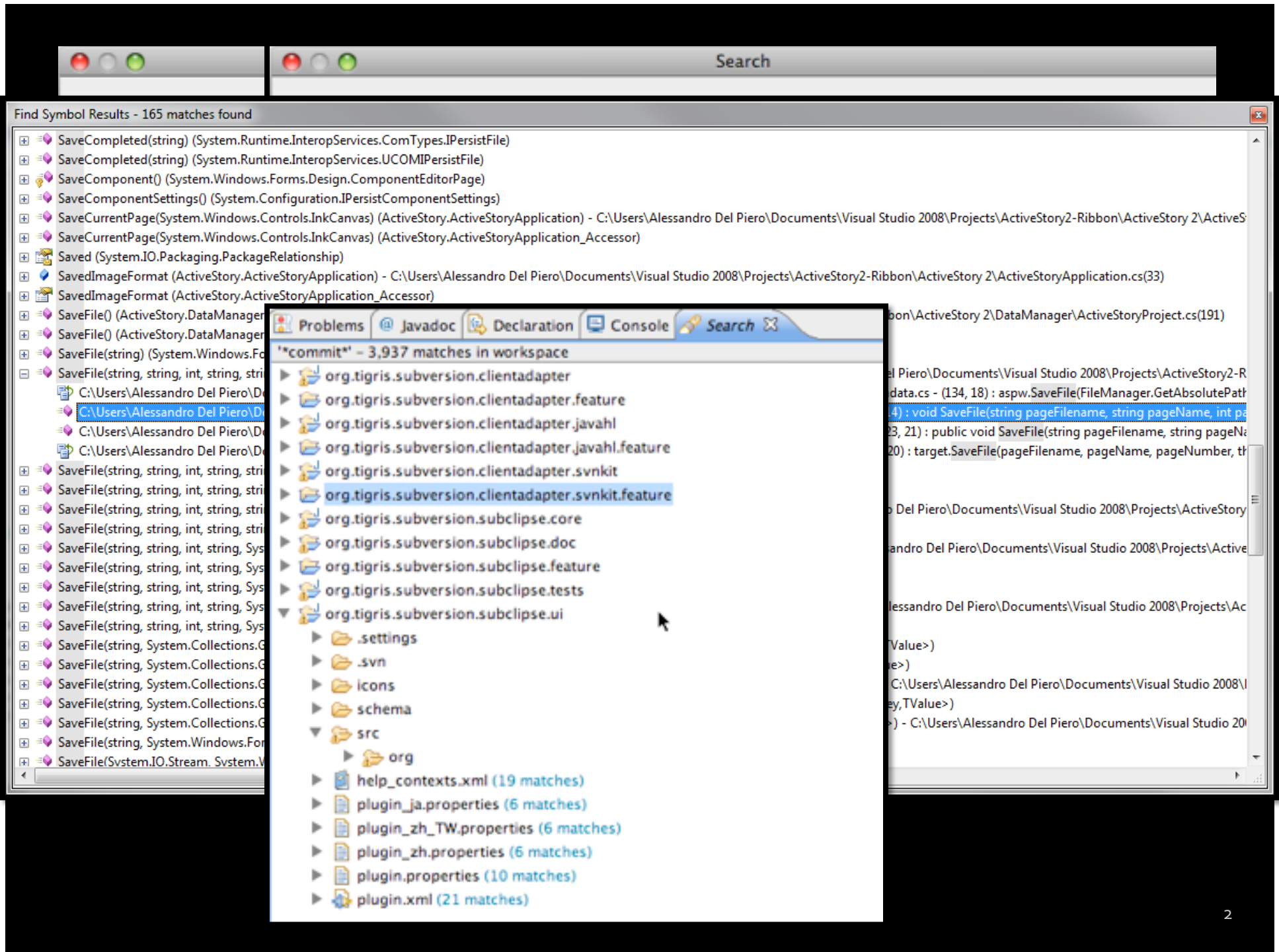


Searching and Skimming: An Exploratory Study

Jamie Starke, Chris Luce, Jonathan Sillito
University of Calgary



Related work

- Current Development Environments
 - Robillard 2004, DeLine 2005, Ko 2006, LaToza 2007
- Tool impact on Program Comprehension
 - Singer 1998, Storey 1999
- Searching Techniques
 - Poshyvanyk 2006, de Alwis 2008

Research Questions

- How effectively can programmers express (using today's tools) the information they are seeking?
- How effectively can programmers determine which of their matches returned from their searches are relevant to their task?
- In what ways can tools be improved to support their information needs and exploring the results of their searches?

Methodology

- 10 participants (2 Graduate Students, 8 Industrial Developers)
- 2 change tasks from Subclipse
- Paired with researcher
- Data
 - Quantitative
 - Data: Number of Results, exploration time, results opened in source editor, etc
 - Statistics of Tool usage and Exploration
 - Qualitative
 - Data: Transcripts and Videos
 - Understand participant Behaviours and Searching patterns

Findings: 5 Key Observations

Full details available in the paper

Hard to Map Concepts to Queries

What you want vs. What you can ask for

Mapping Concepts to Queries

- Example

“I guess a pretty simple case of just taking the first item in a list rather than you know, iterating over the entire list say”

Mapping Concepts to Queries

- Example

“I guess a pretty simple case of just taking the first item in a list rather than you know, iterating over the entire list say”

- What do you search for?

Mapping Concepts to Queries

- Example

“I guess a pretty simple case of just taking the first item in a list rather than you know, iterating over the entire list say”

- What do you search for?
- What limitations do you face in defining your query?

Mapping Concepts to Queries

- Example

“I guess a pretty simple case of just taking the first item in a list rather than you know, iterating over the entire list say”

- What do you search for?

- What limitations do you face in defining your query?

“this only allows [...] one word”

General Searches

Query	Episodes	Average Results
commit	8	1407
commit	3	1585
icon	2	2096
image	2	715
getImage	2	515
space	1	2584
decorator	1	1232
pressed	1	430
modified	1	193

General Searches

Query	Episodes	Average Results
commit	8	1407
commit	3	1585
icon	2	2096
image	2	715
getImage	2	515
space	1	2584
decorator	1	1232
pressed	1	430
modified	1	193

General Searches

Query	Episodes	Average Results
commit	8	1407
commit	3	1585
icon	2	2096
image	2	715
getImage	2	515
space	1	2584
decorator	1	1232
pressed	1	430
modified	1	193

General Searches

Query	Episodes	Average Results
commit	8	1407
commit	3	1585
icon	2	2096
image	2	715
getImage	2	515
space	1	2584
decorator	1	1232
pressed	1	430
modified	1	193

Average

418 (episodes /
w results)

General Searches

Query	Episodes	Average Results
commit	8	1407
commit	3	1585
icon	2	2096
image	2	715
getImage	2	515
space	1	2584
decorator	1	1232
pressed	1	430
modified	1	193

Average

292

(All Episodes)

General Searches

Query	Episodes	Average Results
commit	8	1407
commit	3	1585
icon	2	2096
image	2	715
getImage	2	515
space	1	2584
decorator	1	1232
pressed	1	430
modified	1	193

General Searches

Query	Episodes	Average Results	Avg. Highly Relevant
commit	8	1407	2
commit	3	1585	3
icon	2	2096	2
image	2	715	1
getImage	2	515	1
space	1	2584	5
decorator	1	1232	2
pressed	1	430	1
modified	1	193	1

General Searches

Query	Episodes	Average Results	Avg. Highly Relevant	% Highly Relevant
commit	8	1407	2	0.14
commit	3	1585	3	0.19
icon	2	2096	2	0.10
image	2	715	1	0.14
getImage	2	515	1	0.19
space	1	2584	5	0.19
decorator	1	1232	2	0.16
pressed	1	430	1	0.23
modified	1	193	1	0.52

General Searches

- What they thought was the problem was
 - “I guess a pretty simple case of just taking the first item in a list rather than you know, iterating over the entire list say”

General Searches

- What they thought was the problem was
 - “I guess a pretty simple case of just taking the first item in a list rather than you know, iterating over the entire list say”
- What they searched for
 - commit

General Searches

- What they thought was the problem was

“I guess a pretty simple case of just taking the first item in a list rather than you know, iterating over the entire list say”

- What they searched for

commit

- What did that get them

4770 Results

4 Relevant

Working with General Searches

# Results Opened	Time Spent (s)	# Results Total	# Episodes Occurred
0	47	824	17
1	83	354	33
2	183	7	4
3	171	64	2
4	61	54	1

Working with General Searches

# Results Opened	Time Spent (s)	# Results Total	# Episodes Occurred
0	47	824	17
1	83	354	33
2	183	7	4
3	171	64	2
4	61	54	1

Working with General Searches

# Results Opened	Time Spent (s)	# Results Total	# Episodes Occurred
0	47	824	17
1	83	354	33
2	183	7	4
3	171	64	2
4	61	54	1

Working with General Searches

# Results Opened	Time Spent (s)	# Results Total	# Episodes Occurred
0	47	824	17
1	83	354	33
2	183	7	4
3	171	64	2
4	61	54	1

Working with General Searches

# Results Opened	Time Spent (s)	# Results Total	# Episodes Occurred
0	47	824	17
1	83	354	33
2	183	7	4
3	171	64	2
4	61	54	1

Working with General Searches

# Results Opened	Time Spent (s)	# Results Total	# Episodes Occurred
0	47	824	17
1	83	354	33
2	183	7	4
3	171	64	2
4	61	54	1

Working with General Searches

# Results Opened	Time Spent (s)	# Results Total	# Episodes Occurred
0	47	824	17
1	83	354	33
2	183	7	4
3	171	64	2
4	61	54	1

Working with General Searches

# Results Opened	Time Spent (s)	# Results Total	# Episodes Occurred
0	47	824	17
1	83	354	33
2	183	7	4
3	171	64	2
4	61	54	1

Average
Spent (s)

86

Working with General Searches

# Results Opened	Time Spent (s)	# Results Total	# Episodes Occurred	Avg Time Spent (s) / Result
0	47	824	17	0.057
1	83	354	33	0.23
2	183	7	4	26
3	171	64	2	2.6
4	61	54	1	1.1

Few results navigated

- Participants navigated to only a small number of results



Few results navigated

- Participants navigated to only a small number of results



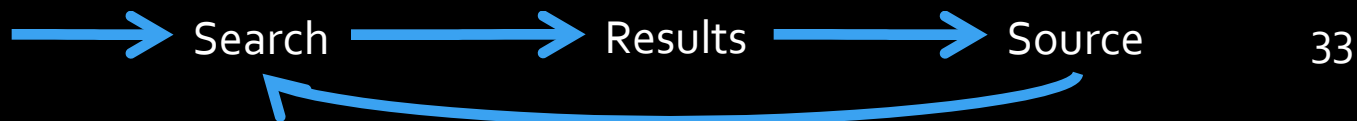
Few results navigated

- Participants navigated to only a small number of results



Few results navigated

- Participants navigated to only a small number of results



Why so few results Explored?

- Too many results

“it's not necessarily feasible to kind of go through them one by one”

- Too many irrelevant results

“is pretty overloaded, [...] I did not want to spend a lot of time going through each search result”

Selective Exploration

- Consider Package Name
- Consider File Name
- Consider Context

Selective Exploration

- **Consider Package Name**
- Consider File Name
- Consider Context

Package Scoping

# Opened Results	Package
1	org.eclipse.compare.internal
6	org.tigris.subversion.clientadapter
5	org.tigris.subversion.subclipse.doc
37	org.tigris.subversion.subclipse.ui

Package Scoping

# Opened Results	Package
1	org.eclipse.compare.internal
6	org.tigris.subversion.clientadapter
5	org.tigris.subversion.subclipse.doc
37	org.tigris.subversion.subclipse.ui

Package Scoping

# Opened Results	Package
1	org.eclipse.compare.internal
6	org.tigris.subversion.clientadapter
5	org.tigris.subversion.subclipse.doc
37	org.tigris.subversion.subclipse.ui

Package Scoping

- Problem Understanding

“seems like a user interface issue”

“I would just ignore all of the other packages entirely and jump directly to the UI”

Selective Exploration

- **Consider Package Name**
- Consider File Name
- Consider Context

Selective Exploration

- Consider Package Name
- **Consider File Name**
- Consider Context

Considering the Name

File Name	# Episodes
ImageDescriptors	4
CommitAction	3
SVNWizardCommitPage	2
SVNUIPlugin	2
SVNDecoratorPreferencesPage	2
SVNAction	2
messages.properties	2
ChooseUrlDialog	2
AddAction	2
add.html	2

Considering the Name

File Name	# Episodes
Image Descriptors	4
Commit Action	3
SVNWizard Commit Page	2
SVNUIPlugin	2
SVN Decorator PreferencesPage	2
SVNAction	2
messages.properties	2
ChooseUrlDialog	2
Add Action	2
add .html	2

Considering the Name

File Name	# Episodes
ImageDescriptors	4
CommitAction	3
SVNWizardCommitPage	2
SVNUIPlugin	2
SVNDecoratorPreferencesPage	2
SVNAction	2
messages.properties	2
ChooseUrlDialog	2
AddAction	2
add.html	2

Considering the Name

- Inclusion

``what would I put in the comments, how would I call my variables, how would I call my class names and function names''

``any number of terms [...] that is related to my hypothesis

Considering the Name

- Inclusion

“what would I put in the comments, how would I call my variables, how would I call my class names and function names”

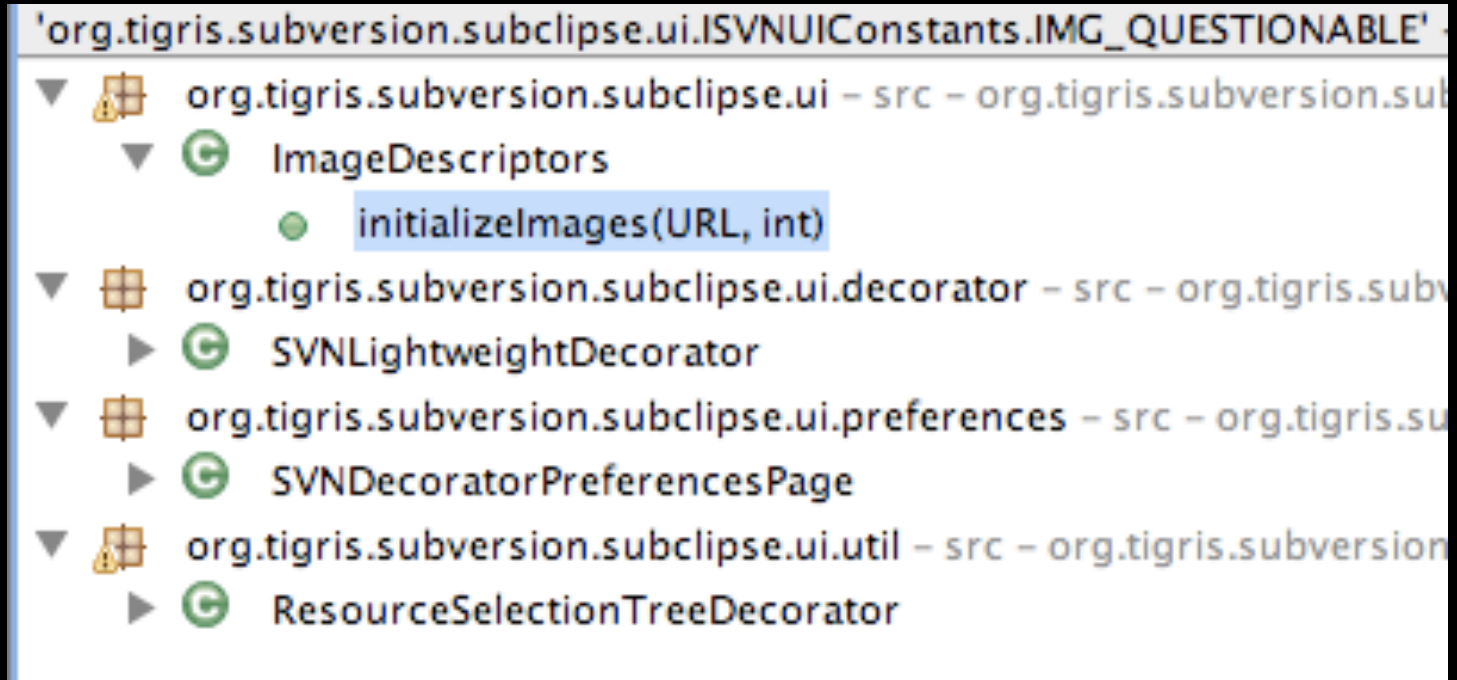
“any number of terms [...] that is related to my hypothesis”

- Exclusion

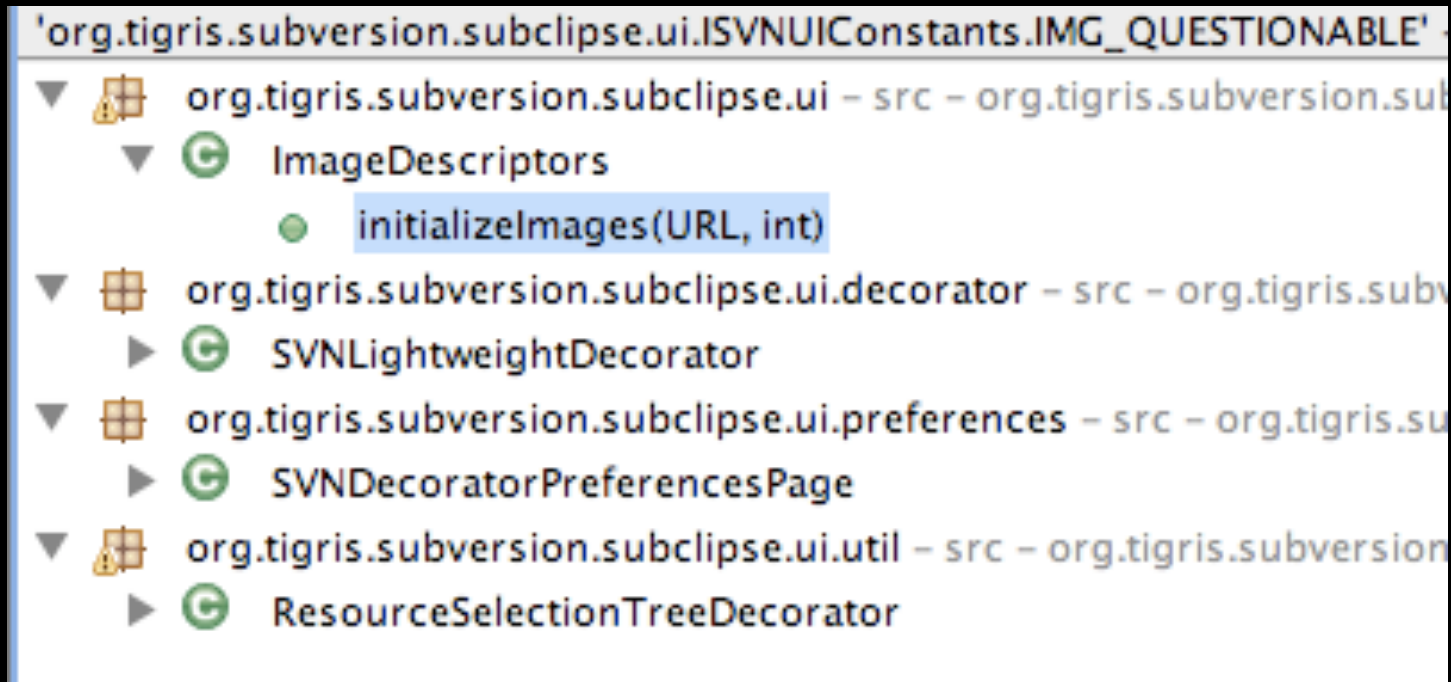
“when they say .html [...] most likely they are auto generated files”

“properties doesn't mean anything to me”

Considering the Name



Considering the Name



“decorator was the more appropriate term in terms of java code”

Selective Exploration

- Consider Package Name
- **Consider File Name**
- Consider Context

Selective Exploration

- Consider Package Name
- Consider File Name
- **Consider Context**

Considering Context

The image shows a file explorer window with the following directory structure:

- org
 - tigris
 - subversion
 - svnclientadapter
 - ISVNPromptUserPassword.java (2 matches)

Below the file name, two search results are shown:

- 108: * @return true if OK was pressed
- 136: * @return true if OK was pressed

Considering Context

- Positive

“the little context it provides when you expand everything [...] gives you a pretty good idea of what you're actually looking at”

Considering Context

- Positive

“the little context it provides when you expand everything [...] gives you a pretty good idea of what you're actually looking at”

- Negative

“you don't get any other information other than the name itself”

Selective Exploration

- Consider Package Name
- Consider File Name
- **Consider Context**

Selective Exploration

- Consider Package Name
- Consider File Name
- Consider Context

Research Goal

- How effectively can programmers express (using today's tools) the information they are seeking?
- How effectively can programmers determine which of their matches returned from their searches are relevant to their task?
- In what ways can tools be improved to support their information needs and exploring the results of their searches?

Implications

- Support for skimming
- Ranking and Grouping of results
- Support for exploring result sets

Support for Skimming

- Result Context is not always obviously available and does not provide comprehensive support for the way participants skim.
 - Context available by default
 - Block level context for structural information

Support for Skimming

- Result Context is not always obviously available and does not provide comprehensive support for the way participants skim.
 - Context available by default
 - Block level context for structural information
- More Information in Source Editor, but programmers don't tend to return to results

Ranking and Grouping of Results

- Ranking of results is only alphabetic, and in some cases suggested to participants that results were more important than they were
 - Many possible rankings
 - Confidence values may provide guidance
- Participants use Packages to narrow down to a smaller portion of the system
 - Provide better grouping of results

Support for Exploring Result Sets

- Participants spent very little time exploring only a small number of results
 - Provide better interaction to support developers in making better use of the results and possibly view them as a basis for ongoing exploration

Implications

- Support for skimming
- Ranking and Grouping of results
- Support for exploring result sets