

LIS590TX

Assignment 3 – Multi-document Summarization

Motivation:

provide you with first hand experience with writing a manual summary based on a collection of news wire texts. Please complete this assignment *before* reading the assigned papers and write down any insights you gain from going through this process manually.

Instructions:

Read these instructions before beginning the assignment.

collection of 25 news stories that address the stated information need in given topic and query statements automatically. For this assignment you will complete this task manually, but your key deliverable from this assignment is not just the summary that you produce, but also your reflections on the process that you used and your thoughts on how the summary can be produced in an automated way.

and query stated below. You will copy and paste the entire row from the tab marked *Documents* to the tab marked *Relevant Sentences*. You may change the order of these sentences ie the rows, but do not make any change to the column order in the *Relevant Sentences* tab.

Step 2: In the second step you will select a subset of the sentences identified in step 1 that best addresses the stated information need stated below. The subset you select must not exceed 250 words. This is called an *extractive* summary.

Step 3: Using information from the original document or the sentences that you collected in step write the best summary you can that addresses the information need stated below

The Input:

The topic and query statements from the Document Understanding Conference 2006, topic identity 6

Topic: Impacts of global climate change

Query: What are the most significant impacts said to result from global climate change?

The relevant documents:

This file contains the text from 25 articles (see the tab marked Documents). The columns are:

Doid: The unique document identifier

GS: Global sentence identifier - unique for all sentences

T: Type of text where H = heading, L=location and T=text

P: Paragraph Identifier - included so that you can see where the paragraphs were in the original document

S: Sentence Identifier - unique within a paragraph

DS: Document Sentence Identifier - unique sentence identifier within the document

Sentence text: The actual text of the document

Text Mining course, but you may not distribute these news texts beyond our Text Mining course without talking to me first.

Assignment deliverables :

(0) Enter your last name below :

Smith

(1) Name this excel file with your last name and submit with the relevant sentences(step 1), the extractive summary (step 2) and your final summary (step 3) completed in the appropriate tabs. Don't change the tab names or column ordering.

(2) A two page single spaced summary (min 10 point font size) that captures your reflections on the process of writing a manual summary. Include how long each step took you to complete and the degree to which you found step 1 useful (this may be negative) . One of the pages should describe how you would complete these process automatically. For the latter you may use any material that we have already covered in the Text Mining course or beyond.