

Instructions for Annotating Answers for Multi-Perspective Question Answering

Introduction

The research question of Multi-Perspective Question Answering (MPQA) is to discover efficient algorithms that can accomplish the task of answering questions about beliefs, opinions, and evaluations embedded in natural language texts. For the purpose of evaluating such systems we need a collection of sample documents together with a set of questions. In addition, for each question we need information about what parts of each document constitute an answer. Traditionally, question answering (QA) collections have used two kinds of information about what constitutes an answer to the questions in the collection: the textual form of the answer and/or the segments of documents in the collection that can constitute an answer. Information about what segments in the documents constitute (or can contribute to) an answer to each of the questions is largely used in evaluation of QA systems for at least two reasons: firstly, it can help assure that the system has found a real answer to the question as opposed to being lucky in picking out the correct answer in the wrong context; and, secondly, it can give credit to a system for finding the place in the collection where a question is answered although the system may not be able to convert the string to the exact string representation required for the answer (e.g. answers to yes/no questions).

For the task of MPQA, the latter reason for relying on information about the text segments that answer questions is even more important considering that it is much harder to convert an answer segment to an exact answer string even when the answering segments are found in the text. Therefore, we will augment the MPQA collection by adding annotations (marking up) to every text segment that can contribute to the answer for any question in the collection. This document explains what and how to annotate, as well as the idea behind the annotations.

We will use an annotation of type ANSWER to designate text segments that constitute answers. An ANSWER annotation will be represented as an xml marking in the source document and will contain five attributes: *question topic*, *question number*, *confidence*, *confidence comment*, and *partial answer*. You will use the *question topic* and *question number* attributes to identify the question that the annotated text segment answers. These attributes will have to be set explicitly for all ANSWER annotations that you add, while all other annotations will have default values.

We will give an example of an ANSWER annotation. Assume question 3 of topic VENEZUELA is: “When was Chavez elected president of Venezuela?” and it is answered in the sentence “Chavez was elected president in 1991. The ANSWER annotation that we will add to the document from which the sentence came will look as follows: “Chavez was elected president <ANSWER: *topic*=Venezuela, *question #*= 3, *confidence*=5, *conf comment*=”, *partial*=false in 1991>.”

In the next section we give some general instructions about the annotations, followed by a section that describes all attributes of an ANSWER annotation.

General Instructions

This section gives general directions of what text spans should be annotated as answers.

What constitutes an answer

The most important decision that has to be made during annotation is what constitutes an answer for the question. For the purpose of this QA collection, *we will annotate as an ANSWER every text segment that can contribute to the derivation or construction of an answer to the question.* We have identified the following difficulties that you are likely to encounter during annotation:

- 1) **Text segments can answer questions only indirectly.** For instance, the question “Did most Venezuelans support the 2002 coup?” can be answered by the following two text segments (among others), both of which should be annotated as answers: “Protesters ... failed to gain the support of the army” and “... thousands of citizens rallied the streets in support of Chavez.” The above two answers are indirect in two different ways. First, the subject in the above two sentences is not “most Venezuelans” as the question asks. Rather, we can infer from the fact that thousand of citizens were against the coup that it was not the case that most Venezuelans supported the coup (and in the case the inference is questionable since thousand of people may protest although most Venezuelans may still support the coup). Second, the answer is indirect because we must infer from the fact that thousand of people rallied in support of Chavez that they were supporting the President of Venezuela who was overthrown by the coup, and thus they were against the coup. That is, even if the segment was “... most Venezuelans rallied the streets in support of Chavez,” we would still have to infer that most Venezuelans did not support the coup and use a certain amount of discourse knowledge in the inference.

When annotating indirect answers use your best judgment. If you can determine the answer from a text segment using common sense, then the segment should be annotated as an answer. Use the *confidence* attribute to indicate how confident you are that an answer can be inferred from the text segment. In the above example, both segments should be annotated as ANSWER with the former segment having a lower confidence score (maybe a 2 on the scale 1-5) than the second segment (maybe a 3 or a 4).

- 2) **Answers to some questions require combining information from more than one segment in the text.** For instance, the question “Are the Japanese unanimous in their opinion of the Kyoto protocol?” might be answered by combining information from a segment in a document that states that some Japanese support the Kyoto protocol with information from a segment in the same document that states that some Japanese do not support the protocol. While neither of the two

segments would have been sufficient to answer the question, when the information from the two segments is combined, that is enough to give an answer.

In situations such as the one described above, annotate as ANSWER any text segment that contributes to an answer. For instance, in the case of “Are the Japanese unanimous in their opinion of the Kyoto protocol?” question, annotate any text segment that expresses an opinion of Japanese source, regardless of whether or not the document alone contains sufficient information to answer the question. Set the *partial* attribute of the ANSWER annotations to **true** to indicate that a given answer segment does not provide a sufficient answer in isolation.

In addition, some segments may answer a question partially, without any need for combining their information with that in different segments. For instance, a partial answer to the question “When was the Kyoto protocol ratified?” maybe the segment “the protocol will be ratified in the near future,” which although not answering the question completely gives a lower bound on the date of the ratification. Such segments should be annotated as answers with the *partial* attribute set to **true**.

- 3) **Sometimes it may be hard to know whether the sources in the document match the entities about which questions ask.** For example, a question mentioned in 1) asks about the opinion of “most Venezuelans.” The two answers given in the example mention “the army” and “thousands of protesters.” Using our background knowledge and our common sense we can conclude (and again this inference is questionable) that “most Venezuelans” did not support the coup.

Similarly, the question given as example in 2), “Are the Japanese unanimous in their opinion of the Kyoto protocol?” asks about the opinion of “the Japanese.” It is not clear what should count as the opinion of “the Japanese” – the opinion of Japanese government sources, the opinion of Japanese news sources, or the opinion of any person or organization from Japan.

A general guideline in the situation described above is to use your common sense and background knowledge and annotate anything that you believe to match the source in the question. If you are unsure of whether a source in a document can be associated with the source of the question, use the *confidence* attribute to indicate that and add a comment to the *confidence comment* field explaining the issue.

- 4) **Sometimes answers may be given in future tense or conditional statements,** especially since the document collection includes documents over a certain time span. For instance, the question “When was the Kyoto protocol ratified?” might be answered by “... will sign the protocol next month,” or the statement “will **likely** draw the ire of X” could answer a question about X’s opinion. Again, future tense and conditional statements should be annotated as ANSWER and the *confidence* and *confidence comment* attributes should be used to indicate how confident are you that the segment answers a question and what is the reason for

the reduced confidence. The confidence score that you assign to the ANSWER annotation should indicate how sure you are that the event occurred in the future given the text span. For instance, the text segment, "the protocol will be signed next March" answering the question "When was the Kyoto protocol signed?" should receive only medium *confidence* score. Although we may believe the author of the document, the event described is in the future and a certain turn of events may have changed the actual date of signing. On the other hand, if the question was "Was there a solar eclipse in 1990?" and it was answered by a segment "the next solar eclipse will be in August 1999", the ANSWER annotation should receive a high *confidence* score since although the event that the segment describes is in the future, we know that with a high probability the solar eclipse occurred as predicted.

- 5) **Sometimes text segments may express opinions only indirectly through the style and choice of language.** In such situations, it may be hard to pinpoint the exact place in the text at which the opinion is expressed. However, you should be able in such text segments to attribute the expression of opinion to a specific word or phrase. Annotate all words and/or phrases that you think express the opinion for which the question asks. If you cannot pinpoint any particular word or expression, then do not annotate any part of the text as an answer. For example, a question asking about Bush's opinion of Saddam Hussein, could be answered by the segment "Saddam has been oppressing the country for far too long, Bush iterated." In the above segment "oppressing" and "far too long" are the phrases that indicate Bush's negative attitude towards Saddam and should be annotated as ANSWER.

For answers that have difficulties different from the ones described above use your best judgment. The general guideline is to use your common sense and best judgment. If you can infer that a given segment answers a question either fully or partially, then annotate it as an ANSWER and use *confidence* and *confidence comment* fields to indicate if you feel uncertain about the answer.

Minimal Spans

When annotating a text span as an answer to a question use the minimum span that answers the question. For instance, the question "What is the Kiko Network?" should be annotated as answered by the text segment "a Tokyo environment umbrella organization," which is the minimal segment answering the question as opposed to the longer segment "a Tokyo environment umbrella organization representing about 150 Japanese groups," which also constitutes a legitimate answer to the question but is not minimal. We believe that annotating only minimal answer segments will make our evaluation of the QA annotations easier.

Below we summarize all attributes of an ANSWER annotation and how they should be used.

Answer Annotation Attributes

Question Topic

Use to indicate the topic of the question that the segment answers. Should be one of *kyoto*, *mugabe*, *humanrights*, or *venezuela*.

Question Number

Use to indicate the number of the question that the segment answers.

Confidence

Use the confidence attribute to indicate how sure you feel that the segment answers the question, matches the source, or is a full vs. partial answer. Should be one of the following: 1, indicating low confidence, 2, indicating moderate confidence, 3, indicating medium confidence, 4, indicating high confidences, and 5, indicating very high confidence. DEFAULT VALUE: 5, VERY HIGH CONFIDENCE.

Confidence Comment

Use this attribute to indicate why you feel less confident that the segment answers the question. This attribute is free form and can be filled with any text that presents an explanation. The attribute is especially important to include for answer segments that present difficulties different from the ones described above. DEFAULT VALUE: "", EMPTY STRING.

Partial Answer

A Boolean attribute that should be set to *true* if the segment presents only a partial answer to the question and to *false* otherwise. DEFAULT VALUE: FALSE.