Input: Corpus of segmented articles covering a key event

Salient Discriminative Co-occurrence Detection (4.1)

Salient Terms → **Synonym Expansion** → **Discriminative Co-Occurrence (DC)**

protesters --- demonstrators...

broke

vandalizing ----- ransacked...

Hong Kong

flag

broke

vandalizing, ransacked...

flag

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

Episode 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**.

SIMILARITY

✓ semantic

✓ DC

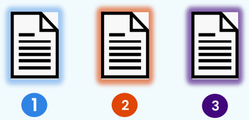
✗ semantic

✗ DC

✓ semantic

✓ DC

Candidate Episode Estimation (4.3)



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

Episode 1: broke, glass panels, ...

Episode 2: slogans, draped, ...

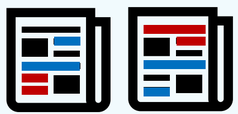
Cluster partions across top articles to form episodes

LLM-guided **episode fluency**

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