

Guidance for Human Evaluators

File format

The attached excel file contains:

1. url: you can access the news article by clicking the url in the first column. Useful for understanding the key ideas.
2. reference: news highlights written by professional editors. They serve as a golden standard for evaluation.
3. Outputs of four models, A,B,C,D: sentences extracted as summary by the models. The sentences are separated by "".

Evaluation

1. Informative ranking: whether the summary includes key points of the news article.
2. Coherence **ranking**: please compare the four outputs, and assign rank 1 to the most **coherent** summary, rank 2 to the second best, rank 3 to the third best and rank 4 to the worst. We especially care about the coherence between adjacent sentences. If two adjacent sentences are unlikely to appear together, they are considered as not coherent. The more non-coherent sentence pairs are, the less coherent the summary is. Coherence is *irrelevant* to length of summary.

o Examples:

- Not coherent, because it looks like there should be another sentence before the first sentence

" McGuire , 25 , said .</s>She and Rogers , 34 , own the Breckenridge Cannabis Club , a recreational marijuana dispensary in the historic and scenic ski town of Breckenridge , Colorado .</s>The two said they brought in more than \$ 47,000 in sales , roughly 30 times their normal daily sales of medical marijuana .

- Not coherent, because we don't know what "it" refers to. It applies to other pronouns as well.

It comes in a yellow - labeled bottle with a fire - breathing demon on it .</s>It tastes like Big Red chewing gum .</s>It 's Fireball Cinnamon Whisky , and lately it 's been as hot as its name .</s>" Fireball is number one , definitely .</s>On an average night , we probably go through three or four bottles . "</s>Fireball , which did n't exist in its current form a decade ago , is the fastest - growing big brand of liquor in America .

- Coherent, "(CNN)" or other "(XXXX)" like headings indicates this is the first sentence of the news. We can understand who "He" is referring to, and the logic flow is natural.

(CNN) James Best , best known for his portrayal of bumbling sheriff Rosco P. Coltrane on TV 's " The Dukes of Hazzard , " died Monday after a brief illness .</s>He was 88 .</s>Best died in hospice in Hickory , North Carolina , of complications from pneumonia , said Steve Latshaw , a longtime friend and Hollywood colleague .

3. Overall ranking: rank according to how satisfied you are with the summary.

Notice:

1. You could give the same rankings to multiple models. For examples, if model B and C's output are identical, you could give rankings like: A=3, B=1, C=1, D=2. If you believe two outputs have the same quality although they are different, you can also give them the same rankings.
2. The sentences are segmented with nltk package, which may not starts or ends properly. "" only serves as a sentence separator, which do not have any semantic meaning.