TREC 2020 News Track

Guidelines v2.1, 27 May 2020

Track coordinators: Ian Soboroff, Shudong Huang, Donna Harman (NIST)
Google group: https://groups.google.com/forum/#!forum/trec-news-track
Slack: trectalk.slack.com#news-2020 (restricted to active participants)
Register to participate at https://trec.nist.gov/pubs/call2020.html

NOTE: due to COVID-19, topic development for the track has been very delayed. The schedule will be updated when topics are released. I am aiming for a release by June 15 but it's not yet definite. I apologize to those of you who planned for an early track this year.

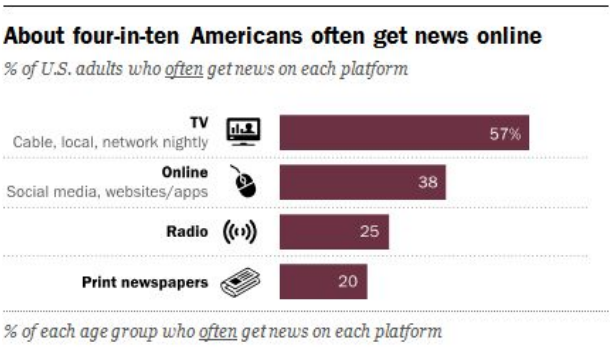
Changelog

Here are the big changes for 2020:

- The schedule is moving back into Spring semester: topics will be released in ~~mid-April~~ mid-May, and runs will be due ~~at the end of May~~ a month later.
- The entity ranking task has ended, and is being replaced by a new Wikification task. (The background linking task remains unchanged.)
- There will be 50 new topics for this year, developed for the track. The topics and relevance judgments from last year are available for tuning.
- Based on the 2019 analysis of duplicate documents in the collection, we are releasing v3 of the Washington Post collection with duplicates removed.

Motivation

While news and newswire has been a common genre in IR experimentation for a very long time, the evaluation tasks in IR have rarely if ever supported the "news user" -- a consumer of news that is not an analyst. According to a Pew Research study in 2016, roughly 38% of Americans get their news online, with the fraction increasing for younger consumers.



Moreover, since online delivery of news has shifted the focus away from the provider or publisher towards the story, news production has been dramatically democratized. If everyone can produce professional looking news, then understanding the context and background of information becomes a harder task for the consumer.

This track is not about detecting "fake news". Rather, we are envisioning new information access tools that help the user understand the context of a story, wherever they are reading it. In conjunction with The Washington Post, we are developing tasks around how news is presented on the web and thinking about how to enhance that learning experience. The larger question is, what roles can IR play in this new, noisy, adversarial online news domain?

Data

The data for the track is the TREC Washington Post Collection, first released in 2018. This is five years of articles, 2012 - 2017, and you can get it at no cost from NIST after completing the requisite usage agreement. The data URL is https://trec.nist.gov/data/wapost/

The files are "JSON-lines" format, that is, each document is a single long line of JSON. The articles are broken into content paragraphs, with interspersed media such as images and videos referenced by URL. Those URLs point back to the Washington Post website and should persist at those URLs for the foreseeable future.

The collection as originally released contained numerous duplicate and near-duplicate documents. The track in 2018 and 2019 used v2 of the collection, which had exact duplciates removed. For 2020, the collection will be updated to v3 which removes near-duplicates as well.

Task 1: Background Linking

The goal background linking task is to develop systems that can help users contextualize news articles as they are reading them. For example, news websites nearly always link to related articles in a sidebar, at the end of an article, from within the text of the article, or all three. We want to look at a particular case for linking: given that the user is reading a specific article (the query article), recommend articles that this person should read next that are the most useful for providing **context and background** for the query article.

Note that links already present in the Washington Post article collection are **not** training data for this task. In our conversations with the Post, their current practice is largely driven by the author of the article and does not follow any fixed guidelines or goal. Hence, we are designing this task as a specific kind of news recommendation task that would be useful in any news reading context, including the Post's website.

From our conversations with Post journalists about linking for background and context, every author has their own guidelines in their head, but three common rules emerged:

1. No wire service articles. (That is, from Associated Press (AP), AFP, etc)
2. No opinion or editorials.
3. The list of links should be diverse.

The assessors will judge wire service articles as **not relevant**. For (2), we decree that articles from the "Opinion", "Letters to the Editor", or "The Post's View" sections, as labeled in the "kicker" field, are **not relevant**. (3) is complicated, as seen in the TREC Web Diversity track and the NTCIR Intent Mining tasks, and we are leaving this aside to a future iteration of the track.

Results will be pooled and judged by NIST assessors on the following scale:

0. The linked document provides little or no useful background information.
1. The linked document provides some useful background or contextual information that would help the user understand the broader story context of the query article.
2. The document provides significantly useful background ...
3. The document provides essential useful background ...
4. The document MUST appear in the sidebar otherwise critical context is missing.

Input

There will be 50 new topics for 2020, developed on the Washington Post collection. The topics will mimic the standard TREC topic format:

```
<top>
<num>Number: xxx </num>
<docid>f30b7db4-cc51-11e6-a747-d03044780a02</docid>
<url>https://www.washingtonpost.com/local/public-safety/homicides-remain-st
eady-in-the-washington-region/2016/12/31/f30b7db4-cc51-11e6-a747-d03044780a
02_story.html</url>
</top>
```

"Docid" references the "id" field in the Washington Post corpus documents. "Url" references the "article_url" field in the documents. Both indicate the query article.

Output

Submissions should be standard TREC format, that is, trec_eval results file format:

```
1 Q0 2707e25a-cfaf-11e6-a87f-b917067331bb 1 37.5 myrun
1 Q0 513673ee-d003-11e6-b8a2-8c2a61b0436f 2 33.2 myrun
...
1 Q0 f8ded480-cdef-11e6-b8a2-8c2a61b0436f 99 0.5 myrun
2 Q0 350e3d74-cf94-11e6-a87f-b917067331bb 1 55.2 myrun
...
```

Systems may retrieve up to 100 documents per topic. The first field is the topic id ("<num>" in the topic), the second field is a literal "Q0", the third field is the document ID of the linked document, the fourth field is the rank (ignored), the fifth field is the score, and the sixth field is the runtag. Note that trec_eval sorts by descending score and breaks ties using document IDs.

Metrics

The primary metric for the background linking task will be nDCG@5, where the gain value is 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.

Example

Query article: Love in the time of climate change: Grizzlies and polar bears are now mating (May 23, 2016)

This article describes and analyzes a phenomenon where grizzlies and polar bears are mating to create a new species known as pizzlies or grolars. It explains why this is happening and points out that it happens (or has happened) to other species as well. Articles along these lines are good background links. For example:

- Coywolves, coyote-wolf hybrids, are prowling Rock Creek Park and D.C. suburbs (July 1, 2014)
- Humans and Neanderthals may have interbred 50,000 years earlier than previously thought (February 17, 2016)

However, the following article is of less relevance and should be ranked lower because it's not about interbreeding.

- Why do seals keep trying to have sex with penguins? (November 18, 2014)

Task 2: Wikification

In addition to providing links to articles that give the reader background or contextual information, journalists sometimes link mentions of concepts, artifacts, entities etc to internal or external pages with in depth information that will help the reader better understand the article. In 2018 and 2019, we explored this idea with an *entity ranking task*: given automatically-identified named entities in the article, rank them in order of utility to a user similarly to the background linking task. In 2019, we concluded that this task was quite simple: news articles are judicious in the entities they name, and their occurrence order in the article is a very good proxy for good performance in the task as formulated.

For 2020, we are changing Task 2 to *Wikification*. Wikification is to the hyperlinking of entities, concepts, or references to another resource that provides more information on the linked thing,

in the spirit of linking in Wikipedia. In Post articles, authors frequently allude to other Post articles, news articles from other sources, or external entities, which they link directly into the prose text of their articles. This task will explore whether systems can automatically wikify the Post source article usefully for the news reader, who as in Task 1 we envision as seeking background or contextual information on what they are reading. While for background linking, the context is the entire article, for wikification, the context is a short, explicit extent in the article that will be decorated with a hyperlink.

The input for this task will be the same source article as in the Background Linking task. Systems will provide output in the following line-oriented format:
`tid score cnum start len link`

Where

- `Tid` is the topic ID,
- `Score` is a numeric score for ranking,
- `Cnum` is the sequence number of the content block, starting from zero, of the linked passage,
- `Start` is the character offset in the content block, starting from zero, for the anchor text start,
- `Len` is the length of the anchor text, in characters. Len must be greater than zero.
- `Link` is the target of the hyperlink being proposed by the system.

Systems may link either to Washington Post articles in the collection, or to Wikipedia articles in the Wikipedia dump provided. A Post article link will be in the format `post_v3:docid`, and the semantics of a Post article link is that the anchor text is referring to or alluding to something covered or explained in the linked article, which would be useful to the reader in understanding or contextualizing the source article. A Wikipedia article link will be in the format `enwiki:name`, following the identifiers in the Wikipedia dump. The semantics of a Wikipedia link are as in the original entity ranking task: the linked article provides useful background or context that would benefit the reader in understanding the source article.

Ranked links will be pooled and judged by NIST assessors as to whether the link is correct, and useful to the reader, on the following scale:

-2. The link target is incorrect or makes no sense to the reader.
0. The link target makes sense but provides little or no useful background information.
1. The linktarget provides some useful background or contextual information that would help the user understand the broader story context of the query article.
2. The link provides significantly useful background ...
3. The entity link provides essential useful background ...
4. The entity link MUST appear in the sidebar otherwise critical context is missing.

Caveat

This is a new evaluation task, and we fully anticipate problems. For example, different systems will likely return the same links with different anchors, and different systems will likely return nearly-overlapping anchors pointing to different links. Year one of this task will give us the data to study this problem.

Metrics

The primary metric for this task will be nDCG@5, where the gain value is 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.

Rules

Manual or automatic runs

When you submit your runs, you will be asked to indicate if the run is manual or automatic. An **automatic** run involves no manual intervention, but runs fully automatically from the topic file. In contrast, **manual** runs can involve human intervention – including manual query formulation, manual relevance feedback, and reweighting/reranking by hand.

Duplicate document handling

We have cleaned the document collection to remove exact and near-duplicate documents. Following adhoc practice, we will judge any remaining duplicate documents independently.

External resources

We will provide a Wikipedia dump from August 20, 2017, which coincides with the end of the epoch of the Washington Post collection. This dump will be formatted identically to dumps used in the CAR track. Wikipedia links in Task 2 will be with respect to this collection.

Past relevance judgments

Last year's topics and their assessments are available from the TREC data page (https://trec.nist.gov/data/news2018.html and https://trec.nist.gov/data/news2019.html). You can use them in tuning your systems.

Key dates

Guidelines released: January 2020

Topics released: TBA, targeting June 15
Runs due: TBA
TREC: November 18 - 20, 2020