

Keyword Extraction Based on Implicit Feedback

Jakub Kříž

Supervisor: Tomáš Kramár

The goals

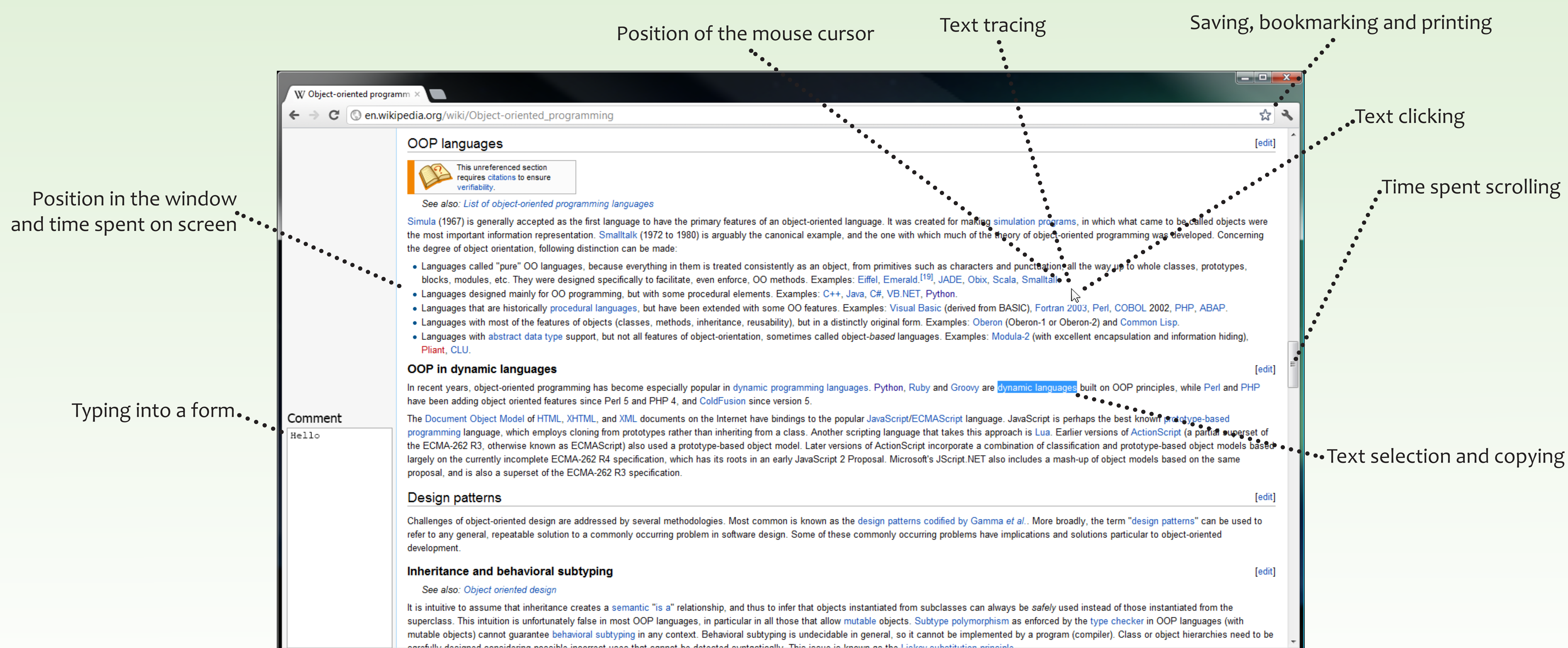
- collect and process the implicit feedback from the users on the internet
- use the feedback to infer which documents the user liked and disliked
- use the feedback to infer which parts of the documents interested him the most
- make keyword extraction more accurate

Implicit Feedback

- everything the user does in the web browser
- collected via implicit interest indicators
- when an indicator is detected the user is likely to be interested in that part of the document
- detected and calculated via JavaScript

The implicit interest indicators

- *global indicators* show interest in the entire document
- *local indicators* indicate the interest in a part of the document
- varying accuracy and reliability



Processing the Indicators

- the document is broken down to text elements - headings, paragraphs
- each element starts with a ranking of 0
- when an indicator is detected the ranking of the target element is increased
- the increase depends on reliability and accuracy of the indicator
- copied and selected keywords are extracted directly from the text and using tf-idf from the ranked elements

Evaluation

- we evaluate the two methods used for keyword extraction
 - » users considered 100% of the keywords extracted from the copied text to accurately represent their interest in the document
 - » users considered 62% of the keywords extracted from the selected text
 - » experimental evaluation of the extraction from ranked elements is currently in progress