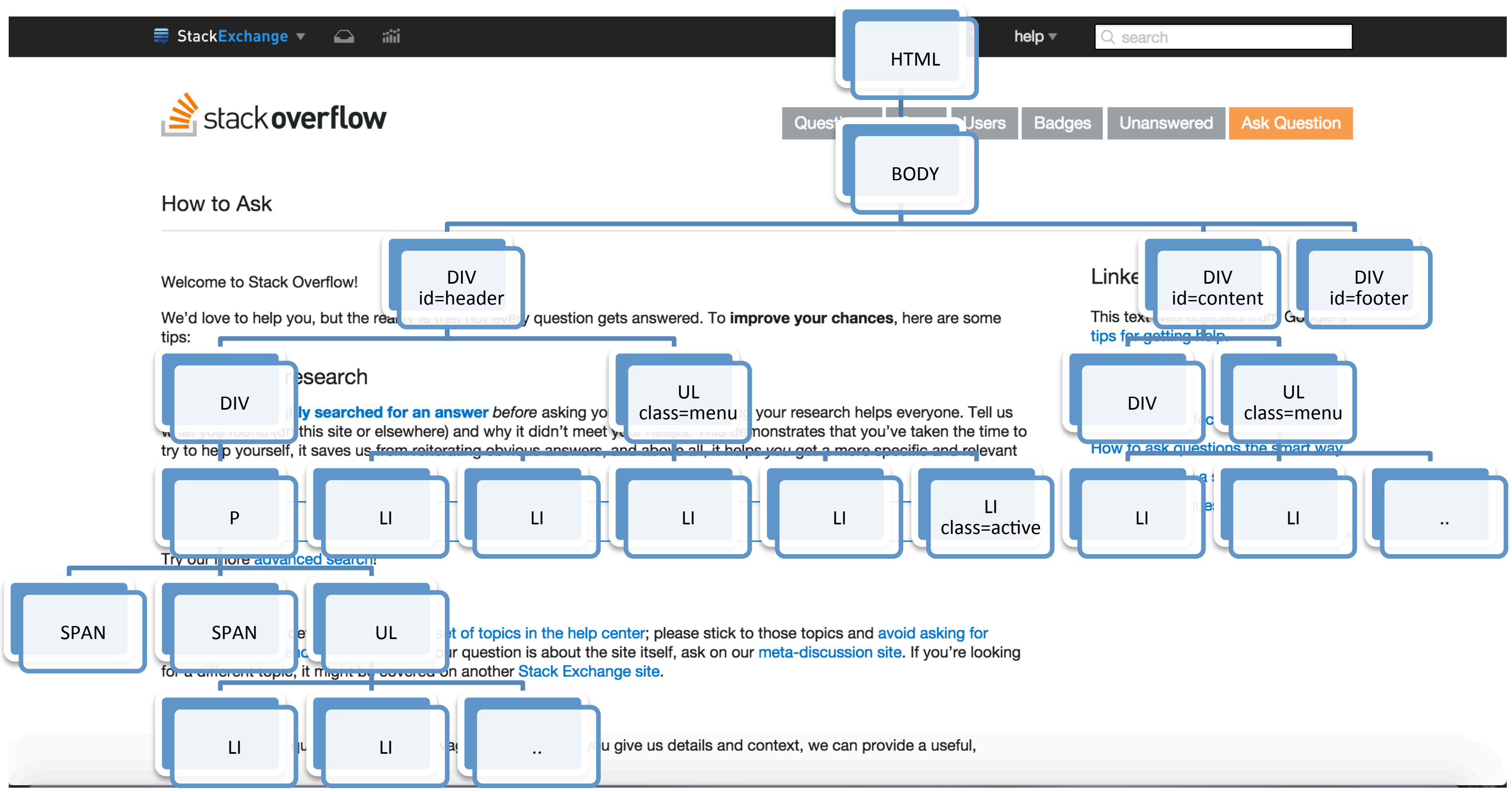
Sample Task

Change the background color of gray menu items when user hovers over any of them.

Web Applications

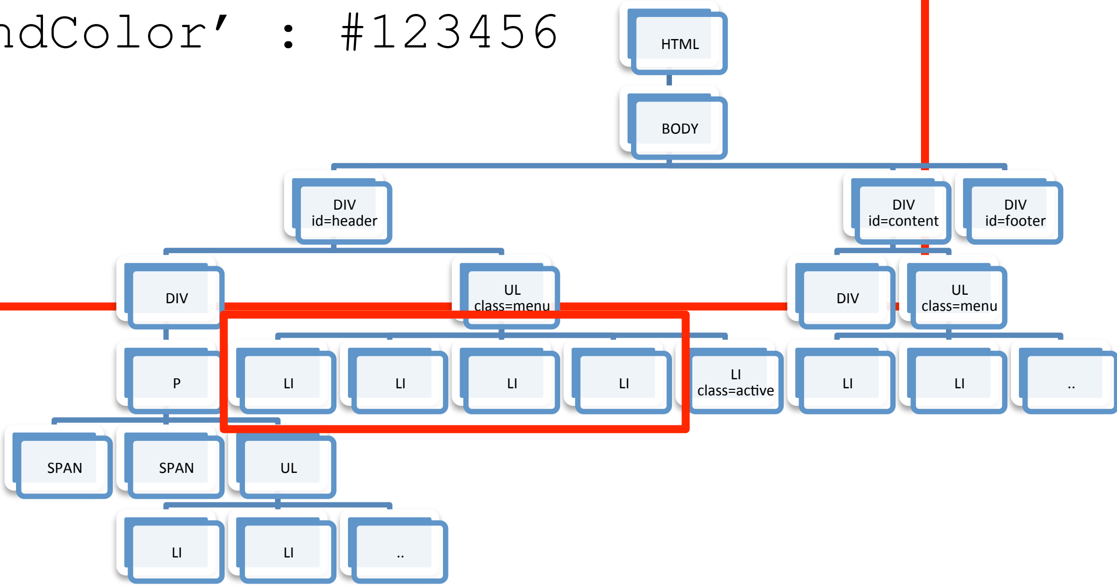


DOM Tree


HTML
BODY
DIV id=header
DIV id=content
DIV id=footer
DIV
UL class=menu
DIV
UL class=menu
P
LI
LI
LI
LI
LI class=active
LI
LI
..
SPAN
SPAN
UL
LI
LI
..
StackExchange
help
search
stackoverflow
Ask Question
Users
Badges
Unanswered
How to Ask
Welcome to Stack Overflow!

JavaScript Code

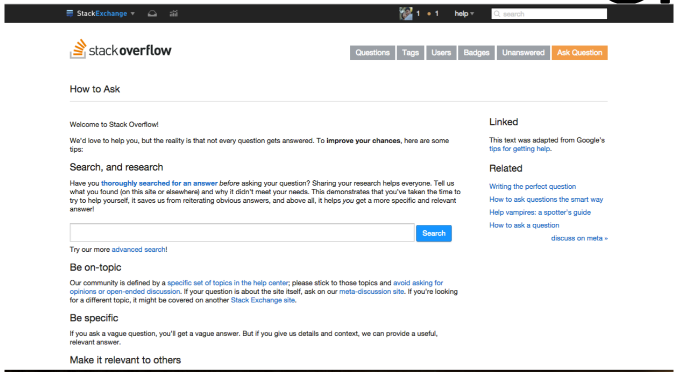
```
var elems = $('#header .menu li:not(active)');

elems.each(function() {
    $(this).hover(function() {
        $(this).css({
            'backgroundColor' : #123456
        });
    });
}
```



Goal

- Assist developers in writing JavaScript code that interacts with the web application.

Challenge – 1


StackExchange 1 • 1 help search

stackoverflow Questions Tags Users Badges Unanswered Ask Question

How to Ask

Welcome to Stack Overflow!

We'd love to help you, but the reality is that not every question gets answered. To **improve your chances**, here are some tips:

Search, and research

Have you **thoroughly searched for an answer** *before* asking your question? Sharing your research helps everyone. Tell us what you found (on this site or elsewhere) and why it didn't meet your needs. This demonstrates that you've taken the time to try to help yourself, it saves us from reiterating obvious answers, and above all, it helps *you* get a more specific and relevant answer!

Search

Try our more advanced search!

Be on-topic

Our community is defined by a specific set of topics in the help center; please stick to those topics and avoid asking for opinions or open-ended discussion. If your question is about the site itself, ask on our meta-discussion site. If you're looking for a different topic, it might be covered on another Stack Exchange site.

Be specific

If you ask a vague question, you'll get a vague answer. But if you give us details and context, we can provide a useful, relevant answer.

Make it relevant to others

Linked

This text was adapted from Google's tips for getting help.

Related

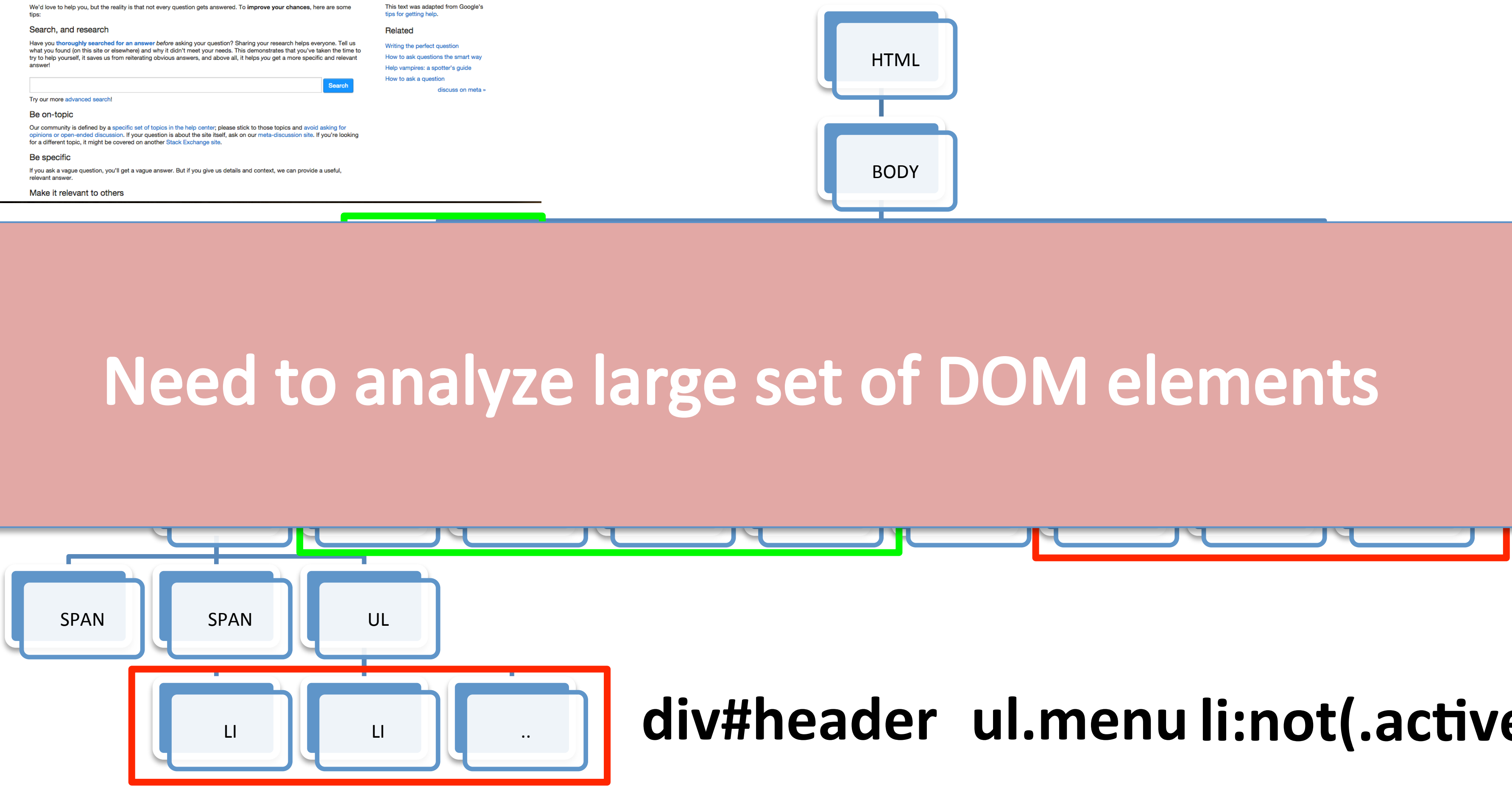
Writing the perfect question

How to ask questions the smart way

Help vampires: a spotter's guide

How to ask a question

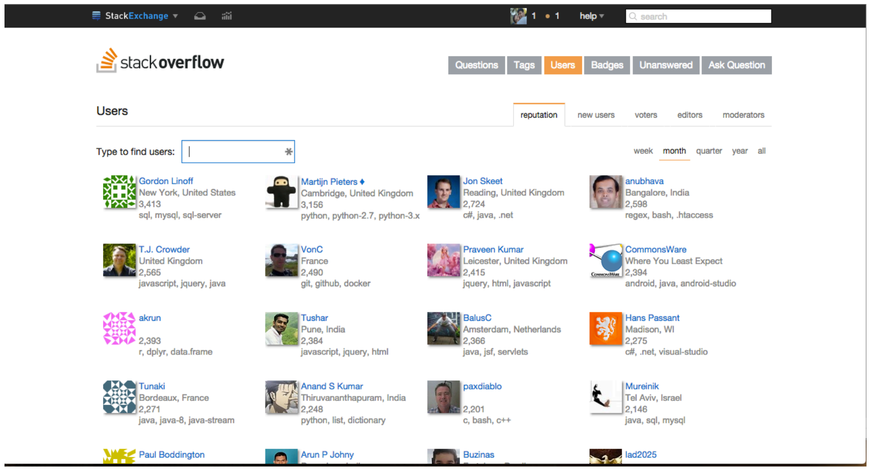
discuss on meta »



Need to analyze large set of DOM elements

div#header ul.menu li:not(.active)

Challenge - 2



Users

Type to find users:

Gordon Linoff
New York, United States
3,413
sql, mysql, sql-server

Martijn Pieters ♦
Cambridge, United Kingdom
3,156
python, python-2.7, python-3.x

Jon Skeet
Reading, United Kingdom
2,724
c#, java, .net

anubhava
Bangalore, India
2,598
regex, bash, .htaccess

T.J. Crowder
United Kingdom
2,565
javascript, jquery, java

VonC
France
2,490
git, github, docker

Praveen Kumar
Leicester, United Kingdom
2,415
jquery, html, javascript

CommonsWare
Where You Least Expect
2,394
android, java, android-studio

akrun
2,393
r, dplyr, data.frame

Tushar
Pune, India
2,384
javascript, jquery, html

BalusC
Amsterdam, Netherlands
2,366
java, jsf, servlets

Hans Passant
Madison, WI
2,275
c#, .net, visual-studio

Tunaki
Bordeaux, France
2,271
java, java-8, java-stream

Anand S Kumar
Thiruvananthapuram, India
2,248
python, list, dictionary

paxdiablo
2,201
c, bash, c++

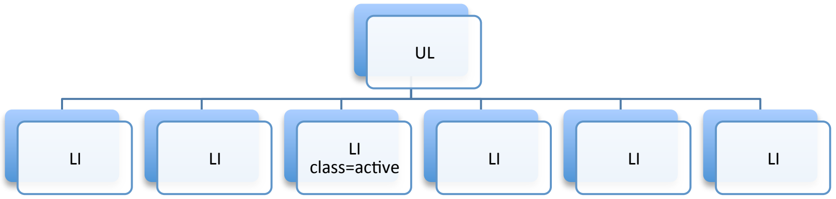
Mureinik
Tel Aviv, Israel
2,146
java, sql, mysql

Paul Boddington

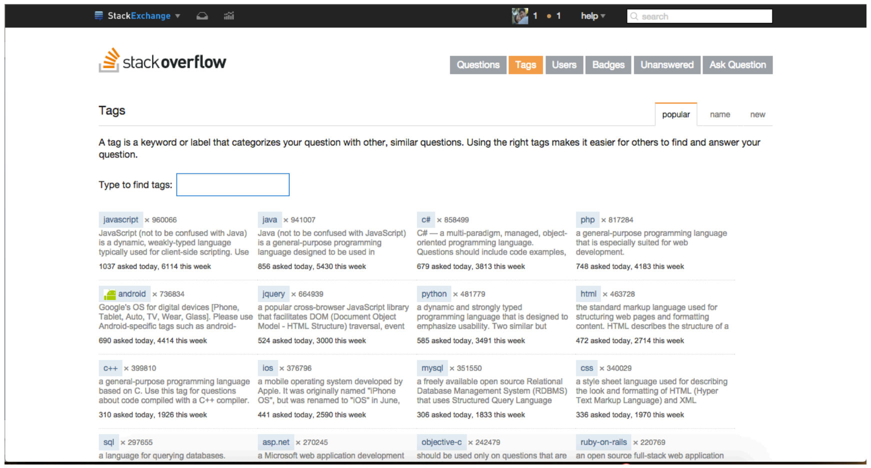
Arun P Johny

Buzinas

lad2025



Challenge - 2



Tags

A tag is a keyword or label that categorizes your question with other, similar questions. Using the right tags makes it easier for others to find and answer your question.

Type to find tags:

javascript × 960066
JavaScript (not to be confused with Java) is a dynamic, weakly-typed language typically used for client-side scripting. Use
1037 asked today, 6114 this week

java × 941007
Java (not to be confused with JavaScript) is a general-purpose programming language designed to be used in
856 asked today, 5430 this week

c# × 858499
C# — a multi-paradigm, managed, object-oriented programming language. Questions should include code examples,
679 asked today, 3813 this week

php × 817284
a general-purpose programming language that is especially suited for web development.
748 asked today, 4183 this week

android × 736834
Google's OS for digital devices [Phone, Tablet, Auto, TV, Wear, Glass]. Please use Android-specific tags such as android-
690 asked today, 4414 this week

jquery × 664939
a popular cross-browser JavaScript library that facilitates DOM (Document Object Model - HTML Structure) traversal, event
524 asked today, 3000 this week

python × 481779
a dynamic and strongly typed programming language that is designed to emphasize usability. Two similar but
585 asked today, 3491 this week

html × 463728
the standard markup language used for structuring web pages and formatting content. HTML describes the structure of a
472 asked today, 2714 this week

c++ × 399810
a general-purpose programming language based on C. Use this tag for questions about code compiled with a C++ compiler.
310 asked today, 1926 this week

ios × 376796
a mobile operating system developed by Apple. It was originally named "iPhone OS", but was renamed to "iOS" in June,
441 asked today, 2590 this week

mysql × 351550
a freely available open source Relational Database Management System (RDBMS) that uses Structured Query Language
306 asked today, 1833 this week

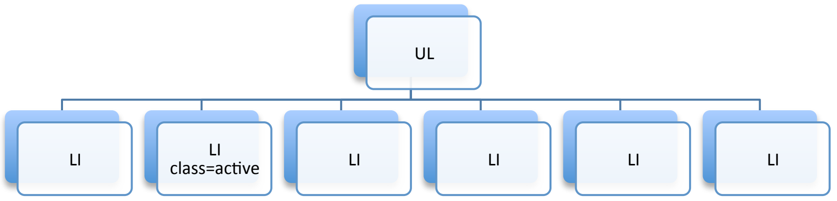
css × 340029
a style sheet language used for describing the look and formatting of HTML (Hyper Text Markup Language) and XML
336 asked today, 1970 this week

sql × 297655
a language for querying databases.

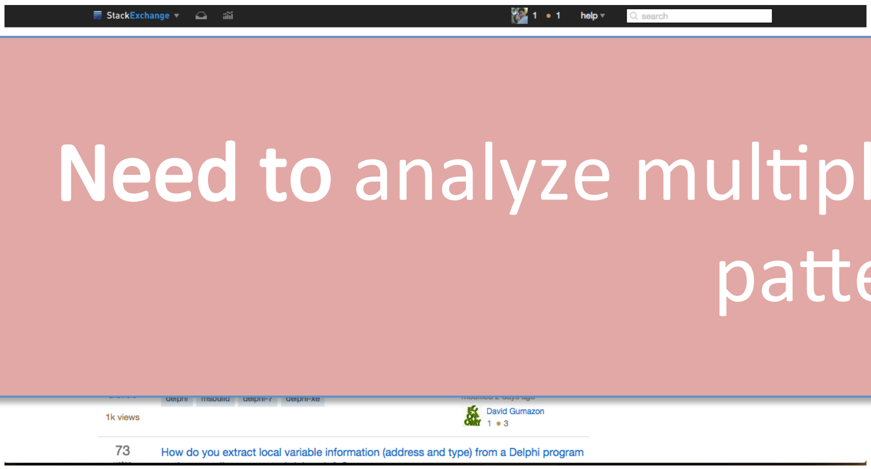
asp.net × 270245
a Microsoft web application development

objective-c × 242479
should be used only on questions that are

ruby-on-rails × 220769
an open source full-stack web application



Challenge - 2



1k views

David Gumazon

1 • 3

73 How do you extract local variable information (address and type) from a Delphi program

Need to analyze multiple DOM states to detect patterns.

Problem Statement

- Writing robust DOM elements locators is a challenging and time consuming task.
  - Developers need to detect patterns in order to avoid hard-coding the selectors.

- Significant amount of JavaScript errors are caused by DOM-JS interactions [ESEM'13]
  - 65% of JavaScript faults are DOM related

Prior Work

JavaScript Code

Test Code

None of the prior work addressed the problem of synthesizing DOM element locators for multiple DOM elements.

CSS Code

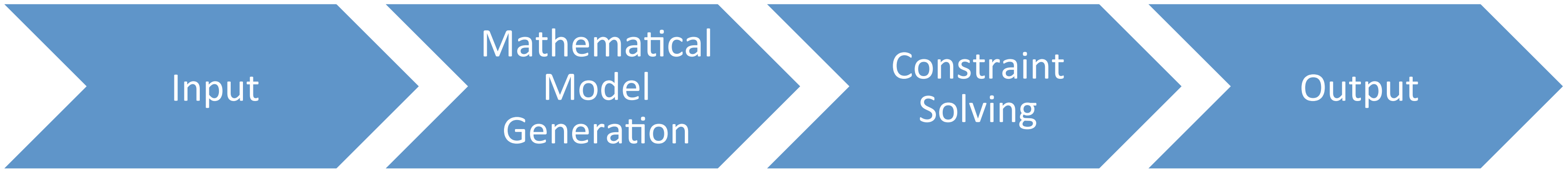
Duplicate detection
Code Maintenance
[Mesbah, ICSE'12]
[Mazinanian, FSE' 14]

Proposed Solution

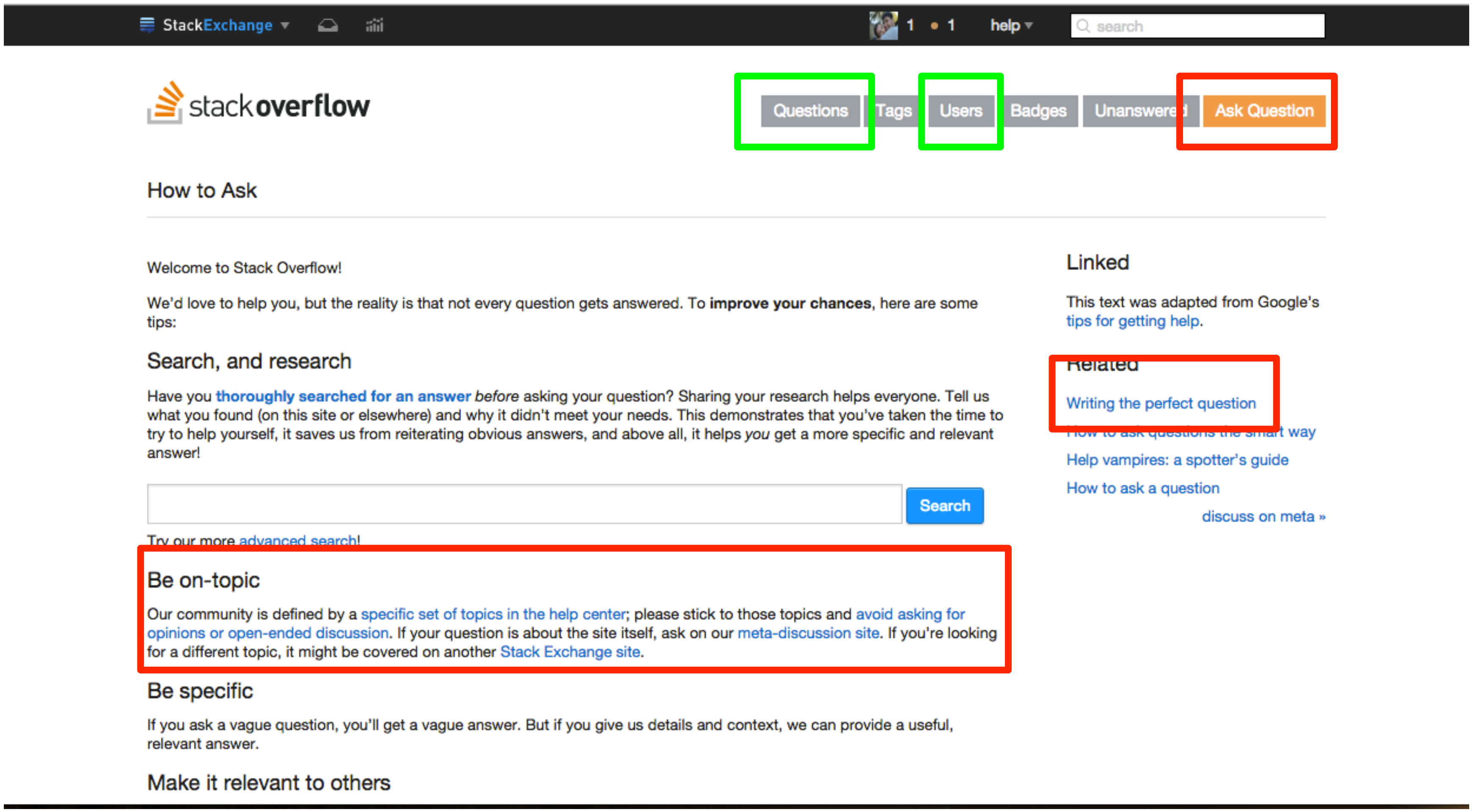
- Utilize Program Synthesis techniques to synthesize DOM Element locators
  - Positive and Negative input examples
  - Generate constraints
  - Use existing SAT solvers to solve these constraints

- Prior work
  - String Manipulation [Gulwani et al.]
  - Data Manipulation [J. Landauer et al.]

Approach Overview



Phase 1 – Input DOM Elements



How to Ask

Welcome to Stack Overflow!

We'd love to help you, but the reality is that not every question gets answered. To **improve your chances**, here are some tips:

Search, and research

Have you **thoroughly searched for an answer** *before* asking your question? Sharing your research helps everyone. Tell us what you found (on this site or elsewhere) and why it didn't meet your needs. This demonstrates that you've taken the time to try to help yourself, it saves us from reiterating obvious answers, and above all, it helps *you* get a more specific and relevant answer!

Search

Try our more advanced search!

Be on-topic

Our community is defined by a specific set of topics in the help center; please stick to those topics and avoid asking for opinions or open-ended discussion. If your question is about the site itself, ask on our meta-discussion site. If you're looking for a different topic, it might be covered on another Stack Exchange site.

Be specific

If you ask a vague question, you'll get a vague answer. But if you give us details and context, we can provide a useful, relevant answer.

Make it relevant to others

Linked

This text was adapted from Google's tips for getting help.

Related

Writing the perfect question

How to ask questions the smart way

Help vampires: a spotter's guide

How to ask a question

discuss on meta »

Input | Mathematical Model Generation | Constraint Solving | Output

Phase 1 – Input Constraints

Constraints	Value
Max Length	3
Ignore	body
…	…

Phase 2 – Mathematical Model Generation

Positive Elements	Negative Elements
`body #header .menu ul li`	`li.active`
`li`	`ul li.active`
`ul li`	`ul li`
`.menu li`	`li`
`#header li`	`#header li`
`body .menu li`	`#content ul li`
`body div li`	`#content li`
`#header ul li`	`#content .menu li`
`#header .menu li`	`.menu li`
`#header .menu li:not(active)`	`body .menu li`
	`ul.menu li`
	`div li`

Phase 2 – Mathematical Model Generation

Positive Elements	Negative Elements
`body #header .menu ul li`	`li.active`
`li`	`ul li.active`
`ul li`	`ul li`
`.menu li`	`li`
`#header li`	`#header li`
`body .menu li`	`#content ul li`
`body div li`	`#content li`
`#header ul li`	`#content .menu li`
`#header .menu li`	`.menu li`
`#header .menu li:not(active)`	**`body .menu li`**
	`ul.menu li`
	`div li`

Constraints	Value
Max Length	3
Ignore	body

Phase 2 – Mathematical Model Generation

Positive Elements	Negative Elements
`li`	`li.active`
`ul li`	`ul li.active`
`.menu li`	`ul li`
`#header li`	`li`
`#header ul li`	`#header li`
`#header .menu li`	`#content ul li`
`#header .menu li:not(active)`	`#content li`
	`#content .menu li`
	`.menu li`
	`ul.menu li`
	`div li`

Phase 2 – Mathematical Model Generation

Positive Elements (DNF)	Negative Elements (DNF)
`(li) ∨ (ul li) ∨ (.menu li) ∨ (#header li) ∨ (#header ul li) ∨ (#header .menu li) ∨ (#header .menu li:not(active))`	`(li.active) ∨ (ul li.active) ∨ (ul li) ∨ (li) ∨ (#header li) ∨ (#content ul li) ∨ (#content li) ∨ (#content .menu li) ∨ (.menu li) ∨ (ul.menu li) ∨ (div li)`

Phase 2 – Mathematical Model Generation

Positive Elements (DNF)	Negative Elements (DNF)
`(li) V (ul li) V (.menu li) V (#header li) V (#header ul li) V (#header .menu li) V (#header .menu li:not(active))`	`~(li.active) ^ ~(ul li.active) ^ ~(ul li) ^ ~(li) ^ ~(#header li) ^ ~(#content ul li) ^ ~(#content li) ^ ~(#content .menu li) ^ ~(.menu li) ^ ~(ul.menu li) ^ ~(div li)`

Phase 2 – Mathematical Model Generation

Conjunctive Normal Form

```
((li) V (ul li) V (.menu li) V (#header li)
  V (#header ul li) V (#header .menu li) V
      (#header .menu li:not(active)) ^
~(li.active) ^ ~(ul li.active) ^ ~(ul li) ^
~(li) ^ ~(#header li) ^ ~(#content ul li) ^
  ~(#content li) ^ ~(#content .menu li) ^
  ~(.menu li) ^ ~(ul.menu li) ^ ~(div li)
```

Phase 3 – Constraint Solving

- `div#header ul.menu li:not(active)`
- `#header .menu li:not(active)`
- `div#header .menu li:not(active)`

Phase 4 - Output

- Rank the synthesized selectors using the following criteria:
  - Universality [QUATIC’10]
  - Abstractness [QUATIC’10]
  - Input Constraints

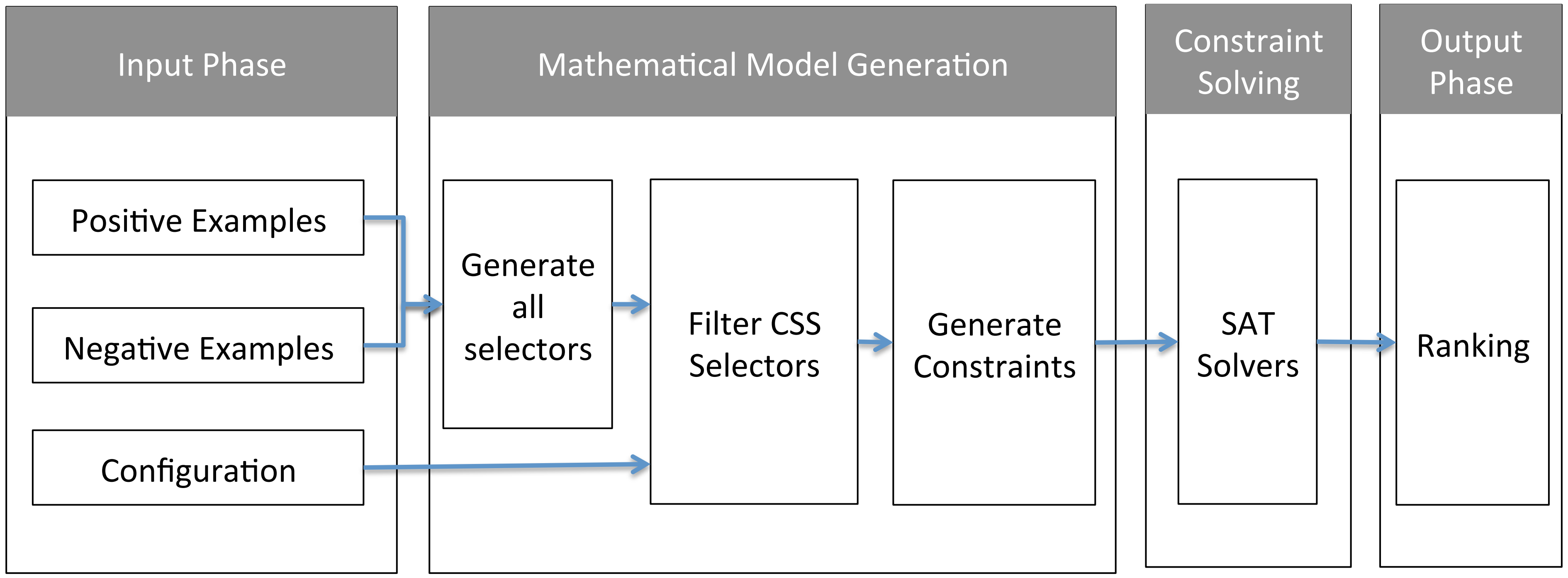
```
div ul.menu li:not(active)
```

```
              Universality = 1 / 3 = 0.33
Abstractness = totalDOMElements * 3 / numElementsInSubtress
```

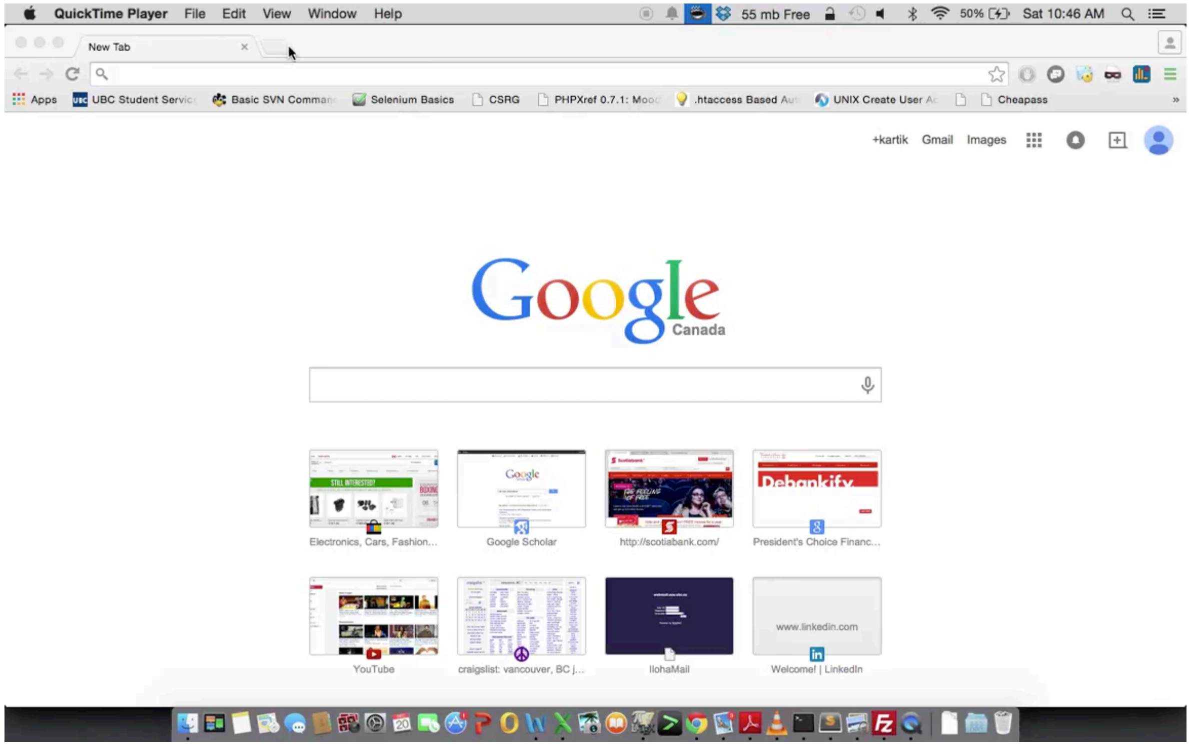
Phase 4 - Output

- `#header .menu li:not(active)`
- `div#header .menu li:not(active)`
- `div#header ul.menu li:not(active)`

Approach Summary



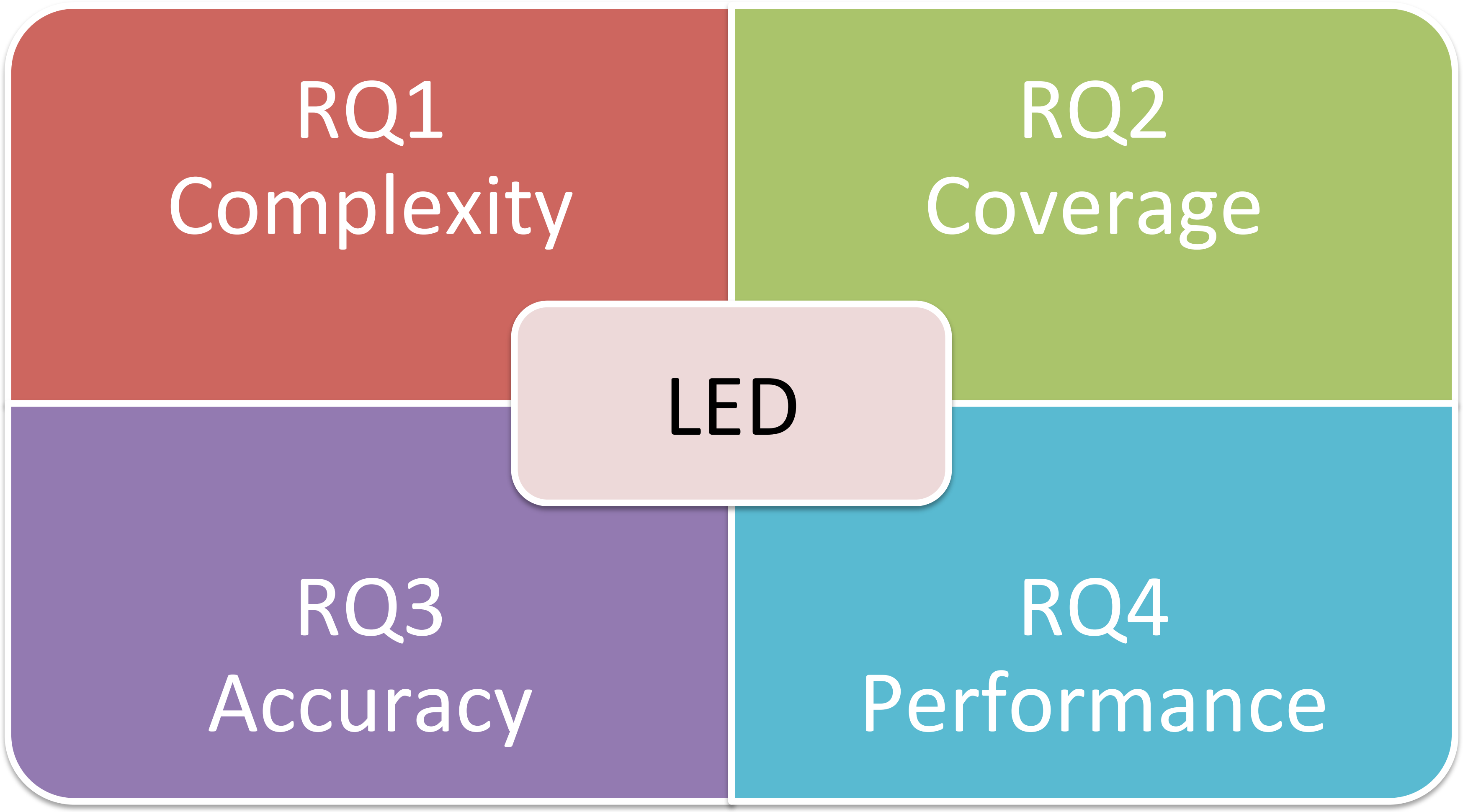
LED: Live Editor for DOM



https://github.com/saltlab/led

Tool Demo: Friday 8:30 AM

Evaluation


RQ1
Complexity
RQ2
Coverage
LED
RQ3
Accuracy
RQ4
Performance

RQ1 - Complexity

- Intercepted DOM API calls within JavaScript code
- Analyzed the DOM element locators used by developers

RQ1 - Results

No. of selected Elements	Percentage	Length of DOM element locator	Percentage
1	78.17%	1	65.85%
2-5	11.97%	2	21.83%
6-10	1.41%	3	2.46%
11 - 100	8.10%	4	9.51%
> 100	0.35%	>5	0.35%

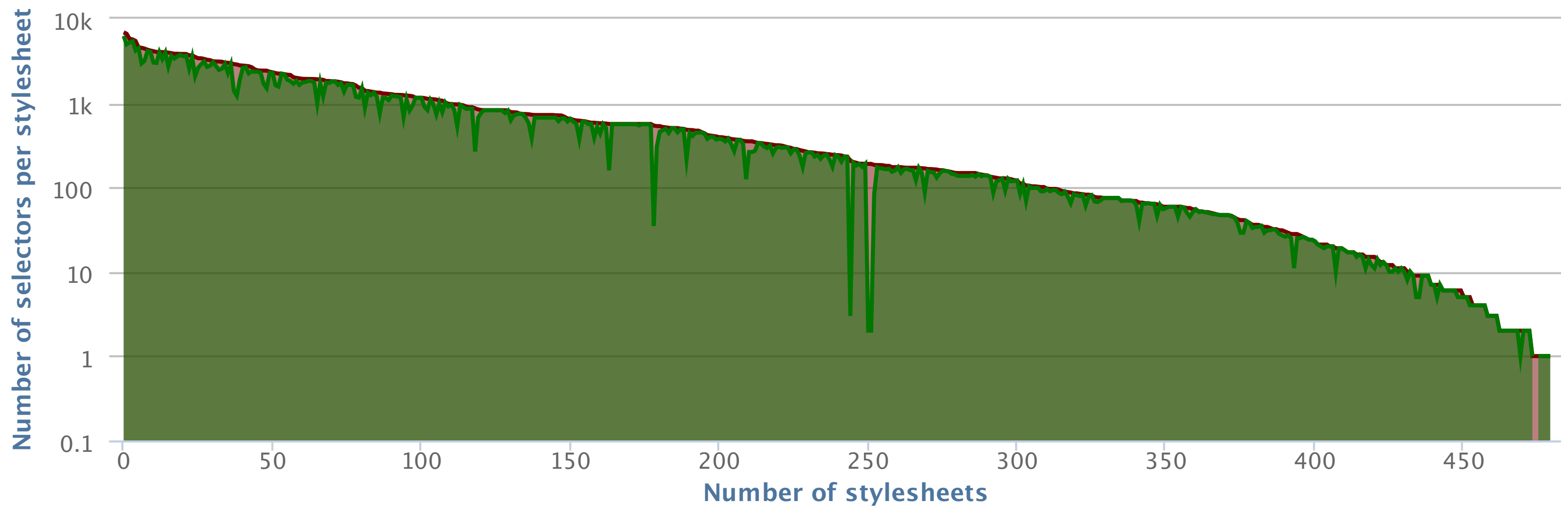
22% select multiple DOM elements

35% are a combination of multiple DOM element locators

RQ2 - Coverage

- Crawled Alexa’s top 200 websites
- Analyzed DOM element locators used in Stylesheets

RQ2 - Results



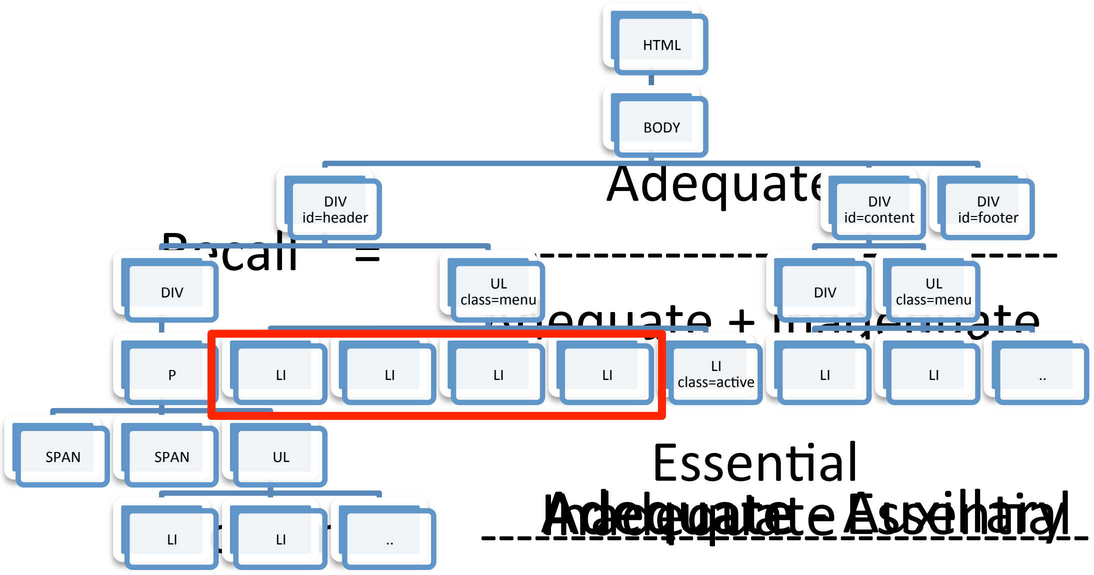
86% supported DOM element locators

RQ3 - Accuracy

- Intercepted DOM API calls within JavaScript code
- Used the selected DOM elements as positive examples
- Added random negative examples
- Synthesized the CSS selectors
- Compared to the one used by the developer.

RQ3 - Accuracy

Category	Type
Inadequate	
Adequate	Essential
	Auxillary



Recall = Adequate / Adequate + Inadequate

Essential
Adequate Essential
Inadequate Auxillary
Essential + Auxillary

RQ3 - Results

Trial	No. of +ve examples	No. of –ve examples	Recall	Precision
0	<= 5	0	98.21%	48.03%
1	> 5	0	100%	47.85%
2	<= 5	5	98.05%	91.84%
3	> 5	5	100%	92.05%

Recall = 98%
Precision = 92%

RQ4: Performance

Search Scope	Average time per application (seconds)			
	Phormer	Gallery3	Wordpress	Average
Limited	0.05	0.08	0.46	0.20
Local	0.06	0.10	0.48	0.21
Global	0.07	0.11	0.49	0.22

Average time = 0.2 seconds
Max avg. time per application = 0.49 seconds

Contributions

1. Discussed the challenges behind DOM element locator synthesis
2. Utilized program synthesis techniques to synthesize DOM element locators for multiple DOM elements
3. Implementation in an open source tool called LED
4. Empirical evaluation to assess LED

https://github.com/saltlab/led

98% Recall and 92% Precision

Max time of 0.49 seconds