Observation of Open Source Programmers’ Information Seeking

Khaironi Yatim Sharif, Jim Buckley
Department of Computer Science
University Of Limerick
khaironi.yatim.sharif@ul.ie, jim.buckley@lero.ie

Abstract
Several authors have proposed information seeking as an appropriate perspective for studying software evolution, and have characterized information seeking empirically in commercial software evolution settings. However, there is little research in the literature describing the information seeking behavior of Open Source programmers.

This work describes a holistic schema for Open-Source (OS) programmers’ information seeking, generated through open-coding of questions in OS developer mailing lists. It then reports on a study of the JDT mailing list, showing the types of information sought within this group and, for a subset of the data, it characterizes the responses obtained.

1. Introduction
Information-seeking has been defined as the searching, recognition, retrieval and application of meaningful content [1]. It has been recognized as a core subtask in software understanding, within the activity of software maintenance [2].

There are several previous works (for eg see [3]) that inform on the types of information that programmers seek during maintenance, but most of these studies employ an existing ‘information-types’ schema developed by [4]. As this schema was developed through a theoretical review of the information available in small pieces of code, it is possible that it ignores other artifacts produced by the development team or artifacts of larger code-bases.

In contrast, Ko [5] observed commercial programmers while working in-vivo and identified the types of information they sought through this observation. The work reported on here mirrors this approach in that it describes the information types that programmers seek, based on open-coding of their communications. However, in this instance, it is Open Source (OS) programmers being observed.

Studying OS development in this way is important because OS development is widely distributed in nature, and this would seem to make information seeking more difficult. Unfortunately, there is little research to inform on information seeking among OS programmers (a notable exception is [3], but this was also based on the schema presented in [4]).

This research describes the types of information OS programmers seek, based on open-coding of the questions in their mailing lists. It then assesses the JDT mailing list over a period of 3 years to illustrate the information types predominantly sought by the programmers and, for a subset of this data, uses the list’s ‘thread’ facility to assess if they obtained responses and the timescale of these responses.

2. The Information-Seeking Schema
This current schema was developed by the first author through open coding analysis of the questions contained in three OS mailing lists: the BSF (2002-2003), the JDT (2002-2004) and ECS (2001-2008). This data set resulted in 1117 emails, from which 364 questions were extracted manually. 2 separate dimensions arose from this analysis: Information Focus and Question Strategy (see [6] for details).

Information Focus refers to the external representation referred to in the search. Foci included: System Documentation, Coding Conventions, the Tools and Technologies used, the Protocols adhered to by the group, the System’s Implementation (SI), the System’s Design, File Configuration within the system and the Tasks being undertaken. Within SI there were 2 clear sub-types: where developers directed their query at tracing the cause of a bug or at implementing an enhancement. With respect to Tasks, developers queried if they were complete, and how to test / create enhancements.

Developers often focused their queries at another developer. They Requested Support for outstanding tasks from group members and they often sought to identify the Owner, or expert on, a specified system artifact. Finally, programmers also sought Critique of their work from the group.

Question Strategy refers specifically to the aspect of the information focus, sought by the developer. The 10 strategies identified are presented in Table 1.
### Table 1. Question Strategy (refined from [6])

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Definition and Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>What</td>
<td>Questions which ask what the information focus does.</td>
</tr>
<tr>
<td>How</td>
<td>How an information focus achieves its goal or is employed. How to proceed.</td>
</tr>
<tr>
<td>Why</td>
<td>Asking for a purpose / explanation of the information focus</td>
</tr>
<tr>
<td>Who</td>
<td>Asking for the relevant persons to consult with on the information focus.</td>
</tr>
<tr>
<td>Where</td>
<td>Asking about the location of the focus or about a location within the focus.</td>
</tr>
<tr>
<td>When</td>
<td>Question about time of an occurrence.</td>
</tr>
<tr>
<td>Perm.</td>
<td>Permission to do something.</td>
</tr>
<tr>
<td>Conf.</td>
<td>Seeking Confirmation from others</td>
</tr>
<tr>
<td>Relations</td>
<td>Directs itself at relationships between entities rather than at the entities.</td>
</tr>
<tr>
<td>Instruction</td>
<td>Question that ask others to do something.</td>
</tr>
</tbody>
</table>

### 3. Information seeking in the JDT List

This schema was used to examine the JDT developer mailing list from 2002 to 2004. Each question within the list was extracted and then categorized by Information Focus and Question Strategy.

In terms of Information Focus, results suggest that the developers sought information frequently on the Tools and Technology they used (34%) and, to a lesser degree, on System Documentation (11.5%), Task Implementation (11.5%), SI-Debug (11.5%), and SI-Enhancement (10.5%).

Within the Question Strategy dimension, programmers concentrated on Confirmation (33%), How (31%) questions and to a lesser degree What, Why, Who and Where questions (approximately 7% each).

In most categories, information seeking remained fairly constant over time, when it might intuitively be expected to fall off as expertise grows. This counter-intuitive (non) trend could possibly be explained by new developers joining the group or by the large scale of the project, where programmers are still familiarizing themselves with the code-base and technologies used over the 3 years.

Also surprising was the amount of System Documentation requests, a trend that contradicts existing information seeking findings [2]. It is possible that, due to delocalization, OS programmers may be motivated to produce better documentation or must rely on it more.

As in other work [5], there seems to be a strong team-orientation to the questions. Who questions and Confirmation question made up approximately 44% of all JDT questions over the 3 years.

In reviewing the responses to the JDT questions, for 2002 the most striking feature was the amount with no responses (54%). Excluding these, the average number of responses per question was 2.1, with an average response time of 6.9 days.

In terms of the frequent query types, Tools and Technology were only responded to 20% of the time, but when they were, they were responded to quickly (avg: 3 days). Confirmation requests fared better with 57% receiving a response (avg response time: 3.5 days). Finally How questions had a 33% response rate with a larger time delay (avg: 5 days)

### 4. Conclusion

In applying the information seeking schema to the JDT mailing list, this study has highlighted a number of interesting findings. Specifically, it suggests a severe ramp up with respect to the tools and technology used, a reliance on documentation not as apparent in commercial, co-located developers and, a strong sense of team awareness. It also suggests that ‘Tools and Technology’ information in particular, is difficult to acquire through the mailing list.

### Acknowledgments

Supported by Science Foundation Ireland grant number 03/CE2/I303_1.

### References


7. References
