Input: Corpus of segmented articles covering a key event

Episode Indicative Term Mining (4.1)

Identify salient terms which discriminatively co-occur:

Terms (a,b) frequently co-occur together and infrequently with other terms

(barrier, broke)
(barrier, glass panels)
...

(spray-paint, vandalizing)
(portraits, podium)
...

Episode Segment-level Partitioning (4.2)

1. Hong Kong protesters **broke** through **barrier**...

2. Many of the **glass panels** [...] smashed.

3. Protesters began **spray-painting slogans** and **vandalizing** the **portraits**.

4. They **draped** the **flag** of colonial ~~Hong Kong~~ at the **podium**.

✓ merge

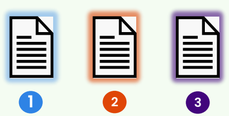
✗ partition

✓ merge

Do terms **discriminatively co-occur?**

Semantically similar?

LLM-Enhanced Episode Estimation (4.3)



Rank articles by partition **quantity** & **quality**

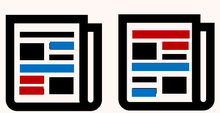
Episode 1: broke, glass panels, ...
Episode 2: slogans, draped, ...

Cluster partitions across top articles to form episodes

Summarize core attributes of each **episode**

Episode-Segment Classification (4.4)

Episode 1: Protesters stormed the Legislative Council
Episode 2: Protesters vandalized the Legislative Chamber



Map confident segments to each episode to form **final episode clusters**

Output: Episode clusters containing relevant article segments