News Track

Summary

The track features modern search tasks in the news domain. As news is getting democratized, understanding the context and relevance of information becomes a harder task for the consumer. The task aims at building new information access tools to help the user understand the context of a story.

Task

There are two tasks for News Track 2020.

1. <u>Background Linking</u> - To retrieve news articles that provide important context or background information, given a story. The goal is to develop systems that can help users contextualize news articles as they are reading them.
2. <u>Wikification</u> - Given a news story, identify short passages in the article that should be hyperlinked to either another article, or a Wikipedia article, in order to provide in-context access to information that would help contextualize or fill in background on the story being read.

Data

The data for the track is the TREC Washington Post Collection which is five years of articles, 2012 - 2017.(https://trec.nist.gov/data/wapost/) The files are "JSON-lines" format. The articles are broken into content paragraphs, with interspersed media such as images and videos referenced by URL. The track in 2018 and 2019 used v2 of the collection, which had exact duplicates removed. For 2020, v3 collection will be used which removes near-duplicates as well.

Evaluation of Runs

Both tasks use the same evaluation metrics. The primary metric is nDCG@5, with the gain value $2^{(r-1)}$ where r is the relevance level from the scale above, and the zero relevance level contributes no gain. Evaluation will use trec_eval so all traditional TREC measures will also be reported to a measurement depth of 100.

Approach to the Task

For the background linking task pseudo relevance feedback can be used to get Named Entities(NEs) from a set of documents. Also, Topic Modelling can be used to get the keywords and retrieval of the news articles based on them. For the Entity Ranking task, BM25 ranking method can be used.