

Two studies of opportunistic programming: interleaving web foraging, learning, and writing code

Joel Brandt^{1,2}, Philip J. Guo¹, Joel Lewenstein¹, Mira Dontcheva², Scott R. Klemmer¹

¹Stanford University HCI Group

Computer Science Department

Stanford, CA 94305

{jbrandt, pg, jlewenstein, srk}@cs.stanford.edu

²Advanced Technology Labs

Adobe Systems

San Francisco, CA 94103

mirad@adobe.com

Presented by David Gonzalez

Key Insights:

Study 1:

- They found that developers spent 19% in the Web.
- Three web session intentions were identified: learning, clarification and reminder.

Study 2:

- Web query types such use of natural language, code or both are relevant indicators for determining the type of web session intention* developers were performing.
- Web query features predicts web page types.

*clarification not included

Problem

- Developers rely on online resources, how often? Why?
- Is it systematic or opportunistic?
- It looks opportunistic, so how are they foraging for information?
- Now that we know what they are looking for and how, can it be generalized?

Study 1: Method and Results

- Think-aloud sessions with 20 Students to understand how programmers leverage online resources.
- Web session intentions can be differentiate it:

WEB SESSION INTENTION:	LEARNING	CLARIFICATION	REMINDER
Reason for using Web	Just-in-time learning of unfamiliar concepts	Connect high-level knowledge to implementation details	Substitute for memorization (e.g., language syntax or function usage lookup)
Web session length Starts with web search?	Tens of minutes Almost always	~ 1 minute Often	< 1 minute Sometimes
Search terms	Natural language related to high-level task "ajax tutorial"	Mix of natural language and code, cross-language analogies "javascript timer"	Mostly code (e.g., function names, language keywords) "mysql.fetch.array"
Example search			
Num. result clicks	Usually several	Fewer	Usually zero or one
Num. query refinements	Usually several	Fewer	Usually zero
Types of webpages visited	Tutorials, how-to articles	API documentation, blog posts, articles	API documentation, result snippets on search page
Amount of code copied from Web	Dozens of lines (e.g., from tutorial snippets)	Several lines	Varies
Immediately test copied code?	Yes	Not usually, often trust snippets	Varies

Table 2. Summary of characteristics of three points on the spectrum of Web use intention.

Study 2: Method and Results

- They performed a web search log analysis on 100,000 queries collected from the Adobe Flex' application.

H1: The first query was exclusively natural language in half of *learning* sessions, versus one third in *reminding* sessions (see Table 3).

H2: *Learning* and *reminding* sessions do not have a significant difference in the proportion of queries with refinements before first viewing results.

H3: Programmers were more likely to visit official API documentation in *reminding* sessions than in *learning* sessions (31% versus 10%, see Table 4). Notably, in *reminding* sessions, 42% of results viewed were Adobe tutorials.

H4: Code-only queries accounted for 51% of all *reminding* queries. Among all (including those not hand-coded) sessions, those beginning with code-only queries were refined less ($\mu = 0.34$) than those starting with natural language and code ($\mu = 0.60$) and natural language only ($\mu = 0.51$). It appears that when programmers perform code-only queries, they know what they are looking for, and typically find it on the first search.

Type of first query	Session type		All hand-coded
	learning	reminding	
code only	0.21	0.56	0.48
nat. lang. & code	0.29	0.10	0.14
nat. lang. only	0.50*	0.34	0.38
Total	1.00	1.00	1.00

Table 3. For hand-coded sessions of each type, proportion of first queries of each type (252 total sessions). Significant majorities across each row in bold, * entry means only significant at $p < 0.05$.

Result click Web page type	query type			All clicks
	code	nat. lang. & code	nat. lang.	
Adobe APIs	0.38	0.16	0.10	0.23
Adobe tutorials	0.31	0.33	0.39	0.34
tutorials/articles	0.15	0.22	0.19	0.18
forums	0.03	0.07	0.06	0.05
unclassified	0.13	0.22	0.27	0.20
Total	1.00	1.00	1.00	1.00

Table 6. For queries of each type, proportion of result clicks leading programmer to Web pages of each type (107,343 total queries). Significant majorities and near-ties across each row in bold.

Questions?

Thanks!

What about the observations motivating Searcher?

API Documentation searchers are mostly reminding sessions? Cool!

What else can be said if we study clarifying sessions in depth?