

Xpantrac Topic Tagging System Tutorial

October 11, 2013

Seungwon Yang <seungwon@vt.edu>
Department of Computer Science, Virginia Tech
114 McBryde Hall (0106), Blacksburg VA 24061, U.S.A.

1. Overall user interface

The user interface (UI) of the prototype Xpantrac (the name came from **eXpansion + extraction** algorithm) system is shown in Figure 1. The interface has three divided sections, namely the Collection pane, the Document pane, and the Topics pane. Users will work on each pane in the order of (a), (b), and (c) to extract topic tags.

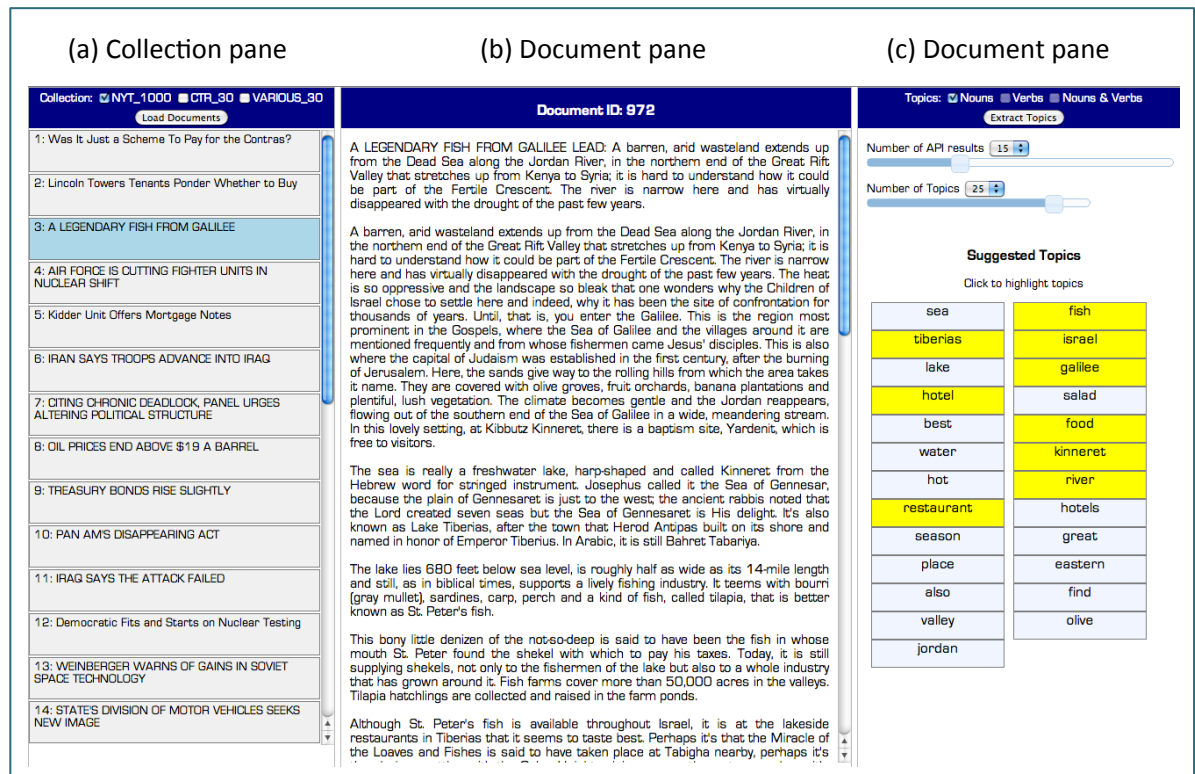


Figure 1. The user interface of the Xpantrac system. (a) Collection pane, (b) Document pane, and (c) Topics pane.

2. Collection pane: Load the documents

Users first load a document collection by adding a check mark for one of the collections: NYT_1000 (*New York Times* articles), CTR_30 (webpages about the Hurricane Isaac disaster), and VARIOUS_30 (various disaster webpages), and clicking the “Load Documents” button shown on the top portion of Figure 2. Make sure the button is pressed.

We use 30 documents from each collection for this study. Once a document collection is loaded, the titles of the documents are shown in rectangular boxes as shown in Figure 2. Users click the box to view the content of the article. The background color of the box changes to sky blue if it is clicked.

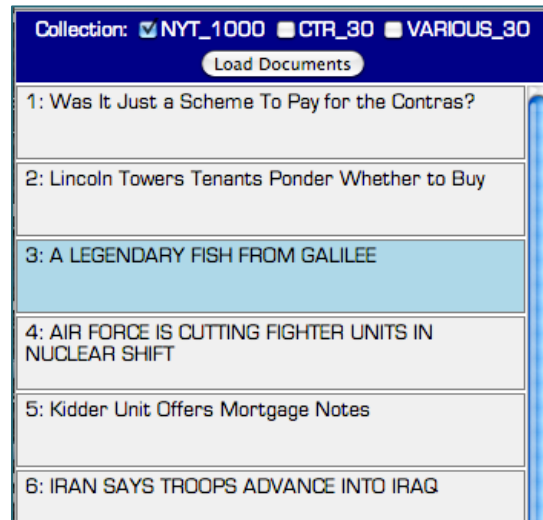


Figure 2. Collection pane. Users load documents to assign tags.

3. Document pane: **View the content of the selected document**

The content of the article is presented in the Document pane, located in the center of the UI. A vertical scroll bar is provided for the longer articles.

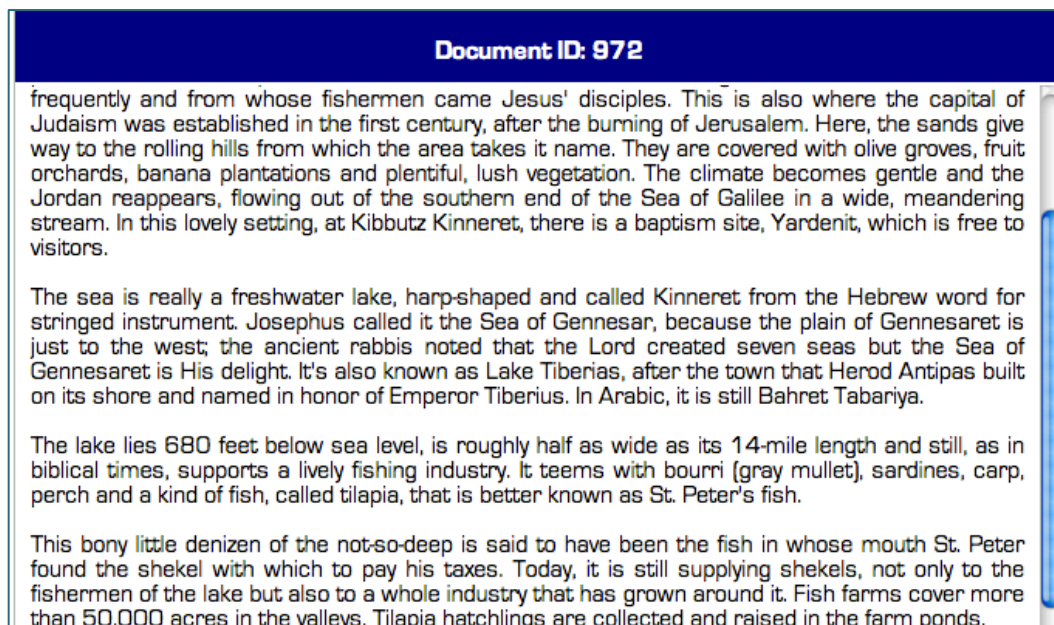


Figure 3. Document pane. Users view the content of the document.

4. Topic pane: **Configure parameters and extract topic tags**

The Topic pane is located on the far right side of the UI. Users configure the parameters and setting before extracting a list of topic tags. First, there are options for whether the users would like to extract topics from (only) nouns, (only) verbs, or from both nouns and verbs. For this study, it is set to (only) nouns, so leave the choices.

There are two slider bars, “Number of API results” and “Number of Topics”, to control parameters of the system. The higher number of API results means the system uses more information in producing the topic suggestions; however, it also means that it takes longer time to extract topics. As the name suggests, “Number of Topics” setting controls how many topic suggestions to be presented. Topics shown on top of the two-column list should be more relevant to the document than the ones at the bottom (e.g., sea → fish → tiberias → Israel →...). Users can click the topic boxes to highlight them for later reference.

The interface is titled "Topics:" and has three radio buttons: "Nouns" (checked), "Verbs", and "Nouns & Verbs". Below this is an "Extract Topics" button. There are two slider controls: "Number of API results" set to 15 and "Number of Topics" set to 25. Below these is the section "Suggested Topics" with the instruction "Click to highlight topics". It contains two columns of topic boxes. The first column has 15 boxes, and the second column has 15 boxes. The topics are: sea, fish, tiberias, israel, lake, galilee, hotel, salad, best, food, water, kinneret, hot, river, restaurant, hotels, season, great, place, eastern, also, find, valley, olive, jordan.

Suggested Topics	
sea	fish
tiberias	israel
lake	galilee
hotel	salad
best	food
water	kinneret
hot	river
restaurant	hotels
season	great
place	eastern
also	find
valley	olive
jordan	

Figure 4. Topic pane. System parameters such as no. of API results and no. of topic suggestions can be changed. Users can mark topics of interest by clicking them.