

## 2005 DUC Topic Development Task for Summarization

Our DUC 2005 task is a question-focussed summarization task and the scenario is as follows: You (the assessor; an educated, adult US native, aware of current events as they appear in a large daily newspaper) would like to find out some information about some topic that you're interested in, and you go to a question-answering service like Google Answers (<http://answers.google.com>). Your question requires someone to look at multiple documents from a collection of news stories to piece together a satisfactory answer. An "answer" will comprise material found in the documents that contributes to satisfying your information need. You would like the answer to be summarized in 250 words. So, your question is a complex question like "What impact has the Chunnel had on the British economy and the life style of the British"; it is neither a simple factoid question like "What year was the Chunnel completed" nor a request for a list of factoids like "What ships have sunk since 1980".

For the 2005 DUC task, NIST will give you a list of old TREC topics and a set of at least 35 documents (TREC newswire articles) for each topic. These documents have previously been judged as being relevant to the topic but, as you may know, assessment of relevance may vary from person to person. You will use *WebAssess* to peruse through enough documents so that you can develop a DUC topic and identify a subset of 25-50 documents that contribute to satisfying the information need expressed in your DUC topic. You will also be using a text editor to write down your DUC topic. (You will be given instructions on how to use *WebAssess* and the text editor if you don't already know how to use these tools.) We would like each person to develop a total of 6-7 DUC topics.

Please follow the following procedure for developing each DUC topic:

1. Pick a TREC topic that you find interesting. Write your initials next to the topic number on the master list that will be kept in the middle of the lab. Do not choose a topic that already has an initial next to it.
2. Familiarize yourself with the information in your document set by reading through at least 50 of the documents (using *WebAssess*) and marking each one as "relevant" if you think it is relevant to the TREC topic, and "not relevant" otherwise. If you mark too many as "not relevant" (i.e., you think you won't find at least 25 relevant documents), then stop, remove your initials from the topic in the master list, and go back to Step 1 to pick another topic. As you are reading the documents, look for one or more aspects that interest you; make a note of these aspects, if you wish.
3. Formulate a DUC topic, which is a request for information about the aspects of the TREC topic that interest you. This can be in the form of a question or set of related questions. Your topic statement should describe in only positive terms the information that you are looking for. It is a description of the desired information, not a list of rules for deciding whether a document is relevant or not. The DUC topic must meet these 4 criteria:
  - (a) The answer can be found in the "relevant" documents.

- (b) The answer is complex, requiring about 250 words.
- (c) At least 25 documents must each contribute some material to the answer.
- (d) *The statement of the DUC topic should be clear and specific enough so that if you give someone your topic statement and they look at the “highly relevant” documents that you mark (see Step 4), they will be able to write a good 250-word answer to your question.* The DUC topic should be a stand-alone topic that can be interpreted without reference to the TREC topic, so include as much background/context as you think is necessary to answer your question and satisfy your information need. Your DUC topic can (but is not required to) borrow specifications from the original TREC topic.

Use the text editor to write down this DUC topic using one file per topic. You may modify your DUC topic statement as you go through Steps 4-5 so that the 4 criteria (a-d) above are met.

4. Go through the “relevant” documents and mark them as “highly relevant” if they contribute to answering your DUC topic. You may stop after finding 50 “highly relevant” documents.
5. If you have at least 25 “highly relevant” documents then go to Step 6. Otherwise, if you can easily modify or expand your DUC topic to include more documents then go to Step 3. Otherwise, see the NIST DUC representative.
6. Write a title for this DUC topic (this could be the same as the original TREC title, if appropriate).
7. Write down the granularity of the desired response for each DUC topic. That is, indicate whether you would like the response to your information request to name specific events, people, places, etc., or whether you want the response to speak in more general terms. The granularity should generally reflect your actual preferences, but in some cases the amount of information in the “highly relevant” documents that you’ve chosen necessitates that the summary make generalizations in order to fit in all the information that contributes to answering the question(s).
  - “specific”: You would like the response to identify specific instances of events, people, places, etc.
  - “general”: You want a high-level response that makes generalizations; you’re not so interested in details of specific events, people, places, etc.