CS 6630: Project Proposal

TED talks topic trend visualization

Hsuan Lee and Chien-Wei Sun

School of Computing, University of Utah

October 12, 2017

1 Basic Info

- Title: TED talks topic trend visualization
- Team members:
  1. First member:

     Name: Hsuan Lee

     e-mail: hsuan@cs.utah.edu

     uID: u1076070
  2. Second member:

     Name: Chien-Wei Sun

     e-mail: cwkenwaysun@gmail.com

     uID: u1141797
- GitHub: `https://github.com/cwkenwaysun/TEDmap`

2 Background and Motivation

TED is a leading organization which provides influential and understandable talk to the world. These talks cover a lot of fields, from anthropology to machine learning,

and also from biology to sociology. We are interested in the relationship between technology and world market, and we want to know if TED somehow shows the trend of popular technology or it provides a platform for topics which do not get much attention in the world.

3 Project Objectives

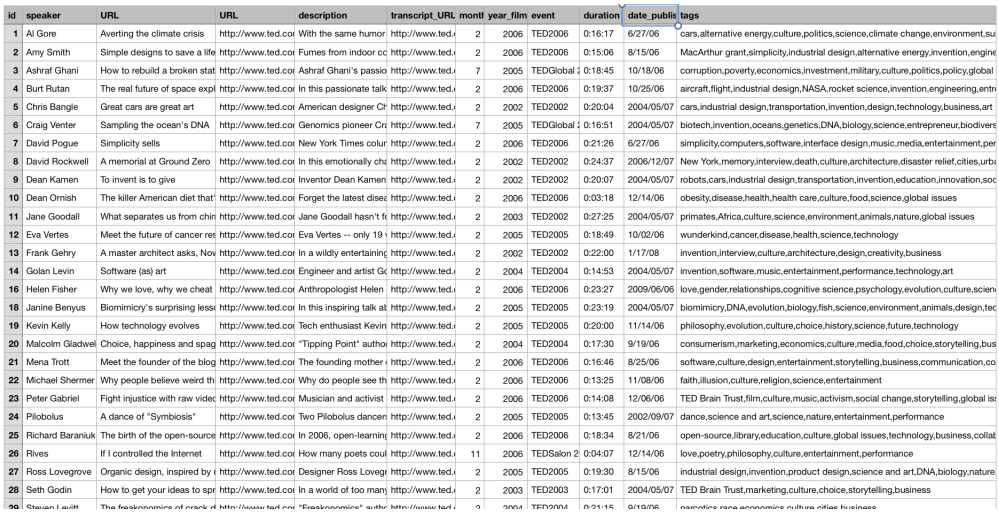
Here are questions we expect to answer at the end of this project:

- What are the trend of category tags appeared on TED talks?
- Is there any relationship between the TED talks and the big events happened in the world?
- If the above answer is yes, how much correlation do they have?
- Can we learn the trend of research on a specific field by analyzing the popularity of keywords? Or it shows the topics which people do not put attention on for now but will become important in the future?

4 Data

link: `https://www.idiap.ch/dataset/ted` These dataset include the video recording from the TED website from 1972 to 2017. For each video, its data contains the following attributes:

id	month filmed
Speaker	year filmed
headline	event
URL	duration
description	date published
transcript URL	tags



id	speaker	URL	URL	description	transcript_URL	month	year_film	event	duration	date_publis	tags
1	Al Gore	Averting the climate crisis	http://www.ted.co	With the same humor	http://www.ted.c	2	2006	TED2006	0:16:17	6/27/06	cars,alternative energy,culture,politics,science,climate change,environment,su
2	Amy Smith	Simple designs to save a life	http://www.ted.co	Fumes from indoor co	http://www.ted.c	2	2006	TED2006	0:15:06	8/15/06	MacArthur grant,simplicity,industrial design,alternative energy,invention,engine
3	Ashraf Ghani	How to rebuild a broken stat	http://www.ted.co	Ashraf Ghani's passio	http://www.ted.c	7	2005	TEDGlobal 2	0:18:45	10/18/06	corruption,poverty,economics,investment,military,culture,politics,policy,global
4	Burt Rutan	The real future of space expl	http://www.ted.co	In this passionate talk	http://www.ted.c	2	2006	TED2006	0:19:37	10/25/06	aircraft,flight,industrial design,NASA,rocket science,invention,engineering,entr
5	Chris Bangle	Great cars are great art	http://www.ted.co	American designer Ch	http://www.ted.c	2	2002	TED2002	0:20:04	2004/05/07	cars,industrial design,transportation,invention,design,technology,business,art
6	Craig Venter	Sampling the ocean's DNA	http://www.ted.co	Genomics pioneer Cra	http://www.ted.c	7	2005	TEDGlobal 2	0:16:51	2004/05/07	biotech,invention,oceans,genetics,DNA,biology,science,entrepreneur,biodivers
7	David Pogue	Simplicity sells	http://www.ted.co	New York Times colur	http://www.ted.c	2	2006	TED2006	0:21:26	6/27/06	simplicity,computers,software,interface design,music,media,entertainment,per
8	David Rockwell	A memorial at Ground Zero	http://www.ted.co	In this emotionally cha	http://www.ted.c	2	2002	TED2002	0:24:37	2006/12/07	New York,memory,interview,death,culture,architecture,disaster relief,cities,urba
9	Dean Kamen	To invent is to give	http://www.ted.co	Inventor Dean Kamen	http://www.ted.c	2	2002	TED2002	0:20:07	2004/05/07	robots,cars,industrial design,transportation,invention,education,innovation,soc
10	Dean Ornish	The killer American diet that'	http://www.ted.co	Forget the latest diseа	http://www.ted.c	2	2006	TED2006	0:03:18	12/14/06	obesity,disease,health,health care,culture,food,science,global issues
11	Jane Goodall	What separates us from chin	http://www.ted.co	Jane Goodall hasn't fo	http://www.ted.c	2	2003	TED2002	0:27:25	2004/05/07	primates,Africa,culture,science,environment,animals,nature,global issues
12	Eva Vertes	Meet the future of cancer res	http://www.ted.co	Eva Vertes -- only 19 y	http://www.ted.c	2	2005	TED2005	0:18:49	10/02/06	wunderkind,cancer,disease,health,science,technology
13	Frank Gehry	A master architect asks, Nov	http://www.ted.co	In a wildly entertaining	http://www.ted.c	2	2002	TED2002	0:22:00	1/17/08	invention,interview,culture,architecture,design,creativity,business
14	Golan Levin	Software (as) art	http://www.ted.co	Engineer and artist Go	http://www.ted.c	2	2004	TED2004	0:14:53	2004/05/07	invention,software,music,entertainment,performance,technology,art
16	Helen Fisher	Why we love, why we cheat	http://www.ted.co	Anthropologist Helen	http://www.ted.c	2	2006	TED2006	0:23:27	2009/06/06	love,gender,relationships,cognitive science,psychology,evolution,culture,scien
18	Janine Benyus	Biomimicry's surprising lesso	http://www.ted.co	In this inspiring talk at	http://www.ted.c	2	2005	TED2005	0:23:19	2004/05/07	biomimicry,DNA,evolution,biology,fish,science,environment,animals,design,tec
19	Kevin Kelly	How technology evolves	http://www.ted.co	Tech enthusiast Kevin	http://www.ted.c	2	2005	TED2005	0:20:00	11/14/06	philosophy,evolution,culture,choice,history,science,future,technology
20	Malcolm Gladwel	Choice, happiness and spag	http://www.ted.co	"Tipping Point" author	http://www.ted.c	2	2004	TED2004	0:17:30	9/19/06	consumerism,marketing,economics,culture,media,food,choice,storytelling,bus
21	Mena Trott	Meet the founder of the blog	http://www.ted.co	The founding mother o	http://www.ted.c	2	2006	TED2006	0:16:46	8/25/06	software,culture,design,entertainment,storytelling,business,communication,co
22	Michael Shermer	Why people believe weird th	http://www.ted.co	Why do people see th	http://www.ted.c	2	2006	TED2006	0:13:25	11/08/06	faith,illusion,culture,religion,science,entertainment
23	Peter Gabriel	Fight injustice with raw videc	http://www.ted.co	Musician and activist	http://www.ted.c	2	2006	TED2006	0:14:08	12/06/06	TED Brain Trust,film,culture,music,activism,social change,storytelling,global is
24	Pilobolus	A dance of "Symbiosis"	http://www.ted.co	Two Pilobolus dancers	http://www.ted.c	2	2005	TED2005	0:13:45	2002/09/07	dance,science and art,science,nature,entertainment,performance
25	Richard Baraniuk	The birth of the open-source	http://www.ted.co	In 2006, open-learning	http://www.ted.c	2	2006	TED2006	0:18:34	8/21/06	open-source,library,education,culture,global issues,technology,business,collab
26	Rives	If I controlled the Internet	http://www.ted.co	How many poets coul	http://www.ted.c	11	2006	TEDSalon 2	0:04:07	12/14/06	love,poetry,philosophy,culture,entertainment,performance
27	Ross Lovegrove	Organic design, inspired by r	http://www.ted.co	Designer Ross Lovegr	http://www.ted.c	2	2005	TED2005	0:19:30	8/15/06	industrial design,invention,product design,science and art,DNA,biology,nature
28	Seth Godin	How to get your ideas to spr	http://www.ted.co	In a world of too many	http://www.ted.c	2	2003	TED2003	0:17:01	2004/05/07	TED Brain Trust,marketing,culture,choice,storytelling,business
29	Steven Levitt	The freakonomics of crack d	http://www.ted.co	"Freakonomics" autho	http://www.ted.c	2	2004	TED2004	0:21:15	9/19/06	narcotics,race,economics,culture,cities,business

Figure 1: Data get from idiap.ch

5 Data Processing

To better understand the impact of TED videos, we develop web crawlers to collect attributes like **rates**(how do people feel after watching a video), **views**(how many time a video has been played), and some potentially valuable data like datetime, redirected urls, and transcripts. We choose to learn and use **Scrapy** (`https://scrapy.org/`) as our crawler.

Furthermore, in order to load data with ease, we transfer our data from csv file to json form. We found this preprocessing can be accomplished painlessly by using **Pandas** (`http://pandas.pydata.org/`) toolkit.

In our research, we plan to visualize the data according to the tags/keywords of the video. It is not efficient to search all the data to find which videos are related with one specific tag on javascript. For practical implementation, we will preprocess the dataset based on tags, which means to use tag as key to create input data.

6 Visualization Design

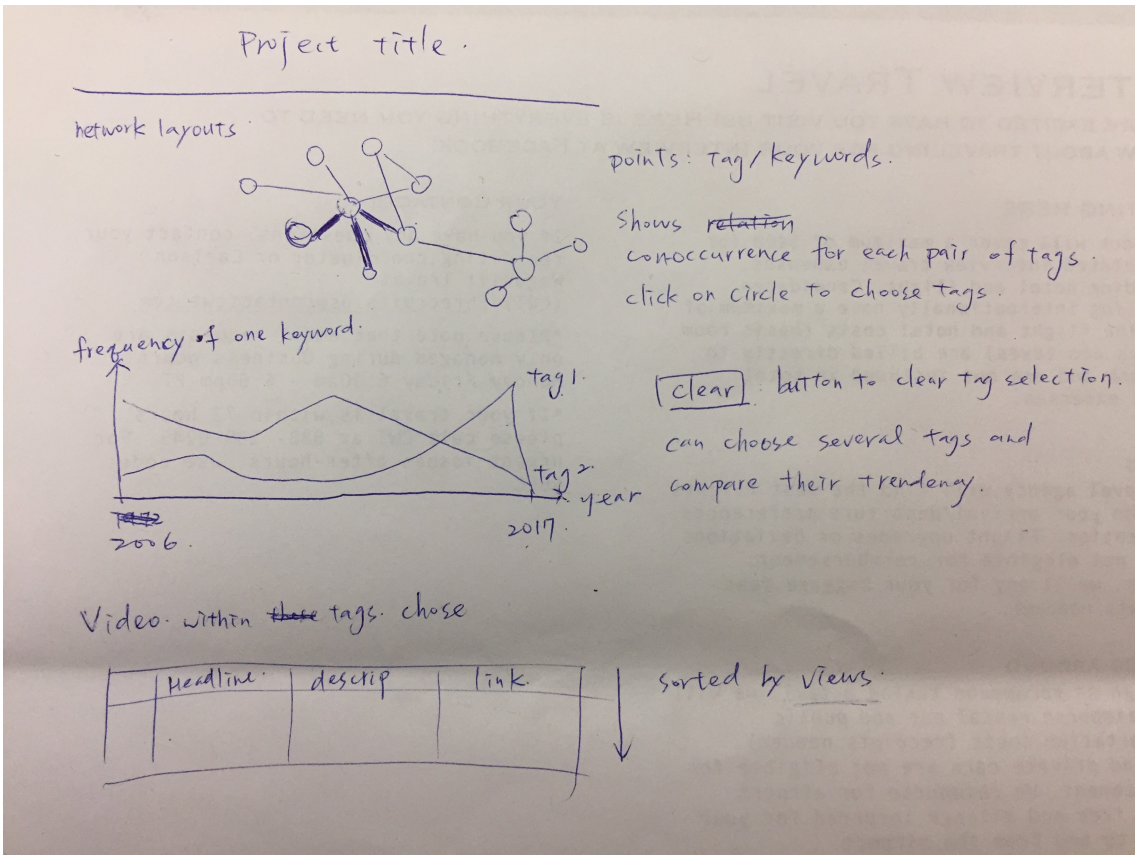


Figure 2: Category of data, and it relationship with the layout

Our design is based on the network layout, as shown in Figure 2. This network is composed of tags, and users can choose several tags they are interested in to through interaction with the network node. Next, the line chart in the middle of Figure 2 would display the tendency of chosen tags versus time/year. The last part helps users to search for TED talks including these tags. Users can see the result is sorted by views or popularity.

We also want to compare the statistic of the tags between years, so we design a bar chart as shown in Figure 3. By making use of the sliding bar on the top, the statistics of two years is displayed. Figure 4 helps us to figure out what attributes are needed in each chart. It also shows the relationship of charts.

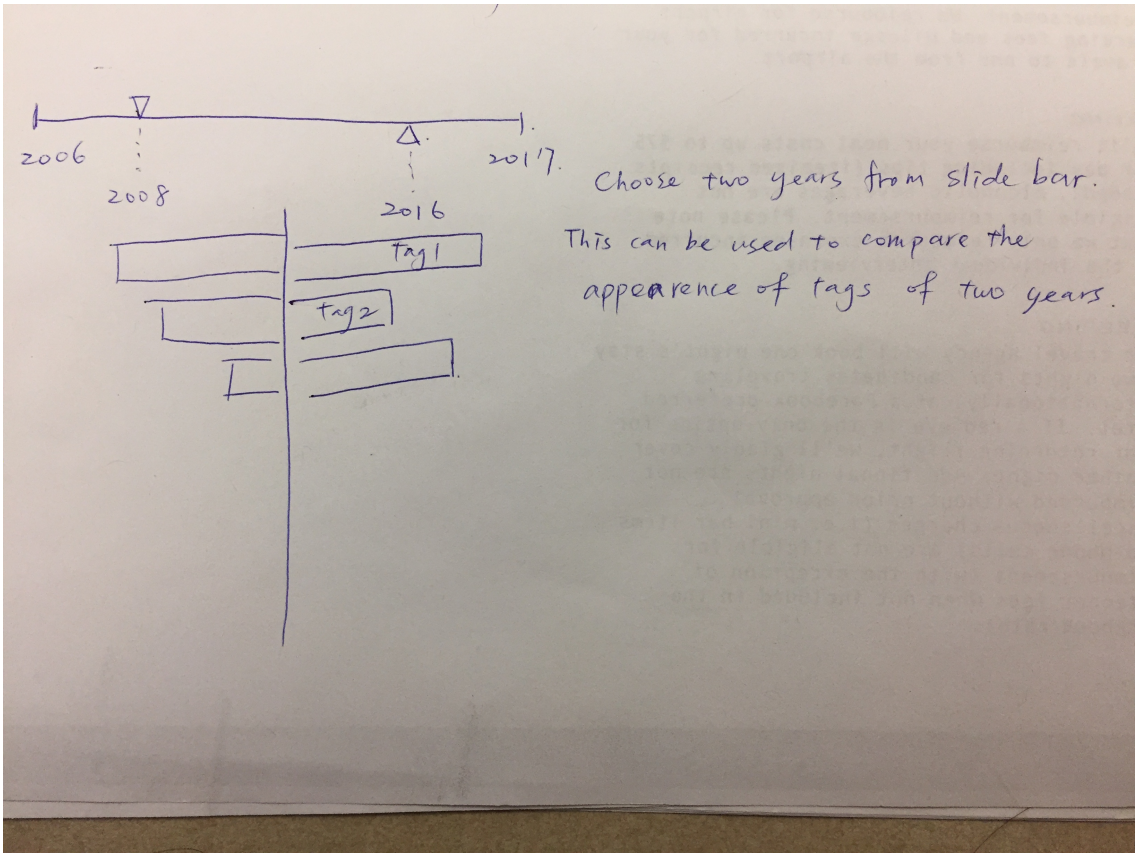


Figure 3: Design of Optional features

7 Must-Have Features

- A network layout shows the relationship of tags. If a pair of tags usually appear simutaneously, we draw their links with larger thickness.

- A line chart which can be used to compare the trend of several tags.

- A table present the videos which contain the tags user choose.

- Semantic zoom: double-click on a tag in network layout to zoom in and show tag name in circle.

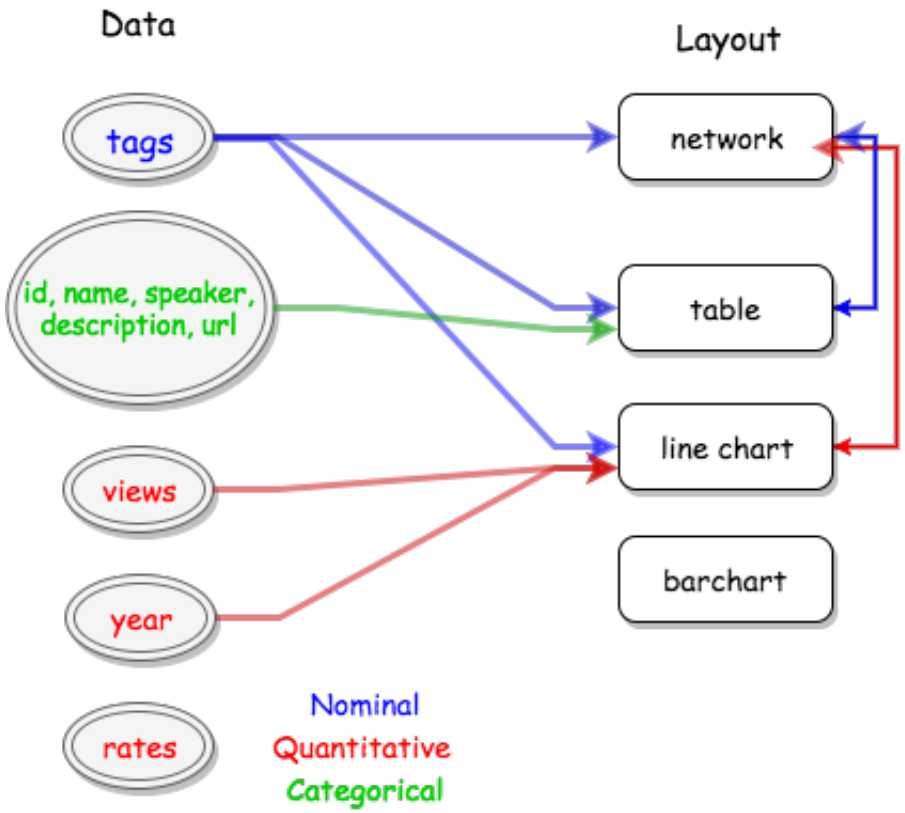


Figure 4: Category of data, and it relationship with the layout

8 Optional Features

- A bar chart tell us the frequencies of tags in one year. We can also choose two years and compare their tags.
- By choosing item from drop-down list, user can view the frequencies of tags on top 10/50/100 viewed/liked talks.
- We have already learned many cool new features of JS ES6 in this class. For further exploration on ES6, we would like to implement some of patterns in this video[1] and also implement one or two design patterns for out own edification.

9 Project Schedule

Week	work distribution	deadline
10/16	1. Layout prototype of webpages complete. 2. Technical investigate 3. Rre-processed data complete	
10/23	1. Network chart and categorize tags complete 2. Test case complete	10/27 Project Proposal Due
10/30	1. Preliminary design review: architectural design satisfies all requirements	
11/6	1. Critical design review 1 2. Milestone report complete	11/10 Project Milestone
11/13	1. Critical design review 2	
11/20	1. Final design review 2. Report outline complete	
11/27	1. Test case review 2. Demo video 3. Report complete	12/1 Final project due

References

[1] Allen, S. `https://www.youtube.com/watch?v=hO7mzO83N1Q`, javascript patterns for 2017,.