Description for the analysis procedure for the article "Understanding initial API comprehension".
Accompanies the ICPC 2023 article "Understanding initial API comprehension".

In this document, we will explain how the data was prepared and coded for the "Understanding initial API comprehension" article.
Due to the sensitivity of the data, we provide only carefully redacted examples of the interviews to ensure no participant or company can be identified.

1. Excerpts

After the interview data has been transcribed, the data is split into excerpts that contain the message of the participant in a compact form.

The interview data itself reflects natural language conversation, and contains many references such as "I had to do it" referring to some it that was mentioned earlier, or "I had to figure that out" referring to some that from earlier on in the discussion.
The data can also be quite messy, and contain a lot of filler words and repeating sentences.

To make the data easier to analyse, the data is split into excerpts, which roughly reflect one sentence or topic said by the participant.

Some of the interviews were also carried out in Finnish, while some of them were carried out in English debending on the participant. In this stage, all the excerpts are written in english to ensure the excerpts are easily comparable.

Example:

In this example, the participant is discussing how they found the API they wanted to use for a task.
A: Interviewer
I: Interviewee

A: "Did you know that this was the API you had to use straight away?""

I: "I had to look it up from somewhere"

A: "Do you remember where you found it?"

I: "Umm. I googled it. "

To make this data easier to use in the following phases of the data analysis process, this data was made into an exerpt like so:
"Participant used google to look up which API they should use for the task at hand"

2. Coding within one interview

In this step, the main elements present in each individual interview, and the connections between these elements were identified.

2.1 Collecting connected excerpts

During the interviews, participants often jump back and forth in time, add details to previous parts of the story or jump forward to explain how things play out in the next steps of the story.
After the interviews have been split into excerpts, the excerpts are organized so that all the excerpts related to the same topic are collected together.

Example

In the example below, we have marked parts of interview data from one interview, which we identified as being related . The excerpts related to these parts are shown in the diagram in section 2.2.

In this example interview, our participant first told us, that they used google to find a suitable API for the task at hand (1), and that "since they knew what to search for" (2), it was easy for them to find a suitable API.

After the interviewee asked them how they knew what to search for, they elaborated that they knew the right keywords (3). Later they also elaborated, that they knew the terminology used to refer to certain types of APIs which allowed them to decide on the right keywords (4).

All these excerpts were related to the same topic - how the particpant used their knowledge to search for a suitable API.
Identification of the related excerpts from the interview data is shown in the example below (translated from Finnish to English):

> I[00:05:06]: I googled these libraries (1). With certain keywords you could find thigs quite easily. I know what to search for so I can find it easily (2).
> [...]
> I[00:05:24]: Actually from google I ended up....and from there I found things that I liked.
> [...]
> I[00:06:02]: I knew to use the right keywords (3). For example, I knew that ORM is a commonly used abstraction technique with databases. In general I knew that there are

> these kinds of libraries that abstract SQL queries. With these kinds of keywords I could find it quite easily... (4)

2.2 Coding the elements of one interview

As the exerpts of each interview are collected and related exerpts are connected, a rough understanding of the elements of the participant's story starts to form.

At this stage of data analysis, the exerpts related to one interview are organized and coded. In this study, the aim is to understand certain aspects of API comprehension, so at this stage we seek to find and code these aspects from the excerpts. Furthermore, we seek to understand how these aspects are connected together.

The aspects of API comprehension we seek to understand are:

- **Actions**: Cognitive, and physical actions performed by the participant to comprehend some aspect of an API
- **Information needs**: Information about an API the participant wanted or needed to have at some point of the API comprehension process
- **Information sources**: The sources participant sought information from or acquired information from during the comprehension process
- **Contextual factors** : Elements that affect the comprehension process in some way
- **Mental models**: Participant's understanding of the API at some point of the API comprehension process
  However, the above list is not comprehensive, and there may be important aspects of API comprehension that are not on the list but emerge from the data. These details are also coded at this stage.

Example

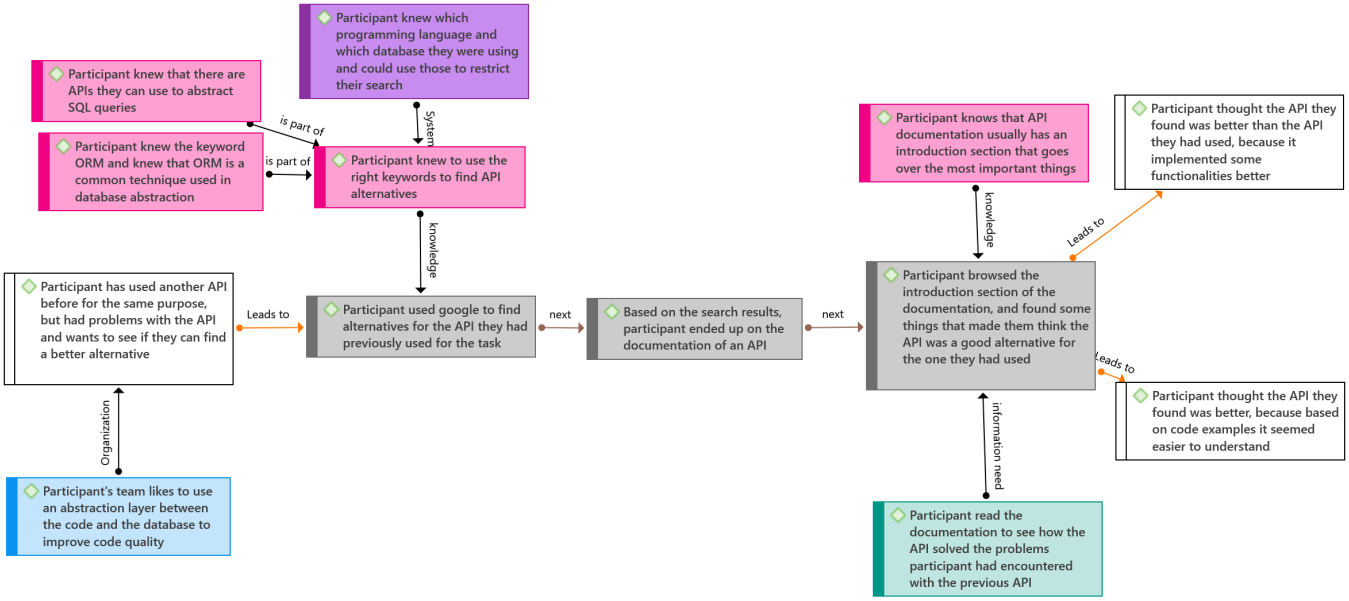
Below is an example of one of a part of the analysis of one interview at this stage.

The diagram shows the exerpts related to this part of the interview. All the excerpts have been organized, related excerpts have been connected together, and the different aspects of the API comprehension process have been coded.

In gray are excerpts related to the participants actions organized into a timeline which shows the orger of actions from right to left. In pink are excerpts related to their knowledge, in blue is an excerpt related to contextual factors coded as organization, in purple excerpt related to the existing system, and in green an excerpt related to the participants information need.

White excerpts are related to the situation that prompted the comprehension process, and the knowledge acquired during this part of the comprehension process.

At this stage of the analysis, the excerpts were used as is. As can be seen from some of the excerpts, they may contain unrelated information or information that may also be related to other codes. However, at this stage we did not split the excerpts.



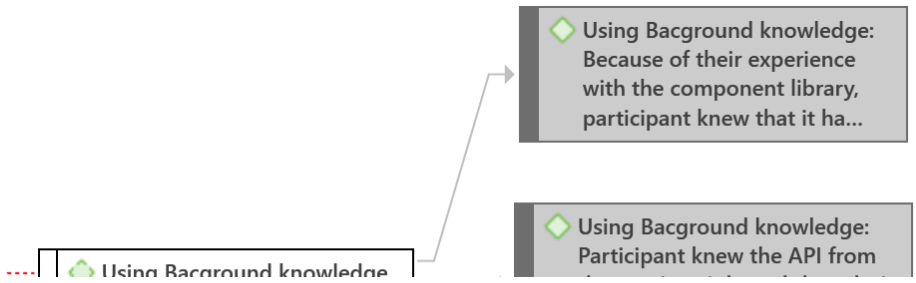
3. Connections across interviews

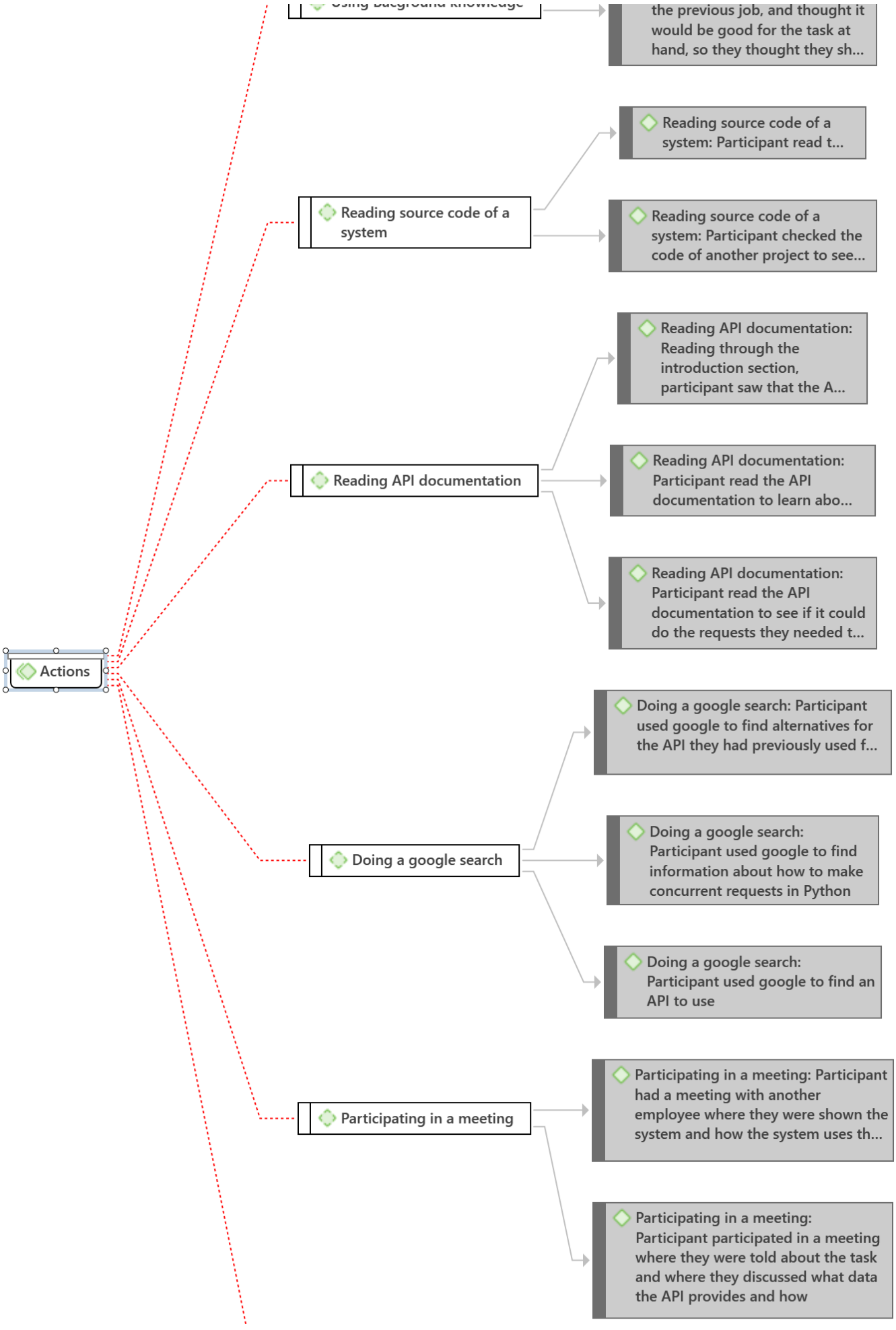
In this step, the elements identified in the previous step are connected across the different interviews. For example, all excerpts related to participant's knowledge or their actions are collected from across all interviews.
These excerpts are then further coded to identify the types of items within each aspect of the API comprehension process.

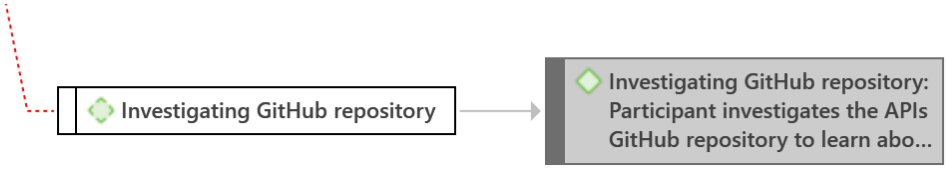
Example

Below is an example of codes in the Action aspect. All the excerpts related to the participant's actions have been collected from across the interviews, and similiar excerpts have been collected together and coded. For example, the excerpt related to a google search action shown in the previous example can now be found from the diagram below under the code "Doing a google search".

From this diagram we can see, that by doing this we were able to find 6 types of actions that were present across the interviews.




the previous job, and thought it would be good for the task at hand, so they thought they sh...
Reading source code of a system: Participant read t...
Reading source code of a system
Reading source code of a system: Participant checked the code of another project to see...
Reading API documentation: Reading through the introduction section, participant saw that the A...
Reading API documentation
Reading API documentation: Participant read the API documentation to learn abo...
Reading API documentation: Participant read the API documentation to see if it could do the requests they needed t...
Actions
Doing a google search: Participant used google to find alternatives for the API they had previously used f...
Doing a google search
Doing a google search: Participant used google to find information about how to make concurrent requests in Python
Doing a google search: Participant used google to find an API to use
Participating in a meeting: Participant had a meeting with another employee where they were shown the system and how the system uses th...
Participating in a meeting
Participating in a meeting: Participant participated in a meeting where they were told about the task and where they discussed what data the API provides and how


Investigating GitHub repository
Investigating GitHub repository: Participant investigates the APIs GitHub repository to learn abo...