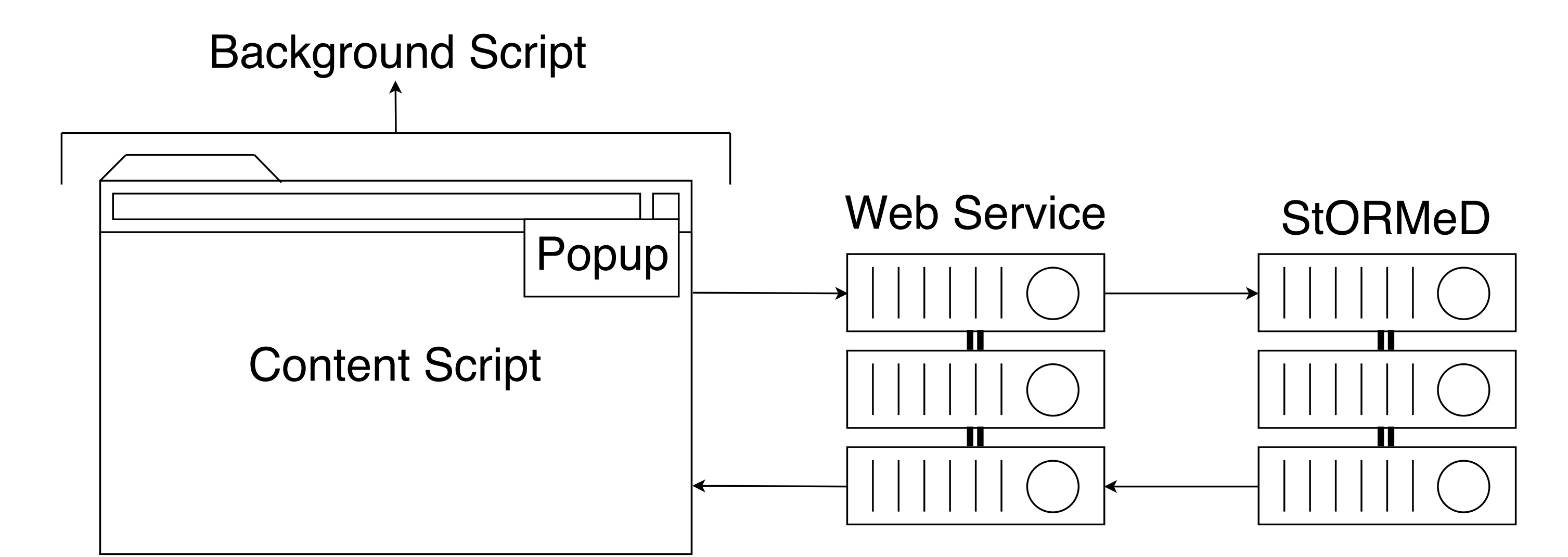




Introduction

To solve the task at hand, developers often rely on the world wide web to find answers. Whether these answers come from Q&A sites, documentation, or tutorials, developers need to aggregate those resources, and select the pieces of information needed to complete the task. Due to the high amount of resources available, this process can lead to information overload which can be mitigated with summarization. In this project, we present WebDistiller, a tool that provides interactive extractive summaries of either a single webpage, or a set of web pages by considering the history of web pages visited by a developer.

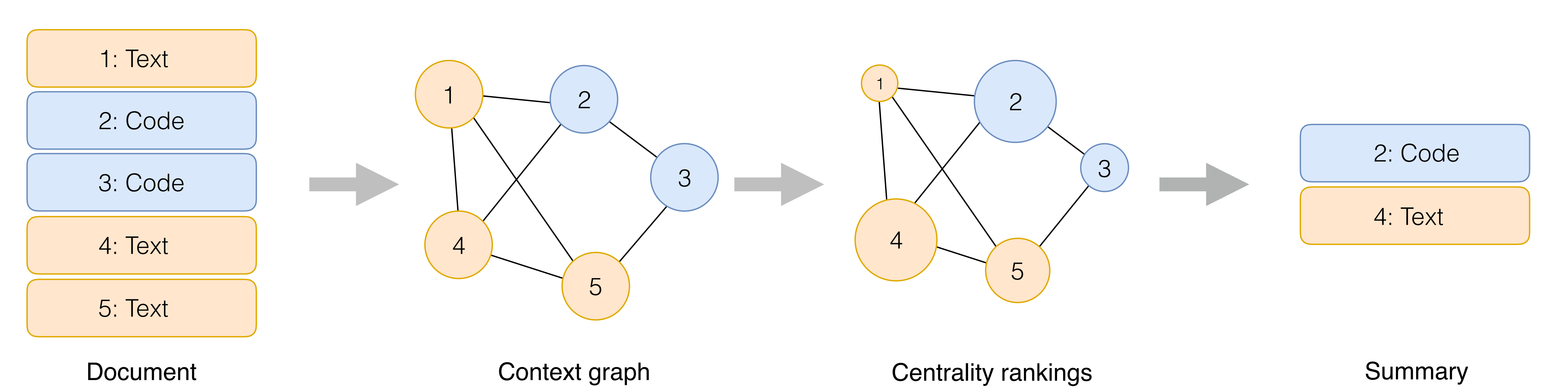
Architecture



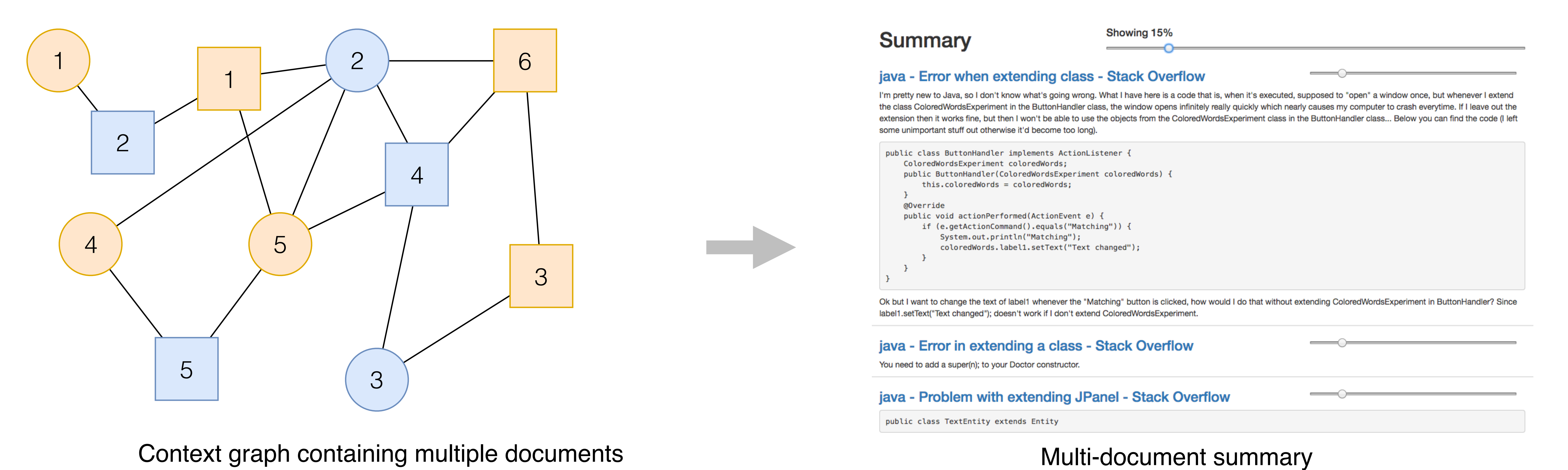
Approach

First, the HTML content of the page is analyzed, extracting both the structure and the text. Thus, the data is packaged and sent to the web service, all using a dedicated Chrome extension. Once the data is sent, the service takes the incoming data, and calculates the importance of each section by determining the degree of centrality. This information is then returned to the extension, which allows the user to filter out the units deemed less important, reducing the amount of information currently displayed on the page, and only showing those that have more to do with the current context.

Single Document



Multiple Documents



User Interface

Session Information

URL

http://localhost:9000

User Id

eMXpKZxuPUM19ydDQp3GJymeAfNehBBB

Save

Content

36%

Generate Summary

Chrome Extension Popup

Conclusion

WebDistiller is a novel approach to solve the issue of information overload in the development context. It leverages an holistic approach to ranking, enabling the global context to play an important role in determining the importance of the different sections when performing summarization of either a single, or multiple documents.

By providing a simple, minimalistic slider, the user experience is kept as simple as possible.

The major limitation is the granularity of the summarisation, as it is not possible to summarize the contents of the single paragraphs.

