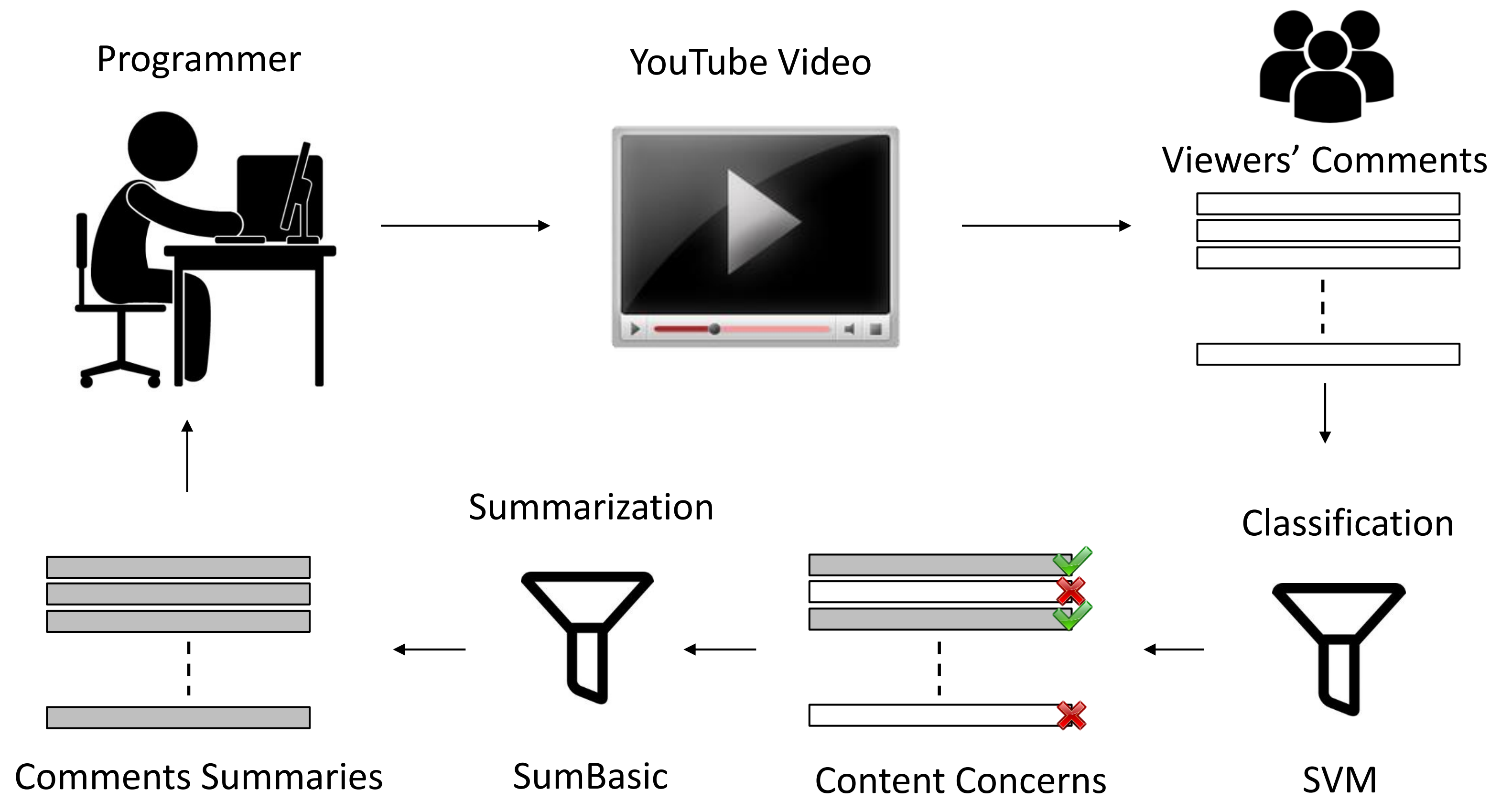
Analyzing YouTube Comments on Coding Tutorial Videos

Elizabeth Poché and Anas Mahmoud
Louisiana State University, Baton Rouge

Abstract

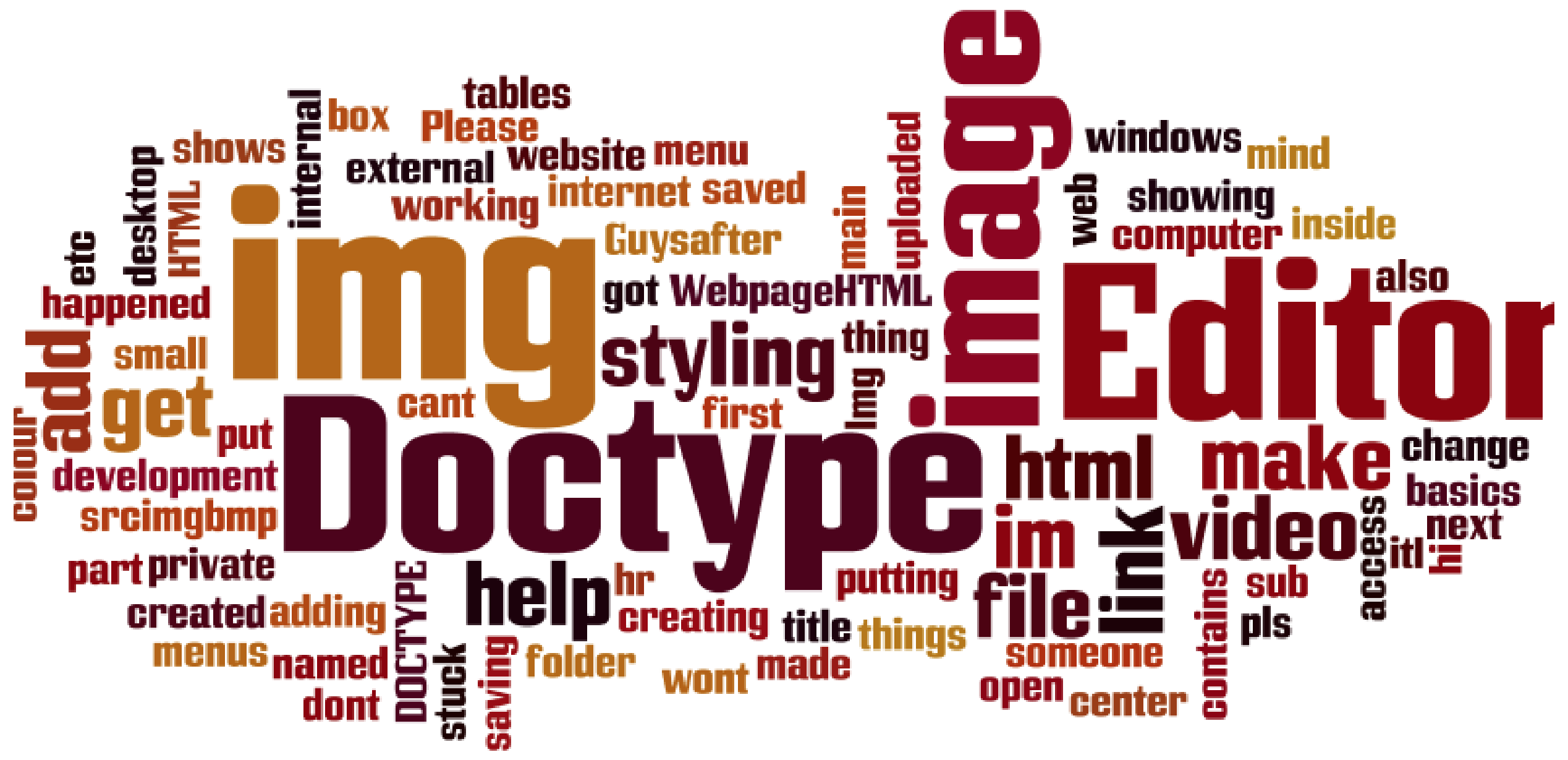
Video coding tutorials enable expert and novice programmers to visually observe real developers write, debug, and execute code. Previous research in this domain has focused on helping programmers find relative content in coding tutorial videos as well as understanding the motivation and needs of content creators. In this paper, we focus on the link connecting programmers creating coding videos with their audience. More specifically, we analyze user comments on YouTube coding tutorial videos. Our main objective is to help content creators to effectively understand the needs and concerns of their viewers, thus respond faster to these concerns and deliver higher-quality content. The results show that Support Vector Machines can detect useful viewers' comments on coding videos with an average accuracy of 77%. The results also show that SumBasic, an extractive frequency-based summarization technique with redundancy control, can sufficiently capture the main concerns present in viewers' comments.

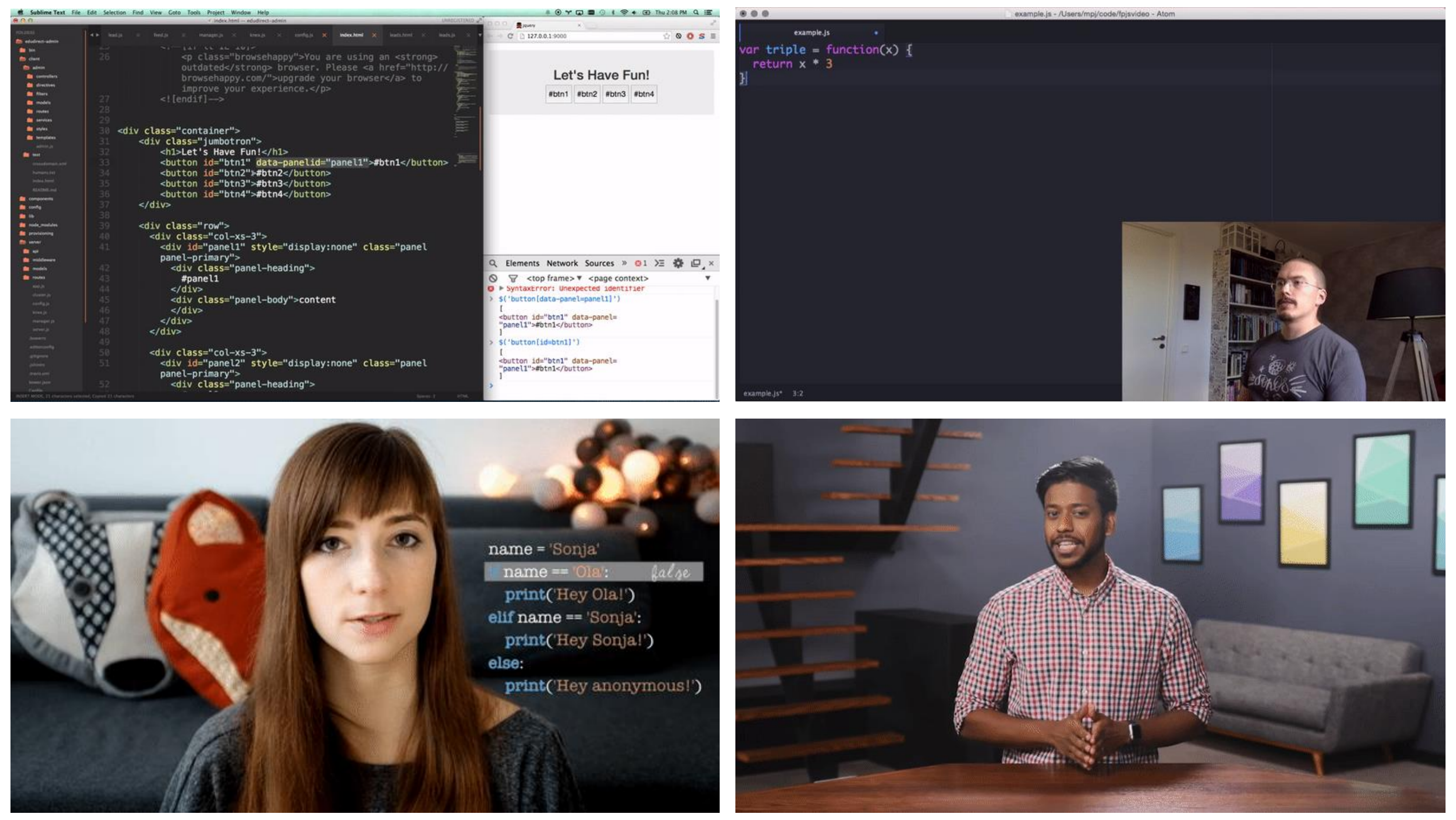
Approach

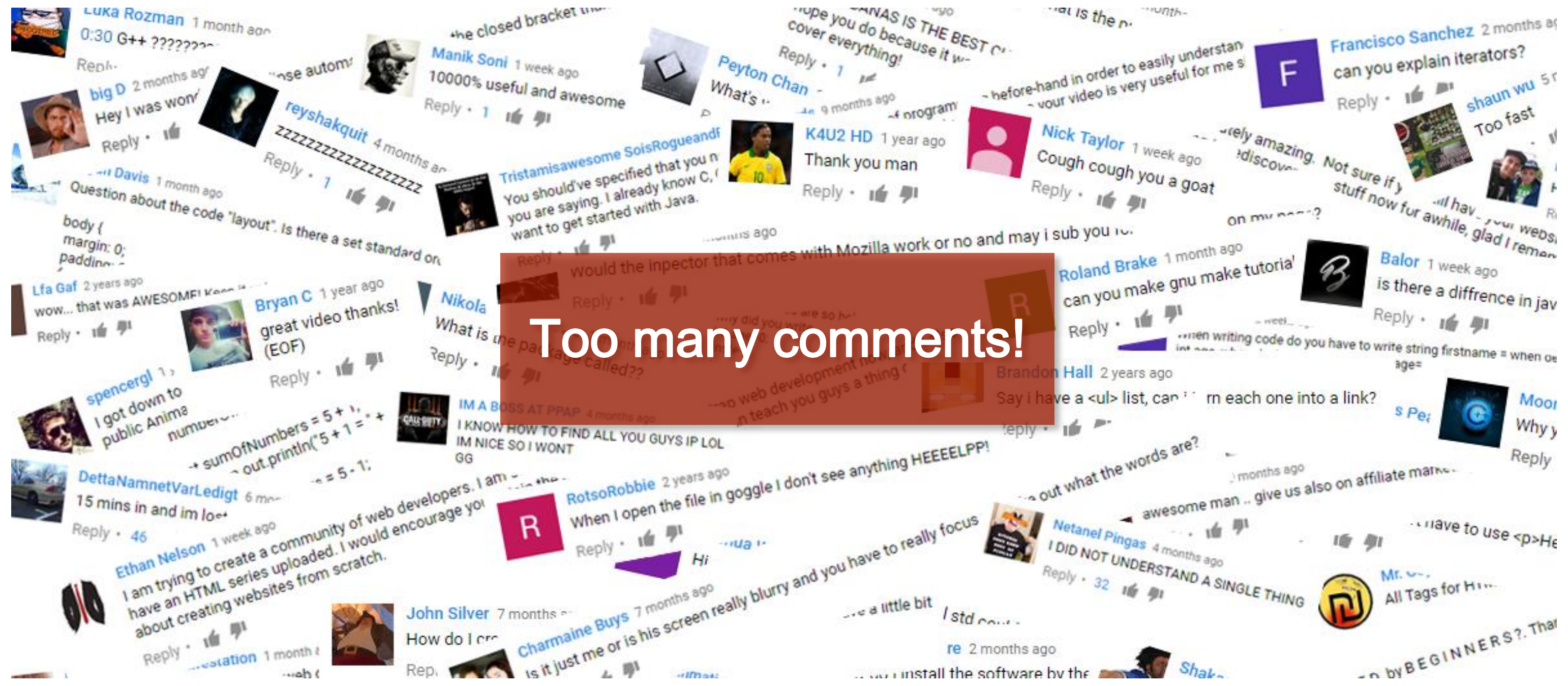


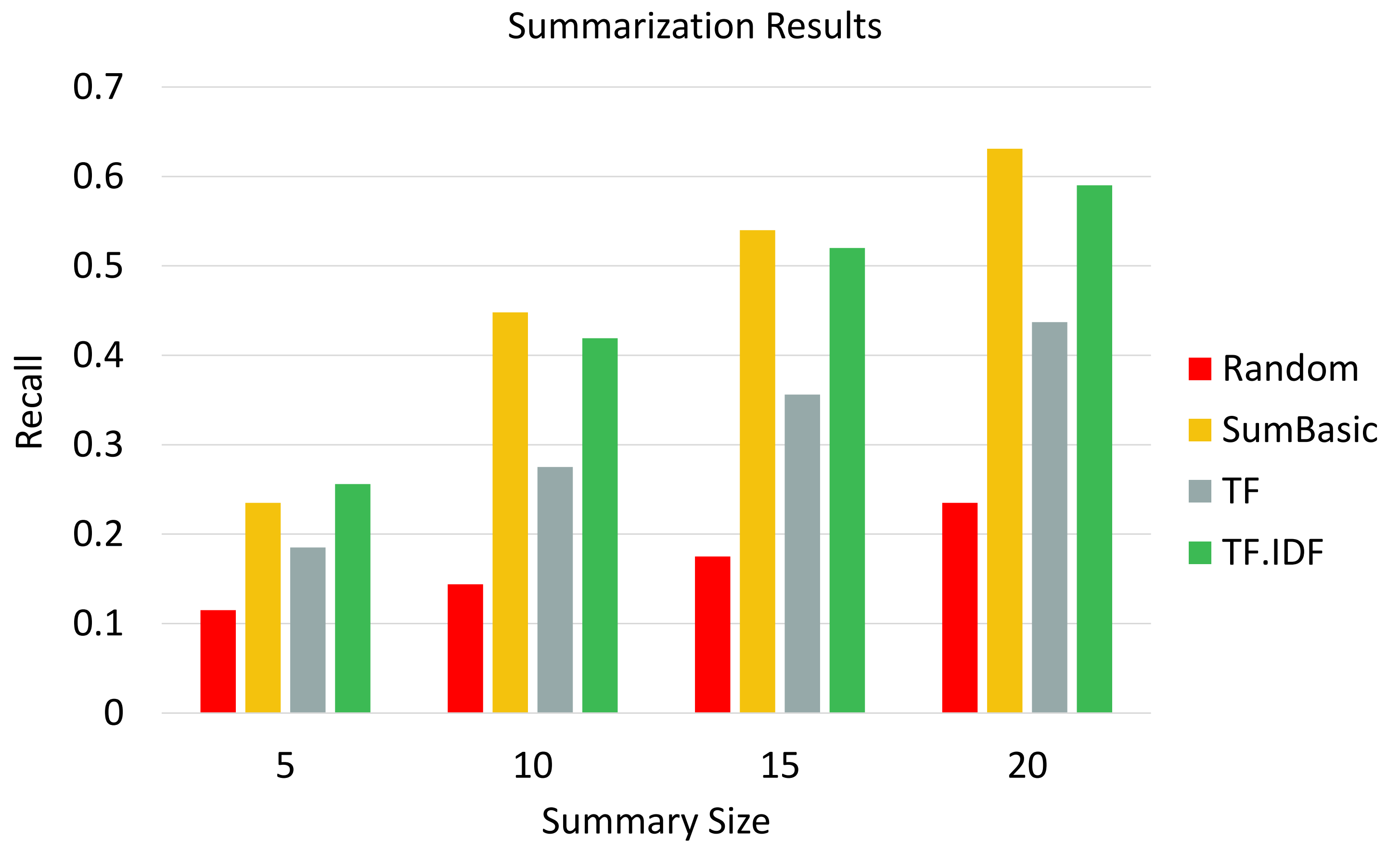
Comments Summarization

We investigate the performance of different text summarization techniques in capturing the main topics raised in the content concern comments on coding videos, including TF, Hybrid TFIDF, and SumBasic.

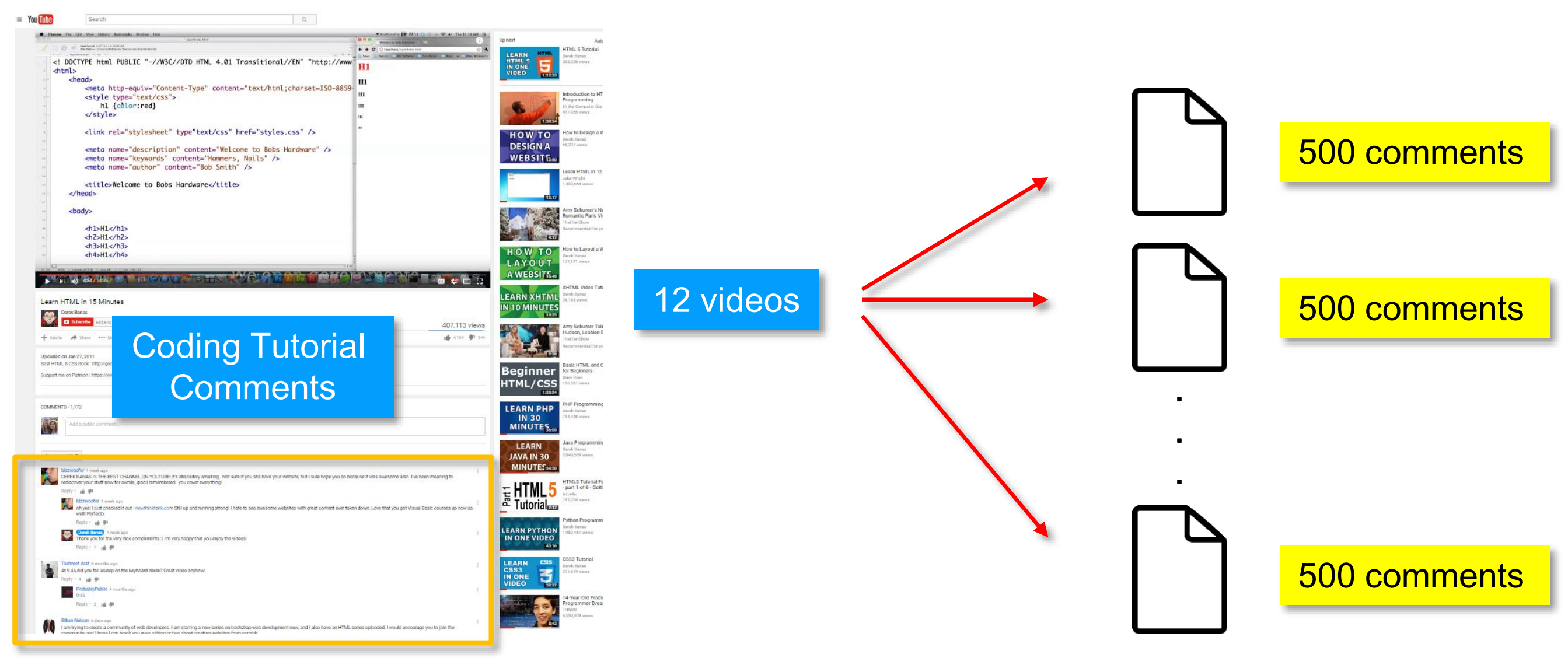




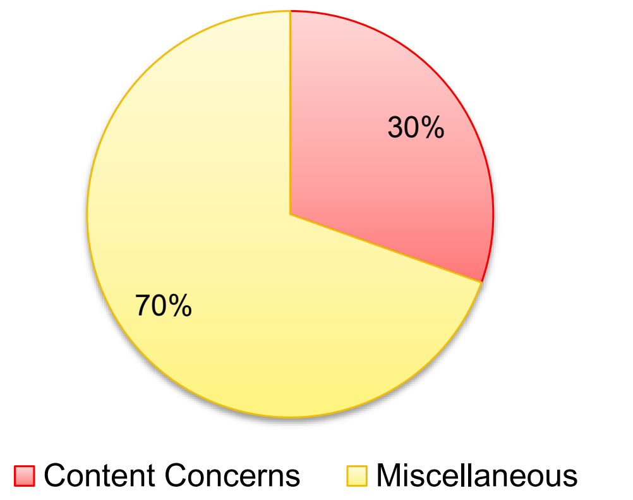




Data Collection and Qualitative Analysis



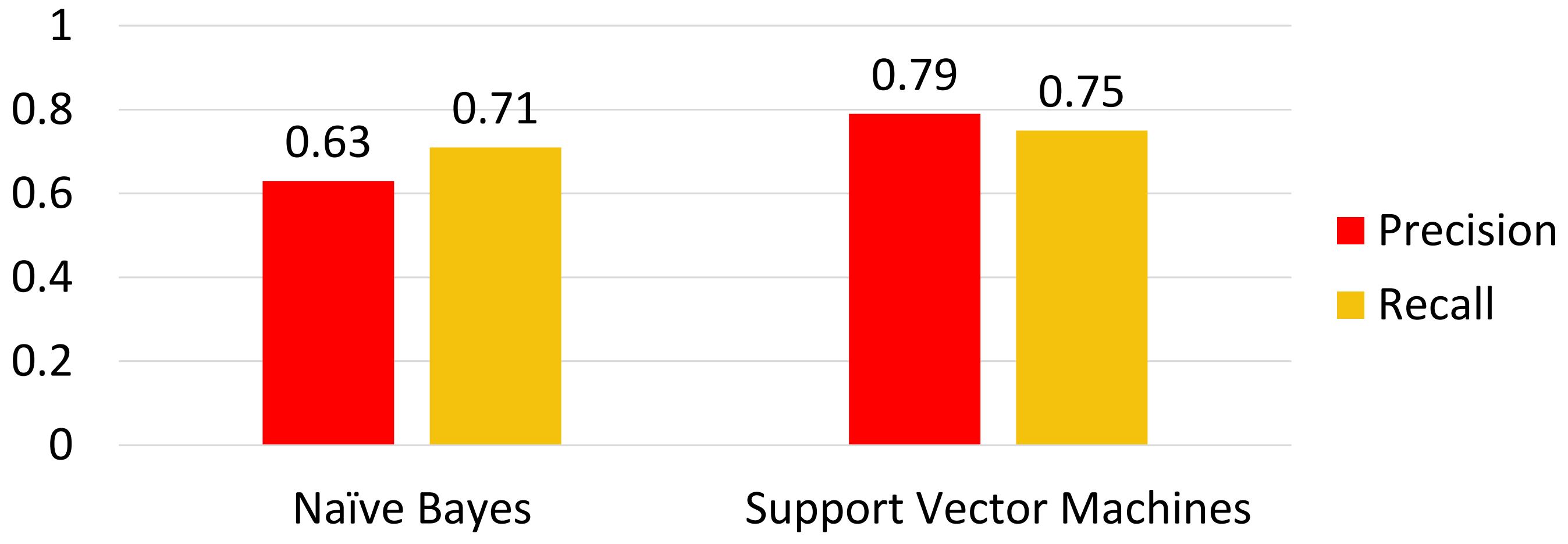
The sampled comments in our dataset were manually examined and classified by a team of five annotators. The team members have an average of 4 years of experience in programming. A tool was created to aid in the manual classification process. Each annotator classified each comment and saved the results to a separate database. The tool then merged the different classifications and majority voting was used to classify each comment. No time constraint was enforced to avoid fatigue.



The results of the manual classification process show that around 30% of the comments were found to be content-related, or useful, meaning that the majority of comments are basically miscellaneous.

Classification

This phase is concerned with automatically classifying input comments into informative (useful concerns) and other (not useful) comments using Naïve Bayes and Support Vector Machines.



Contact

Elizabeth Poché
Department of Computer Science and Engineering
Email: epoche55@gmail.com

Paper

E. Poche, N. Jha, G. Williams, M. Vesper, J. Staten and A. Mahmoud, Analyzing User Comments on YouTube Coding Tutorial Videos, International Conference on Program Comprehension (ICPC), (10 pages) (AR: 19%), 2017