# Review of Developing Teaching Material Using a Combination of Data Mining, Machine Learning, Crowdsourcing Techniques

**Yanbo Chen**
ychen380@illinois.edu

## Introduction

As the computer technology quickly advances, online education becomes more popular and people are more likely to search educational resources from the Internet. The quality of these educational resources varies considerably thus affecting the effectiveness of learning from online materials. Hence, there is a need of developing teaching material using state-of-the-art computer science techniques such as data mining and machine learning to make online education more effective. In this review, three research papers regarding online education technology will be discussed.

## Leveraging Community-Generated Videos and Command Logs to Classify and Recommend Software Workflows

In this paper, the researchers develop a hierarchical approach which combines topic modeling and frequent pattern mining to classify the workflows provided by the software application in terms of community-generated videos and command logs as well as four recommender algorithms to recommend workflows and related videos to users. The initiative of this approach is that the users of software applications are usually unaware of the efficient workflows and following inefficient workflows to complete a task. There are

During the classification step, community-generated videos and command logs are used as the dataset. In this step, the researchers proposed a combination of topic modeling approach and frequent pattern mining approach to identify frequent user tasks. In practice, frequent pattern mining approach is largely affected by the frequent items. Therefore in the proposed hierarchical structure, topic modeling is applied above the frequent pattern mining approach to prevent frequent user activities from drowning out distinct but less frequent user activities in the dataset, which is a key stage in this classification step.

For the recommendation, workflow videos are selected based on the topic modeling stage and the frequent pattern mining stage such that the videos either belong to the topic or contain the selected patterns.

## Tutorons: Generating Context-Relevant, On-Demand Explanations and Demonstrations of Online Code

In this paper, the researchers build a framework called Tutoron that is able to detect and parse explainable code in a web page, specifically for CSS selectors, regular expressions and the Unix command “wget”. And then Tutoron can generate context-relevant, on-demand natural language explanations of code. The initiative of this framework is that programmers usually turn to the web for documentation and example code to solve programming problems.

There are mainly three stages performed by Tutoron to generate the explanations of code. The first stage is the detection stage. In this stage, Tutoron will extract blocks of code from the HTML file of the webpage and divide these code blocks into candidate explainable regions. After that, Tutoron will proceed to the parsing stage so that the detected code snippets are parsed into data structures in preparation for explanation. In the final stage, also called the explanation stage, Tutoron will traverse the data structure output by the parsing stage to generate micro-explanations and demonstrations of the code. It is worth to mention that these stages are performed on a web server which will largely reduce the computational burden for the user.

By using Tutoron, programmers are able to reduce their need to access external documentations and to learn programming concepts more quickly. During the research, the researchers observe an 80% precision of Tutoron detecting code examples. However, micro-explanations are sometimes not enough for the user to solve the problem. The micro-explanations would be more comprehensive if external links can be added such as related links to Stack Overflow. In this way, the user can refer to external links when the micro-explanations is not sufficient for the problem.

**DynamicSlide: Exploring the Design Space of Reference-based Interaction Techniques for Slide-based Lecture Videos**

In this paper, the researchers develop a video processing system, named DynamicSlide, that can extract automatically extract a slide and its relevant sentences from slide-based lecture videos. The initiative of this system is that the user is able to learn materials in the slides more effectively by three major techniques: highlighting the item that is currently being explained, enabling item-based navigation, and user ability to take notes based on slide items.

There are mainly three stages performed by DynamicSlide to achieve the three major techniques mentioned above. The first stage is the shot boundary detection stage. In this stage, the system will measure the difference between two consecutive frames to extract slide shots from the video and recognize texts from the slides to extract text of each slide. In the next stage, the system will try to group the words in the slide based on the position information of the words so that they can form semantic units such as phrases and sentences. During the final stage, the system aims to find an alignment between sentences from narration script and texts identified from the previous stage. It is worth to mention that enabling item-based navigation will allow user to do a global information search so that specific information can be found more easily and that note-taking feature is much more helpful when the user wants to build their mental knowledge model. However, a limitation that is also worth to mention is that the wrongly identified slide shots and texts could be a significant cost to the user. Precision must be assured so that the overall study experience can be smooth for the user when this prototype is getting improved.

**Conclusion**

In this technical review, three research papers are reviewed and two of them proposed

software prototype that can be used to improve the study effectiveness and experience. It can be seen that these prototypes achieve reasonable precision but is still not enough for real study experience. The cost of incorrect information generation could be quite disturbing for the user when the system is considered as the major study resource. However, these prototypes also show a promising study experience that we may achieve in the future. The classic relationship between the professor and students might not be replaced as software and videos instead of classrooms which makes study more viable to common people and effective. Online education is under rapid development and it can be said that it will be the mainstream of study tool for students.

**Reference**


1. Xu Wang, Benjamin Lafreniere, Tovi Grossman (2018, April), Leveraging Community-Generated Videos and Command Logs to Classify and Recommend Software Workflows.
   **http://www.cs.cmu.edu/~xuwang/papers/workflow_chi2018.pdf**
2. Andrew Head, Codanda Appachu, Marti A. Hearst, Bjorn Hartmann (2015, Oct), Tutorons: Generating Context-Relevant, On-Demand Explanations and Demonstrations of Online Code.
   **https://andrewhead.info/assets/pdf/context-relevant-explanations.pdf**
3. Hyeungshik Jung, Hijung Valentina Shin, Juho Kim (2018 Oct), DynamicSlide: Exploring the Design Space of Reference-based Interaction Techniques for Slide-based Lecture Videos
   **https://juhokim.com/files/MM2018-Workshop-DynamicSlide.pdf**