An Exploratory Study of Web Foraging to Understand and Support Programming Decisions
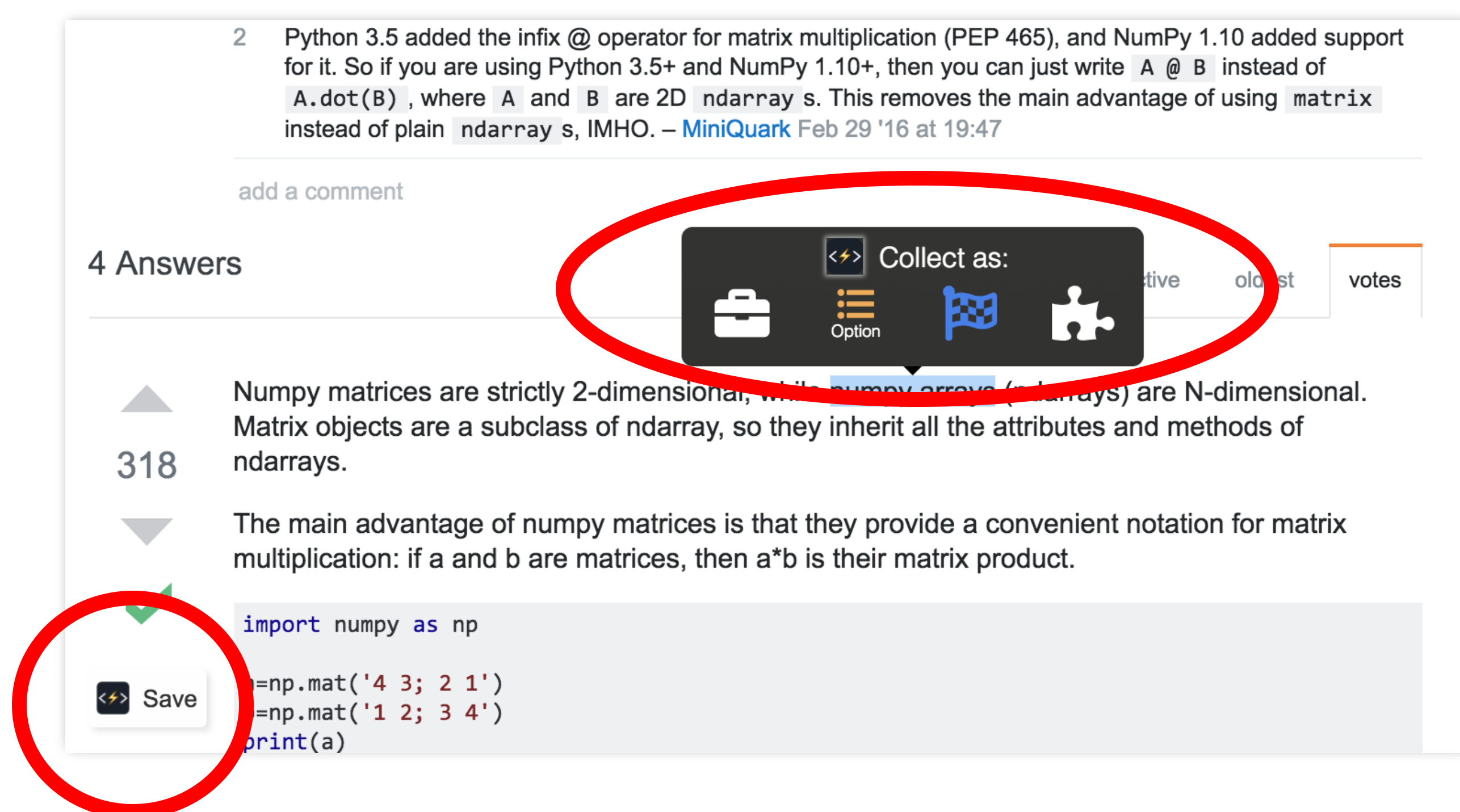
Jane Hsieh, Department of Computer Science, Oberlin College

Michael Xieyang Liu, Brad A. Myers, and Aniket Kittur, Human Computer Interaction Institute, Carnegie Mellon University

INTRODUCTION

Programming is a cognitively demanding process involving a variety tasks such as information gathering, sensemaking and decision-making. Programmers today are becoming more reliant on the growing amount of information available. To understand the types of questions and solutions that programmers seek, we analyze Stack Overflow questions by category and assess the proportion of questions that were compatible with the **comparison table** – our alternative model of representation, designed to reduce the costs of sensemaking and decision-making. Finally, the usability of the comparison table as a representation model is tested using a within-subjects A/B test.

SAMPLING AND CATEGORIZATION
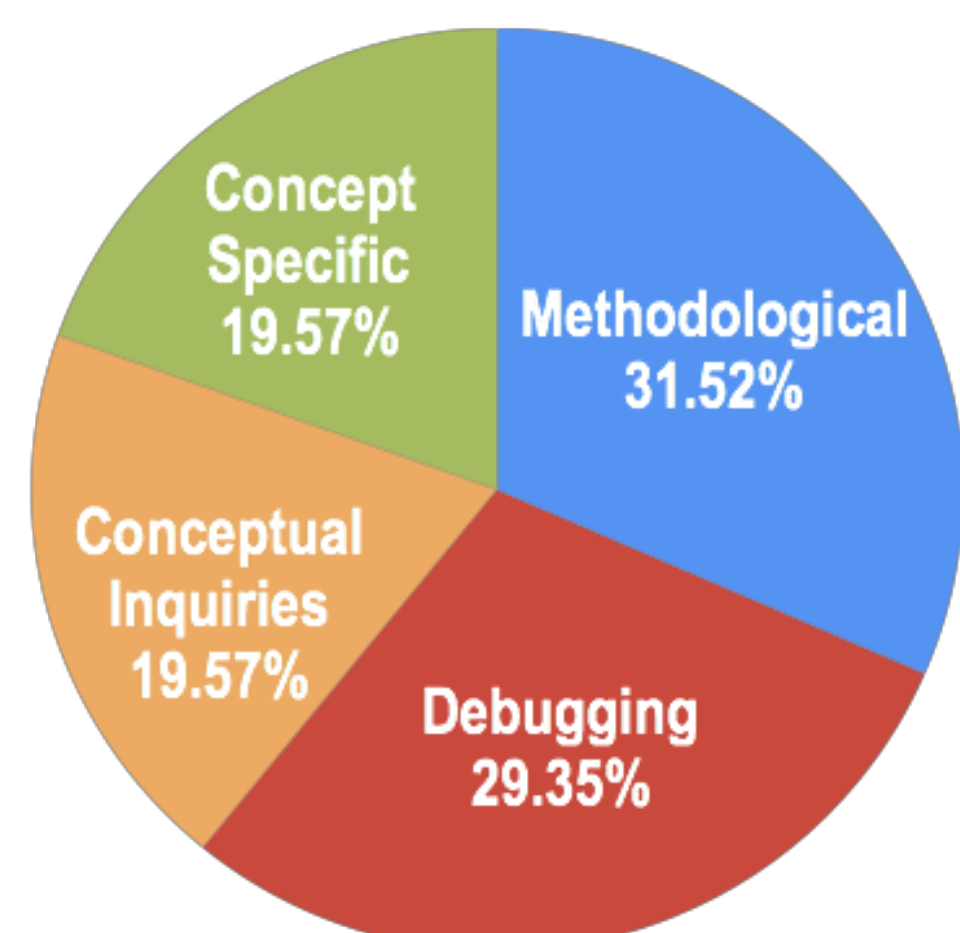
The first sample of 92 Stack Overflow questions were classified into 4 broad categories:

1. **Methodological**, tool-foraging questions where a programmer seeks methods or code snippets.
2. **Debugging** questions with specific context such as error messages or step-by-step instructions.
3. Abstract **conceptual inquiries** which often trigger discussions and debates or summaries of concepts
4. **Concept-specific** questions where a forager seeks understand how particular methods/commands work

Initial Queries	Questions
"how to answers: 10" with active filter	21
"which should views:500000" [b]	20
Hot filter ("hottest" questions today)	20
Month filter	19
"how to" with votes filter	13

- Results from initial analysis: questions from the **methodological** and **concept-specific** categories are more compatible with the comparison table.

To gather questions from popular questions and as well as the long tail of the distribution, we formed more specific queries to find our final samples:

| Sample Queries	Compatible Questions	
Most viewed questions	44 / 50	88%
Questions asked on 6/15/18 with 3 or more answers	44 / 90	49%

MODEL OF REPRESENTATION

A comparison table uses 3 types of collected information:
- **Options** – rows are ways of completing a task
- **Criteria** – columns of requirements needed for solving the problem or completing the task
- **Snippets** – code or other relevant web content

Up and down thumbs rate options according to the criteria, supported by linked snippets.

Above comparison table was constructed using the UNAKITE tool, a Chrome extension that helps end-users capture and organize web content into comparison tables. https://janeon.github.io/help/

PRELIMINARY USER STUDY

Goal: To test whether the model improves speed and/or accuracy of the sensemaking process.
Method: Within-subject A/B test
- A: subjects were asked to answer comprehension questions using a comparison table
- B: subjects were asked to answer comprehension questions using the original Stack Overflow page
- Question and tool-use were both counterbalanced for order
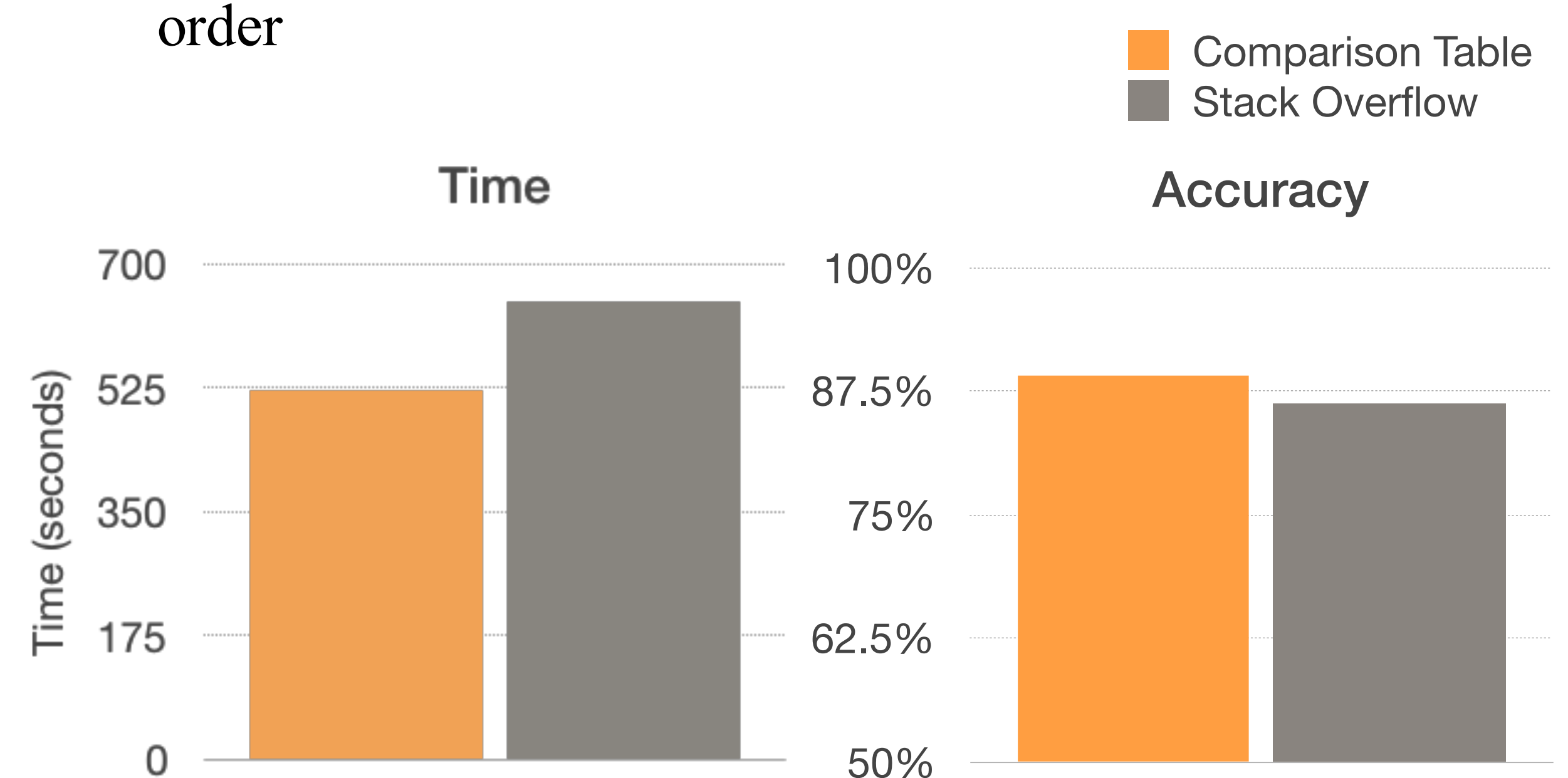
Subjective rating results
- User's likelihood to use the comparison table for a programming task in the future: 4.75 / 7
- Level of difficulty with using the table: 3.5 / 7

CONCLUSIONS AND FUTURE WORK

From these results we have observed initial success with the comparison table as a mental model for representing information found in question pages on Stack Overflow. Assistive tools can be developed to help facilitate and improve the user's sensemaking and decision-making processes. Future work is needed to study to what extent decision questions like these are common in other domains besides programming, and if our proposed tools could help in those contexts as well.