CD^{2}CR:News Articles and Scientific Paper Annotation Guidelines

By James Ravenscroft

Version 1.0 PUBLISHED

18/4/2020

Introduction

The purpose of this task is to identify examples of words and phrases that refer to the same object or concept across two related documents, even though they may use different language and style.

For example, a newspaper report about a new scientific work may say "Red Meat may contain cancer causing chemicals" and the related scientific work may say "an analysis of carcinogenicity of red meat". The task here is to understand that "red meat" refers to the same thing in both sentences and possibly less obviously that "carcinogenicity" and "cancer causing" are the same thing in this context, forgiving any differences in tense and grammar.

As in the example above, this task is generally fairly intuitive with occasional, more difficult scenarios. This document aims to cover general ground guidelines for how you should approach the task and gives some examples of more difficult scenarios and how they should be resolved.

Core Principles

An overview of how the task works

You will be presented with a summary of a newspaper article and the abstract from a related scientific paper. You will also be presented with two sets of words or keywords that represent concepts in each document. By reading the two short documents you should use your human judgement to indicate whether the two sets of keywords refer to the same concept across the two documents or not by clicking the "Yes" or "No" buttons respectively. (see **Figure 1** below for an example of how this will be presented to you during the task). If a task is particularly challenging because the link between the sets of words is not simple then you can click "This task is difficult to think about".

Are *'drinking alcohol'* and *'alcohol consumption'* mentions of the same thing?

Use the below button to add this task to the 'difficult' list. You need to do this before you give a Yes/No/Report answer if applicable.

This task is difficult to think about

Use the buttons below to give a final Yes/No/Report answer

Yes

No

Bad Example

News Summary [link]

The work, published in the British Medical Journal, showed a 14% to 25% reduction in heart disease in moderate drinkers compared with people who had never drunk alcohol. Another article, by the same Canadian research group, showed alcohol increased "good" cholesterol levels. For many years, studies have suggested that **drinking alcohol** in moderation has some health benefits.

Science Abstract [link]

Objective To conduct a comprehensive systematic review and meta-analysis of studies assessing the effect of **alcohol consumption** on multiple cardiovascular outcomes. Design Systematic review and meta-analysis. Data sources A search of Medline (1950 through September 2009) and Embase (1980 through September 2009) supplemented by manual searches of bibliographies and conference proceedings. Inclusion criteria Prospective cohort studies on the

Figure 1: Screenshot of an example within the CDCR Tool.

There may be cases where the two mentions are the same but have different wording but there are also going to be cases where they have identical wording like in **Figure 2** below.

Are *'the moon'* and *'the Moon'* mentions of the same thing?

Use the below button to add this task to the 'difficult' list. You need to do this before you give a Yes/No/Report answer if applicable.

[This task is difficult to think about]

Use the buttons below to give a final Yes/No/Report answer

[Yes]

[No]

[Bad Example]

News Summary [link]

Chinese scientists have identified a new kind of rock on **the moon** . An unmanned Chinese lunar lander, launched in 2013, has explored an ancient flow of volcanic lava and identified mineral composition entirely unlike anything collected by the American astronauts between 1969 and 1972, or by the last Soviet lander in 1976. Japan, India and China have all launched lunar orbiters on their own rockets.

Science Abstract [link]

The chemical compositions of relatively young mare lava flows have implications for the late volcanism on **the Moon** . Here we report the composition of soil along the rim of a 450-m diameter fresh crater at the Chang'e-3 (CE-3) landing site, investigated by the Yutu rover with in situ APXS (Active Particle-induced X-ray Spectrometer) and VNIS (Visible and Near-infrared Imaging Spectrometer) measurements. Results indicate that this region's composition differs

Figure 2: Screenshot of an example where the two mentions are identical

How long does the task go on for?

The task will continue until the system runs out of examples to present to the user. You can take a break at any point because the system will remember where you got to and take you back to the most recent example that you haven't yet annotated.

How the examples are generated and why it's important

The examples are generated automatically using an algorithm that discovers related pairs of news articles and scientific papers and then compares all concepts within the two documents against each other (every 'key phrase' is compared with every other 'key phrase' from the other document). We use an algorithm to provide a crude 'similarity' between pairs of key phrases allowing us to sort them from most similar to least similar in our annotation tool.

This means that you will see pairs that are most similar - possibly exactly the same word or phrase - to begin with and they will become more challenging or less likely to be the 'same' as you work through the examples.

Of course sometimes our automatic system gets things wrong. You might occasionally find that the newspaper article and scientific paper pairing are completely unrelated. If that is the case then simply click the "Report Bad Example" button to remove that example from the system. There are other scenarios when a task is a "bad example". Please refer to the section below for further guidance.

Resolving Complicated Examples

Please use the following instructions to resolve complicated examples.

Good example or bad example?

The following instructions should help you to decide whether to use the "Report bad example" button.

The key phrases are completely unrelated to each other

- Read the two summaries. If they are about the same topic but the key phrases are unrelated then this is simply a 'No' example.
- If the newspaper article and scientific paper summaries appear to be completely unrelated then see "Newspaper and Scientific Paper aren't about the same thing" below.

Newspaper and scientific paper aren't about the same thing

- If the newspaper article and scientific paper appear to be unrelated, please open the links and read the full news article - it might be that the scientific paper was mentioned in passing - in which case the answer might be a "no" rather than a "bad example".
- An example is a bad example if the newspaper article and scientific paper are clearly unrelated to each other after closer examination.

Are *'ammonia'* and *'functional tricuspid regurgitation'* mentions of the same thing?

Use the below button to add this task to the 'difficult' list. You need to do this before you give a Yes/No/Report answer if applicable.

This task is difficult to think about

Use the buttons below to give a final Yes/No/Report answer

Yes

No

Bad Example

News Summary [link]

The first close-up observations from Nasa's Juno spacecraft have captured towering clouds, swirling cyclones and dramatic flows of **ammonia** that drive giant weather systems on the largest planet in the solar system. The $1.1bn probe swung into orbit around Jupiter in July last year on a mission to peer through the thick clouds that shroud the planet and learn how the alien world, and ultimately all of the planets in the solar system, formed around the nascent sun 4.5bn years ago. Facebook Twitter Pinterest Once every 53 days the Juno spacecraft swings close to Jupiter, speeding over its clouds.

Science Abstract [link]

<jats:p>Surgical treatment is the gold standard treatment of **functional tricuspid regurgitation** (FTR) but this carries high risks of morbidity and mortality. Percutaneous procedures are an attractive alternative to surgery for selected patients deemed to be high-risk surgical candidates. A number of tricuspid transcatheter devices have been developed to treat FTR, but at present,

Figure 3: Unrelated scientific paper and newspaper articles - the news article is about discoveries made by the Juno spacecraft, the scientific paper is about a specific surgical treatment.

One or both of the **key phrases** is gobbledygook/unreadable

- If one of the key phrases is missing a few characters but you can still understand what it means from context then please treat the example as "good" and answer with yes or no.
- If the key phrase is completely unreadable even with context from the summary documents then mark the example as "bad"
- Please see "Formatting Issues" below for further guidance

One or both of the **summary documents** is gobbledygook/unreadable

- If either summary document is missing a few words at the beginning or end but you can still understand the general context then please treat the example as "good" and answer with yes or no.
- If there are words missing from the summary, please use the links to the full documents to help you to understand the summary. If this helps then please treat the example as "good" and answer with yes or no.
- If after further examination you cannot understand one or both of the summaries then mark the example as "Bad Example"
- Please see "Formatting Issues" below for further guidance

Formatting issues

Nonsensical word placement

Are 'shift work' and 'Conclusions Shift work' mentions of the same thing?

Figure 3: An example where bad formatting has seeped into the entity text (Conclusions is the heading/title of the section that the mention was pulled from)

Sometimes the automatic process that creates the examples will make a mistake and the formatting of the task may be confusing or incorrect. **Figure 3** shows an example where the section heading 'Conclusions' has been included in the mention from the scientific work.

Since these issues are a side effect of the way that the text has been processed and are unrelated to whether or not the two mentions refer to the same thing, you should mark these examples as **Bad Example**

Mangled/incorrect numbers

As part of the automated process that creates the tasks, numbers are sometimes mangled or broken. In these cases we are still interested in whether or not the mentions refer to the same concept regardless of the numbers.

In **Figure 4** below the news mention is '000 runners' and the scientific mention is 'Participants 3913 of 7048 participants in the Bonn marathon 2010' - by reading the two summaries it is possible to deduce that both phrases refer to the same. Therefore this example results in a "Yes" answer.

Are '000 runners' and 'Participants 3913 of 7048 participants in the Bonn marathon 2010' mentions of the same thing?

Use the below button to add this task to the 'difficult' list. You need to do this before you give a Yes/No/Report answer if applicable.

This task is difficult to think about

Interesting Frame (Append Spreadsheet)

Use the buttons below to give a final Yes/No/Report answer

Yes

No

Bad Example

News Summary [link]

Report claims taking painkillers before or during race offers no benefit compared with taking them afterwards

Using painkillers during a marathon does not reduce pain and could damage a runner's health, doctors have said. Researchers from the Bonn Centre for Pain Management and the University of Erlangen-Nuremberg interviewed **000 runners** at the Bonn Marathon in 2010 about their use of painkillers and the effect they had. Researchers found that runners who used painkillers had a 13% increased risk of "adverse events" including muscle cramps and intestinal cramps.

Science Abstract [link]

Objectives To prevent pain inhibiting their performance, many athletes ingest over-the-counter (OTC) analgesics before competing. We aimed at defining the use of analgesics and the relation between OTC analgesic use/dose and adverse events (AEs) during and after the race, a relation that has not been investigated to date. Design Prospective (non-interventional) cohort study, using an online questionnaire. Setting The Bonn marathon 2010. **Participants 3913 of 7048 participants in the Bonn marathon 2010** returned their questionnaires. Primary and secondary outcomes Intensity of analgesic consumption before sports; incidence of AEs in the cohort of

Figure 4: The number of runners in the news mention is mangled but it is clear that the author is referring to the 3913 runners who participated in the study.

Complicated cases of 'yes' or 'no'

The following instructions should help you to resolve cases where the example is more difficult.

You can click the "This example is difficult to think about" button to alert the research team to particularly challenging examples. If another user has already reported an example then a message to this effect will appear instead of the button.

One of the key phrases is less specific than the other

- Generally speaking journalists use less precise wording to appear to a more general/less specialist audience.

- Please check whether, to the best of your judgement, the news article is referring to the same concept as the scientific work despite linguistic differences. Use the summary context or full document link to help in this judgement. Below are some examples to help:

   a. A news article states that "Sheep able to distinguish pictures of celebrities from unfamiliar faces with near-human accuracy… the connection emerges from work on face recognition skills of a welsh mountain breed..."

   AND

   The related scientific paper says "we trained eight sheep (ovis aries) to recognize the faces of four celebrities…"

   The key phrases are "a welsh mountain breed" and "eight sheep (ovis aries)".

   In this case we can infer from context that both "a welsh mountain breed" and "eight sheep (ovis aries)" refer to the same set of sheep that were trained on the facial recognition task so this example would result in a **YES** answer.

   b. A news article says "the humble shark, in particular, the great white shark is perhaps the most feared…",

   AND

   A related scientific article says "great white sharks (carcharodon carcharias) less likely to attack humans…"

   The key phrases are "the humble shark" and "great white sharks (Carcharodon carcharias)".

   In this case we can infer that the journalist was aware of the difference between sharks in general and great white sharks and that "the humble shark" is in fact different to "great white sharks (Carcharodon carcharias)" resulting in a **NO** answer.

'Technical' differences between the mentions

Please bear in mind that it is the intent of the journalist that is important rather than any specific technical differences between the mentions. Please keep an open mind when you compare example mentions and give the benefit of the doubt to journalists using less precise language or modifiers.

Are *'orangutan numbers'* and *'wild orang-utan populations'* mentions of the same thing?

Use the below button to add this task to the 'difficult' list. You need to do this before you give a Yes/No/Report answer if applicable.

Task already in 'difficult list', added by Maria

Interesting Frame (Append Spreadsheet)

Use the buttons below to give a final Yes/No/Report answer

Yes

No

Bad Example

News Summary [link]

BBC News science reporter Deforestation has had a huge impact on **orangutan numbers** Researchers report that a population crash occurred during the past 200 years, coinciding with deforestation in the same area. The study focuses on orangutans found in the forests of Kinabatangan Wildlife Sanctuary in Malaysia.

Science Abstract [link]

Great ape populations are undergoing a dramatic decline, which is predicted to result in their extinction in the wild from entire regions in the near future. Recent findings have particularly focused on African apes, and have implicated multiple factors contributing to this decline, such as deforestation, hunting, and disease. Less well-publicised, but equally dramatic, has been the decline in orang-utans, whose distribution is limited to parts of Sumatra and Borneo. Using the largest-ever genetic sample from **wild orang-utan populations** , we show strong evidence for a

Figure 5: In this example we compare orangutan numbers and wild orangutan populations. The mentions have different geographical contexts (Malaysia and Sumatra and Borneo respectively) but could be considered the same.

In **Figure 5** above we compare "orangutan numbers" and "wild orangutan populations". The news summary seems to discuss a wildlife sanctuary in Malaysia whereas the scientific work centres on wild orangutan populations in Sumatra and Borneo. It is clear from reading the full news article that the journalist is talking specifically about the study despite the geographical distinction.

Given that our primary focus is journalistic intent and the journalist is adamantly talking about orangutan populations discussed in the study, we would resolve this example to be a "yes".

By reading the full scientific paper (not required as part of the annotation process) it is possible to understand the full context: the orangutans originated in Sumatra and Borneo but were relocated to a malaysian wildlife sanctuary. This explains the inconsistency between the journalist and scientific work and confirms that "yes" was the correct answer.