Section Data Example in the Test Set

Query: How does the qualitative evaluation of extractive summarizers using word clouds elucidate the differences in content focus between the original documents and the summaries?

Text:
Here we use word cloud representations to give an intuitive interpretation of the content in the generated extractive summarizers.We create word clouds for the two best methods in Section 5.3.In this paper, we used an online tool called WordItOut5 to generate the word cloud representations.In all word clouds presented in this paper, a filter is used to display only the words with minimum frequency of 2.

Figure 3 shows a word cloud made by the aggregation of all the summaries generated by the PKUSUMSUM-Centroid method.This gives a sense of the content in those summaries.For contrast, we also generate a word cloud for the original news articles without the content of the generated summaries.Specifically, common words are first removed completely and then the word clouds are built with frequencies of surviving words.In essence, this shows what information remains apart from the generated summaries.

The images clearly show a contrast of content.The summary wordcloud shows “earthquake” as its most prominent word.The image of the articles show less focus.If viewed alone, the reader would not quickly infer the gist of the original content.Similarly, Figure 4 represents “Lead” method.And here we see an even more stark difference.6. Comparing Twitter and News Media information.

Figure: Figure 3

Caption: Figure 3: The word clouds representing summaries generated by PKUSUMSUM-Centroid method (left) and original documents without the content of those summaries (right).

Answer 1: Word clouds highlight the prominent themes in summaries versus original texts by displaying relative word frequencies visually.
Answer 2: Visual contrast in word frequency highlights content focus differences.