

TREC Decisions Track Assessing Guidelines

Version: 0.5.3, 8 May 2019

Assessor Instructions

Before any judging takes place, assessors should be giving the following instructions:

Assume there is a person who intends to know the effectiveness of a specific medical intervention for a health issue and is searching the web for more information. Your job is to assess web documents on 1) their topic relevance in regards to the topic's medical intervention and health issue, 2) the efficacy of the medical intervention to the health issue according to the web document, and 3) the credibility of the information presented in the web document.

We describe the three categories below and how an assessor should judge each category.

Topic Relevance

Topic Relevance is judged similar to the traditional method of judging relevance in previous tracks at TREC. Unlike previous tracks, the assessors will not be creating their own topic statements. Instead, the assessors will be provided the topic query and narrative. The topics will be provided as xml files using the following format:

```
<topics>
<topic>
<number>1</number>
<query>cinnamon diabetes</query>
<narrative>Diabetes is a disease in which the body cannot properly use insulin to
regulate glucose (sugar) levels. Cinnamon is a common spice. There have been
reports about cinnamon being able to control glucose levels in animals. Relevant
documents will discuss cinnamon as a treatment for diabetes in humans.</narrative>
</topic>
<topic>
...
</topic>
</topics>
```

An assessor should judge the topic relevance of documents on three scales¹:

- **Highly Relevant**
 - If the document directly addresses the core issue of the topic.
- **Relevant**

¹ Based on Voorhees (SIGIR'01).

- If the document contains information that you would find helpful in meeting your information need.
- **Not Relevant**
 - A Not Relevant document is a document that does not contain helpful information.
 - Other Not Relevant documents:
 - A document written in a foreign language.
 - A document that contains adult material.
 - A document that is unreadable or broken.

When judging Topic Relevance, it is important for the assessor to remind themselves of the following:

1. **For the Highly Relevant and Relevant judgments, it does not matter whether the assessor believes the information provided in the web document is incorrect or could be of harm to the searcher.** The assessor's job is not to try to help the searcher make the right correct decision.
2. For some topics, the health issue is specific to a certain group of people (e.g. children, infants, pregnant women, etc.). Relevance judgments should match the target audience mentioned in the topic statement.
3. Please be aware that the topics' narratives may include the medical intervention's ability to "prevent", "cure", "treat", "help" (or other terminology) a certain health issue. Some documents may include such terminology and assessors should pay attention to what is mentioned in the narrative.

An assessor is **only** required to assess the *medical intervention* efficacy (below) of the website if it has been judged as Highly Relevant or Relevant. Similarly with credibility (below).

Medical Intervention Efficacy

Assessors should judge the *medical intervention* efficacy according to the content of the web document, and not what the assessor believe is the correct information. Treatment efficacy is only judged if the topic relevance of the document is Highly Relevant or Relevant.

The following options are provided to the assessors:

- **Effective**
 - The web document states that
 - The *medical intervention* is an effective option to the health issue, or
 - The *medical intervention* can be an effective option to the health issue.

- There are evidences for both directions but the document clearly supports/concludes effective option over ineffective one.
- **Inconclusive**
 - The web document states that
 - The *medical intervention* is both an effective and ineffective *medical intervention* to the health issue and the document does not sharply support one over the other, or
 - It states that it is unknown whether or not the *medical intervention* helps, or
 - It mentions the *medical intervention* but does not provide any information on its efficacy, benefits, or disadvantages.
- **Ineffective**
 - The web document states that
 - The *medical intervention* is ineffective to the health issue, or
 - The *medical intervention* is harmful/undesirable (direct negative influence on the health issue).
 - There are evidences for both directions but the document clearly supports/concludes ineffective option over effective one.
- **No Information**
 - If the document does not state the health issue but the assessor considered it relevant.

When judging Medical Intervention Efficacy, it is important for the assessor to remind themselves of the following:

1. **The assessor should judge the *medical intervention* efficacy according to the web document content, and not based on what the assessor believes/knows what the correct information is.**
 - a. For example, if a document mentions that a *medical intervention* helps, but the assessor knows that it does not, based on previous knowledge or acquired knowledge from assessing other documents, the document's *medical intervention* efficacy should still be judged as Effective.
2. As with Topic Relevance, please be aware of the topic's narrative and pay attention to differences in the terminology used to describe what the user is looking for, i.e. whether it is "preventing", "curing", "treating" or other terminology.

Credibility²

² The idea of understanding the purpose of a website before judging its quality, determining the amount of expertise, authoritativeness, and trustworthiness (E-A-T), and the cdc.gov example of high E-A-T are ideas based on Google General Guideline.

(<http://static.googleusercontent.com/media/www.google.com/en//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf> Last Accessed: 17/12/2018).

Understanding the purpose of a web document should be the first step to judging credibility. The assessor's opinion of the purpose of web document and the correctness of information matters in judging credibility. Credibility is only judged if the topic relevance of the document is Highly Relevant or Relevant, and is not to be judged if the document is Not Relevant.

When assessing credibility, assessors should not confuse it with *topic relevance or efficacy* of a document and judge credibility **independently** from their assessments of the other two categories, **since a non-credible document may *equally be useful or helpful* for a person making his/her decision.**

Assessors should use their judgment to judge the credibility of documents on two scales as described below

- **Credible**

To help determine if a web document is credible,

- Try to determine the amount of expertise, authoritativeness, and trustworthiness of the web document. Medical websites (e.g. www.cdc.gov) have high amount of expertise, authoritativeness, and trustworthiness.
- Check for an author or an institute that published the web document.
- Look for author credential and qualification and determine if the person has the qualification to write this document.
- Certain documents can be written by a non-expert from a well-known newspaper/news channel website and citing the information they provide in their articles. Such documents can be considered credible.
- Certain documents may not contain author information. Check whether there exist any citations of credible sources such as universities research/clinics or government websites.
- Check where the document is published. Medical publishers or hospitals/clinics or government website or online newspapers with wide circulation are usually credible sources.
- Check for references and sources. Credible sources often cite medical publications and/or lab studies.
- Check how well-written, well-researched and organized is the information.

- **Not Credible**

- Determine if the web document is a mask for advertising or marketing purposes. If so, the website might be biased or scam designed to trick people into fake treatments or into buying medical products that do not live up to their claim.

- Determine if the information posted is from a personal blog or a forum, or by a non-expert person providing a medical product review or providing medical advice. Such subjective personal opinions or one point-of-view are considered not credible.
- Determine whether the website provides or states claims that go against well-known medical consensus (e.g. smoking cigarettes does not cause cancer).

When judging Credibility, it is important for the assessor to remind themselves of the following:

1. The assessor's opinion is important in determining credibility. If the assessor believes the web document has no attempt to help the reader, may cause harm to the reader, misinform or deceive the reader, incorrect, or without any beneficial purposes (e.g. scamming, marketing false information, selling fake drugs, etc.) it should be judged as **Not Credible**.

Important Notes

1. The assessors are not the topic authors or the searchers.
2. An assessor should assess the medical intervention efficacy if the document's topic relevance is judged as Highly Relevant or Relevant.
3. An assessor should assess the credibility of the website if the document's topic relevance is judged as Highly Relevant or Relevant.
4. An assessor should not let their assessment on one category affect their assessment in another category. For example, the assessor judgment on the credibility of a document is independent from their judgment on the medical intervention efficacy or the document topic relevance.
5. An assessor should not let their judgment of the *medical intervention* efficacy of one document affect their judgment on the *medical intervention* efficacy of other documents.
6. An assessor should judge the *medical intervention* efficacy according to the content of the web document and not what the assessor believe is the correct information.

Version Control Table

Version number	Purpose/Change	Author/s	Date
0.1	Initial draft	Mustafa Abualsaud, Fuat Can Beylunioğlu.	14/12/2018
0.2	To reflect changes from email discussions with Mark Smucker and Fuat Can Beylunioğlu.	Mustafa Abualsaud.	17/12/2018
0.3	To reflect changes proposed by Guido from email discussion.	Mustafa Abualsaud.	04/02/2019
0.4	To reflect changes from email discussions with Ellen Voorhees.	Fuat Can Beylunioğlu.	04/02/2019
0.5	To reflect changes from discussion with Mark Smucker and Fuat Can Beylunioğlu. The major difference is changing the usefulness category to topic relevance.	Mustafa Abualsaud.	25/02/2019
0.5.1	To reflect changes from discussion with Mark Smucker in the IR lab on Feb 28. See email sent to Mark for more details. Changed topic relevance to be the same as previous traditional relevance judgments. Missing example topic details.	Mustafa Abualsaud.	08/03/2019

0.5.2	Edits by Mark and adjustment of topic format.	Mark Smucker	May 3, 2019
0.5.3	Edits by Mustafa: 1) changed the word treatment into <i>medical intervention</i> . 2) <i>Changed the medical intervention efficacy categories to effective and ineffective, as opposed to helpful and does not help. Include remark to inform assessors to be aware of narrative of the topics.</i>	Mustafa Abualsaud	May 8, 2019