



# VISIT THE DOME

Thang Nguyen

Ruta Bakhda

- Task 1: Adjust annotation interface to separate the stance annotation from the sentences of the speech
- Task 2: Create “debate” dataset



# TASK 1

## ○ Previous version

Income Tax By Dan McTeague (7 segments remaining)

Label of Paragraph ?

I recently attended an anti-tax rally in my riding where 2,000 people told budget. ? Canadians have a right to be heard and the Minister of Finance has made it possible through an open the Reform Party, in concert with top officials in organizations such as the Canadian Taxpayers' Federation.

## ○ Up-to-date version

Income Tax By Dan McTeague (7 segments remaining)

Label of Paragraph ?

I recently attended an anti-tax rally in my riding where 2,000 people told the government they do not want to be heard and the Minister of Finance has made it possible through an open and full budgetary process. ?



## TASK 2

- What is “debate” dataset

```
emergency debate  
commons debates after recess  
continuation of debate on address in reply  
supply-procedure resuming on thursday or friday debate on amendment  
subject matter of questions to be debated  
supply-official reporting of debates.  
statement in debate on defence estimates respecting royal military college  
address debate
```



## TASK 2

- The results is placed in `/mnt/nfs/webis20/data-in-progress/wstud-visit-the-dome-ss19/lipad_results`

Name	Date modified	Type	Si
address-debate	24/06/2019 11:13	File folder	
canada-united	24/06/2019 11:19	File folder	
commons-debates-after-recess	24/06/2019 10:31	File folder	
continuation-of-debate	24/06/2019 10:39	File folder	
emergency-debate	23/06/2019 03:23	File folder	
house-of-commons-debates	24/06/2019 11:26	File folder	
oral-question	23/06/2019 08:01	File folder	
other-debate	24/06/2019 11:42	File folder	
statement-by-members	23/06/2019 08:01	File folder	
statement-in-debate	24/06/2019 11:06	File folder	
subject-matter	24/06/2019 10:53	File folder	
supply-official	24/06/2019 11:00	File folder	
supply-proc	24/06/2019 10:46	File folder	



## TASK 3 : ANNOTATION AGREEMENT

- Task : Sample statement by members annotation task
- Sentences : 109
- Annotators : 6
- Cohen's kappa agreement : **0.3**
- $<0$  as indicating no agreement
- 0.00–0.20 as slight,
- **0.21–0.40 as fair,**
- 0.41–0.60 as moderate,
- 0.61–0.80 as substantial,
- and 0.81–1 as almost perfect agreement.



# TASK 4 : NEW ANNOTATION TASKS

- Change in task generation :
  - “Basepk “ folder in WAT-SI task
  - Individual speech generation in csv based in Basepk
- Change in task assignment :
  - Bug :
  - Annotator 1 : **Task 1**, Task 2
  - Annotator 2 : Task 2, task 3, Task 4
  - Annotator 3 : **Task 1**, task 3, task 4
  - Each user has been assigned 3 tasks.
  - Task 1 has been assigned 2 times, we need to assign it to one more user.
  - Annotator 1 has an empty slot but it has been assigned task 1 already.



# SOLUTION

- Counter number of loops it takes to assign task
- If it is more than 3,
  - Get list of unassigned task
  - Get list of annotators with empty slots
  - Get list of annotators with completely filled slots
  - Pick randomly annotator with filled slot
    - Check if it contains unassigned task
    - If not, pick one task from the list of filled slots
    - Switch it with unassigned task





# TASK SET UP

- 16 controversial topics selected
  - Abortion
  - Status of Women
  - Human rights
  - Criminal code
  - Immigration
  - Body Shaming

Almost 100-150 sentences per annotator



# CHANGING LABELS FOR SPEECHES

- Old labels :
  - Conclusion
  - Non conclusion
- New labels:
  - Conclusion
  - Premise
  - Non Argumentative



# CURATION TASK FOR SAMPLE ANNOTATION

- Curation task was set up to find labels with majority vote
  - Total : 109 segments
  - Tie : 30 sentences
  - Manual label assignment for tie.
- 
- Generating csv of separated sentences with curation results
  - Above file will be input in Args.

