

# Multi-Doc Summarization

## COVID-19

Sharon Wang  
Week 2

# Basic Strategy for Summarization

Any sentence that contains either the **query** or any of the CatE generated **phrases** is put into the summary.

- Eliminate sentences that contain vague times, such as days of the week, “last weekend,” “a month ago”, etc.
- Eliminate questions.

| <b>Economy</b> | <b>Death</b> | <b>Transmission</b> |
|----------------|--------------|---------------------|
| stimulate      | fatalities   | community           |
| industries     | recorded     | spread              |
| restart        | count        | exposure            |
| rebound        | 1000-case    | evidence            |
| protectionism  | 1523         | human-to-human      |
| third-largest  | 3869         | person-to-person    |
| sectors        | toll         | clusters            |
| economic       | died         | spreading           |
| economies      | deaths       | exposures           |
|                | fatality     |                     |

precision: 15/20  
= 0.75

precision: 10/37  
= 0.37

precision:  
19/41 = 0.46

recall:  
15/17 = 0.88

recall:  
10/30 = 0.33

recall:  
19/35 = 0.54

# Post-processing on phrases (Clean-up + Ranking)

1. Remove words that have the same **stem** as the query word
2. Remove words in the phrases that have the same stems
3. Given a query word  $q$ , let  $v(q)$  be its word2vec representation. Let  $p_1, p_2, \dots, p_n$  be the phrases generated. Compute  $\cos(v(p_i), v(q))$ . Remove  $v(p_i)$  if  $\cos(v(p_i), v(q)) < k$  for some  $k$  ( $0 \leq k \leq 1$ ).
4. Rank the phrases according to  $\cos(v(p_i), v(q))$ .  $w_i$  is more relevant to  $q$  than  $p_j$  is if  $\cos(v(p_i), v(q)) > \cos(v(p_j), v(q))$ .

# Strategy for Summarization

- After the basic summarization described in Slide 2, use cosine similarity to further narrow down the important sentences.
- For each sentence  $S$  in the corpus that contains either  $q$  or  $p_i$ , compute  $\sum \cos(v(w_i), v(q)) / |S|$  where  $S$  contains  $w_1, w_2, \dots, w_n$ .
- Only add sentence  $S$  to the summary if the summation is greater than a certain value.