Computer Engineering Department
Bilkent University

CS533: **Information Retrieval Systems**
Assignment No. 5: Topic-based Novelty Detection Annotation
March 12, 2010

Due date: March 19, 2010; Friday, by 11:59 pm (only on the Web)

Note: The accuracy of your annotations are important to us. Because of that there will be an additional step to resolve conflicts between annotators (note that each topic is assigned to two different annotators). I do not want to encourage that but if you need more time let me know.

Topic-Based Novelty Detection Annotation

In this assignment, we want you to do some annotation which will be used for dataset construction. The constructed dataset will be used in novelty detection experiments in Turkish news.

An account is created for each one of you at http://139.179.21.199/ttracker2/index.aspx. You are also assigned some topics which we want you to annotate. Number of assigned topics may vary from annotator to annotator but we tried to keep the total number of documents as same. Your user names are your names and passwords are your student ids (Names do not contain Turkish characters and they are lowercase). Some special cases for user name assignments are given in the table at the end of this document (for users with two names or for usernames which were already assigned).

When you login to the system, you will see the screen in Figure 1. This screen shows the topics assigned to you for annotation. You can select one of these topics to start annotation by pushing the radio button near the topic and click "Select" button. We suggest you to finish the topic you started before passing to another one. Also, in order not to forget the previously read documents, it would be better if you finish a topic you started before giving a break (Of course this may not be possible for long topics).
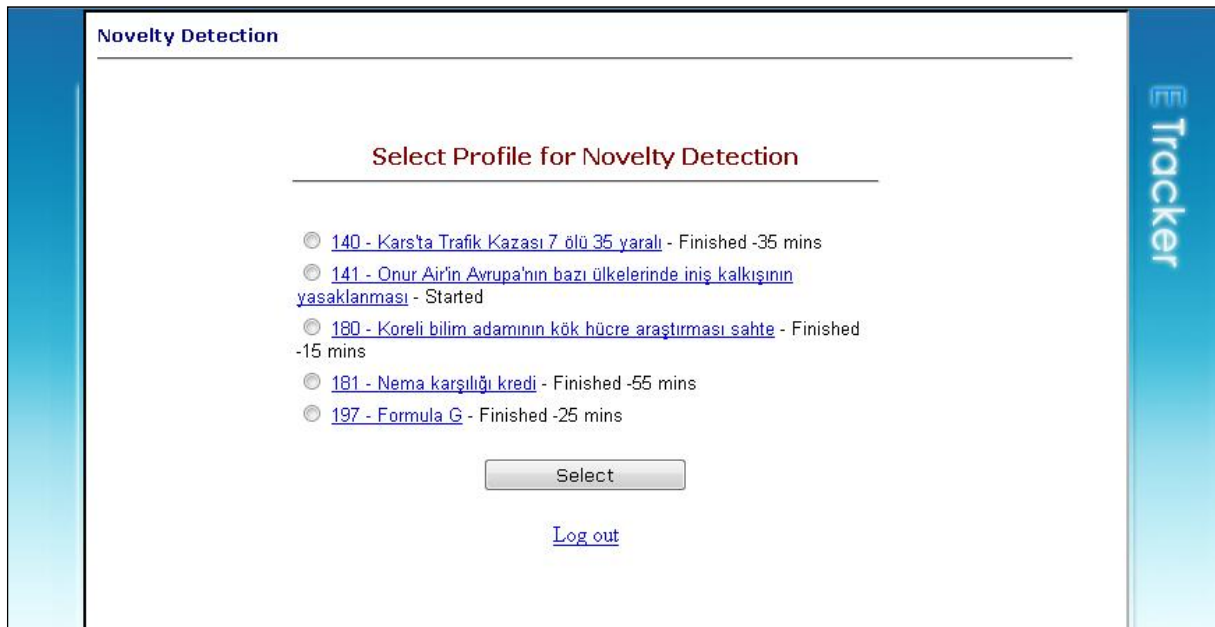

Figure 1. The screen for selecting topics.

After selecting the topic to be annotated, in the next screen the documents included in the topic will be listed. Basically, a topic is initiated with a first story which is the news that reports the topic first in the timeline and tracked by the documents which are relevant to the topic. You can see an example screen shot in Figure 2. We

selected a topic about a flight company's banishment from some of the European countries. We first need to read the "First Story of Topic" (Since it is the first story, there is no labeling for it, it is obviously new). Then, we read document-1 and annotate it as novel if it reports some new information with respect to the first story; otherwise we annotate it as not novel. When we are done with document-1, we read document-2 and annotate it as novel if it reports some new information with respect to the first story and document-1. This process goes like this for all documents incrementally. To sum up, we label a document as novel if it contains new information which has not been provided in the previously read documents. So, you should read the documents in order, our system would not also allow you to read any of the documents if you did not read and labeled all of the previous documents. Please be aware that only 20 documents are shown in one page, so if next page button is available you should go to next page by clicking the button.

Figure 2. Tracking documents of a topic.

When we click the link of a document, the screen in Figure 3 will show up in a new window (or tab). In the left part of the screen, text of the document is presented. You can label the document using the "Novel" and "Not Novel" buttons at the bottom of the screen. "Close" button closes the window. In the right of the screen, a history section is given. The documents you read previously can be seen in this box (This box can be used to skim the previous documents to remember if a piece of information is given or not). The documents that you labeled as novel are highlighted with green in this box.
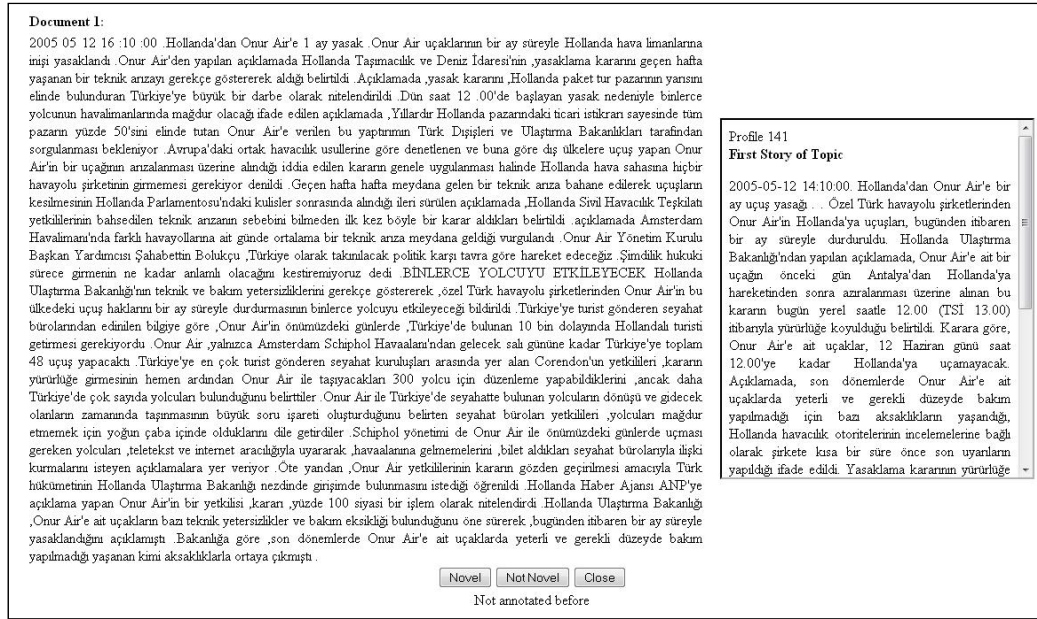
Figure 3. Annotation screen.

When you finish annotating all of the documents, at the last page you can finish annotating by clicking the finish button. We also want you to record the time you spent while annotating the topic. After entering the time with the combobox at the bottom of the screen, you can finish the current topic by clicking the finish button. After finishing, you can still make changes to your annotations. It only allows us to keep track of the finished annotations. An example screenshot is given in Figure 4. Also, at any time you can click the "Save&Close" button to return the screen that lists the topics.

Figure 4. The last page of the document list.

Since any wrong decisions can affect reliability of the experimental dataset and cause incorrect results, we want you to annotate the documents carefully.

Table 1. Username assignments for special cases

Name	Surname	Username
Muhammed Abdullah	Bulbul	abdullah
Ahmet Cagri	Simsek	ahmetcagri
Ahsen Ikbal	Yergok	ahsen
Bugra Mehmet	Yildiz	bugra
Ceyhun Efe	Karbeyaz	ceyhun
Duygu	Sarikaya	duygusarikaya
Omer Faruk	Uzar	omerfaruk
Fahreddin Sukru	Torun	sukru
Utku Ozan	Yilmaz	utku

For problems and other questions please send an email to Cem Aksoy (caksoy at cs.bilkent.edu.tr). Please also don't hesitate to contact me (canf at cs.bilkent.edu.tr).