Why Text Clustering?

- In general, very useful for text mining and <u>exploratory</u> text analysis:
  - ➔ Get a sense about the overall content of a collection (e.g., what are some of the "typical"/representative documents in a collection?)
  - ➔ Link (similar) text objects (e.g., removing duplicated content)
  - ➔ Create a structure on the text data (e.g., for browsing)
  - ➔ As a way to induce additional features (i.e., clusters) for classification of text objects
- Examples of applications
  - – Clustering of search results
  - – Understanding major complaints in emails from customers