Initialization step

> 1. Study the problem context by choosing the data you want to mine

1. Problem Description and Context Analysis

We have chosen a dataset containing information about all audio-video recordings of TED Talks uploaded to the official TED.com website until September 21st, 2017.

We want to use this dataset to start a project of data mining, finding the relationship of views, number of comments, main topics, languages and areas. In addition, we may find a popular topics trend from last ten years and predict what will be discussed at next five years.

2. Difficulties

We have found a obvious relationship between views and number of comments by doing a simple linear analysis, however it seems not easy for us to extract data from the 'tag', while we are attempting to find the relationship between main topics and views.

Besides, we need to choose one or two main tags from the array of 'tag' data, which means it requires a huge amount of work to organize and label data.

> 2. Elaborate the Use-case diagram and detailed description of the most important cases.

> 3. Define the global architecture of the Project

The goals of the application

We tend to develop an application recommending TED Talks and seeing the popular topics trend between any time period.

After register and log in our web application, users would do things as follow:

1. Choose some of the generalized topics which they are interested in.
2. Watch the recommendation from our applicaton.
3. After watching, they may give some advice like 'I like it' and 'I do not like it'. If they do not like it, they may choose a more likely label for the video helping our system to improve the dataset training.

And they can also see the trend in different areas and different time periods, which will be displayed as a histogram or bar chart.

> 4. Global architecture

We decide to develop a web application, so our system contains forward-end and backward-end.

Backward-end

Using a web framework of Python called Django to operating the data given by our data analysis program.

Forward-end

Using the web framework called React to show data gotten from the backward-end.

We could get the architecture diagram like this:

Subtitles

- Problem Description
- Context Analysis
- Difficulties
- GUI
- Use Case Diagrams
    - sequence diagram
    - class diagram
- Global architecture
- Define tasks for each student