



# INEX 2010 Book Track

## Topic Development Guidelines

---

Gabriella Kazai

Version 1.0

### 1 Introduction

This year we will run two new search tasks: “Prove It” and “Best Book to Reference”. In the Prove It search task systems will need to find evidence in books that can be used to either confirm or reject a factual statement. In the Best Book to Reference (or Best Book for short) search task systems will need to return the most relevant books on the general subject area of the factual statement.

This year we will also experiment with using Amazon Mechanical Turk service to gather both topics and relevance judgements – in addition and *in parallel* to collecting these from the book track participants. Our aim is to test the reliability of the crowdsourcing approach so that in future years we can move the test collection creation completely to crowdsourcing.

As in previous years, we are building on a corpus of over 50,000 digitized books and aim to build a test collection to evaluate IR and XML-IR approaches to the two search tasks.

This document provides guidelines on how to create the topics that will form part of the test collection.

### 2 Requirements and deadline

Each participating organisation is asked to create **a minimum of 5 topics, at least 2 of which should be on facts that are also included in Wikipedia**. Topics must be submitted by **Saturday, July 31, 2010** using the candidate topic submission form at:

<http://www.inex.otago.ac.nz/tracks/books/TopicDevelopment.asp>

### Topic Format

A topic is a representation and explanation of an information need. ***It is crucial that all parts of a topic are carefully filled in!***

- |                          |   |
|--------------------------|---|
| <b>&lt;fact&gt;</b>      | A factual statement that we want to confirm or reject based on evidence found in books.   |
| <b>&lt;query&gt;</b>     | The query that you would use in a web search engine to find information relevant to the factual statement.  |
| <b>&lt;wiki&gt;</b>      | The URL of the Wikipedia article that also contains this fact (must be given for at least two of your topics).  |
| <b>&lt;narrative&gt;</b> | Any explanatory information, including why the fact may be of interest, e.g., how it may be related to a task or project, and an <u>exhaustive</u> account of what aspects of the topic are relevant or irrelevant. |

A couple of example topics are given below.

```
<fact> On the 11th of July, 1798, an act was approved for the establishing and organizing a Marine Corps. </fact>
<query> US Marine Corps </query>
<wiki> http://en.wikipedia.org/wiki/US\_Marine\_Corps#Origins </wiki>
<narrative>
  <task>
    I am editing the Wikipedia article on Marine Corps (see http://en.wikipedia.org/wiki/US\_Marine\_Corps#Origins)
    and want to confirm that this fact is correct.
  </task>
  <infneed>All statements in books are relevant that either prove or reject the fact. All facts are relevant even if the
    full date is not given – the year is sufficient. I am only interested in the Marine Corps of the USA. The more details
    are given regarding the how the Marine Corps were established, the better.
  </infneed>
</narrative>
```

An example book in the book corpus that the above fact is confirmed is  
[http://www.booksearch.org.uk/BookViewer.aspx?A=de9511ef-9deb-4878-88d8-25a1cb055e8f&B=394D4A3431E4D63B&C=0&E=US Marine Corps&Page=52](http://www.booksearch.org.uk/BookViewer.aspx?A=de9511ef-9deb-4878-88d8-25a1cb055e8f&B=394D4A3431E4D63B&C=0&E=US_Marine_Corps&Page=52)

```
<fact>The Marabou stork has a water sack in an external bag a little below its throat, which allows it to make long
  flights over arid deserts.</fact>
<query> Marabou water sack </query>
<wiki></wiki>
<narrative>
  <task>I am compiling a list of birds with water sacks and categorizing them based on whether the sack is external
    or internal. Thus, I want to find evidence to back up the fact above or to contradict it – maybe the stork has no such
    sack or has an internal one.
  </task>
  <infneed>Relevant passages must state the location of the Marabu stork's water sack. I am not interested in any
    other stork. Images where the sack is clearly visible or illustration of it are also relevant.
  </infneed>
</narrative>
```

An example book confirming this fact is [http://www.booksearch.org.uk/BookViewer.aspx?A=318a38cc-d65c-41b9-bd2a-a2858ab3d9db&B=5CACA36DAA470D45&C=5&E=Marabou water sack&Page=68](http://www.booksearch.org.uk/BookViewer.aspx?A=318a38cc-d65c-41b9-bd2a-a2858ab3d9db&B=5CACA36DAA470D45&C=5&E=Marabou_water_sack&Page=68).

### 3 Procedure for Topic Development

Please follow the steps below to create your topics. You can use Topic Form in Appendix to record all information about your topic before submitting it through the INEX web site (URL is given in Section 2).

#### Step 1: Find facts in books

Your task is to find **at least 3 general knowledge facts that you believe are true in the books** available at <http://www.booksearch.org.uk>. Both the facts and the books must be in English. A fact should not be longer than a paragraph.

For example, the fact that "The first Electric Railway in London was opened in 1890 and run between the stations: Bank and Stockwell" can be found on page 187 of the book titled "West London" by George Bosworth.

#### Step 2: Find the same fact both in Wikipedia articles AND in books

Your task is to find **at least 2 facts that appear BOTH in a Wikipedia article AND in a book** available at <http://www.booksearch.org.uk>. You can start either by finding a fact on Wikipedia first, then locating the same fact in a book, or you can start by finding a fact in a book and then in Wikipedia.

For example, the [Wikipedia page on Beethoven's Symphony No. 3](#) claims that "Beethoven dedicated the symphony to Napoleon, but when Napoleon proclaimed himself emperor, Beethoven tore up the title". [Page 144 of the book titled Beethoven by Romain Rolland describes this very fact.](#)

### **Step 3: Topic Submission**

To submit your topic, fill in the online Topic Submission Form on the INEX website - see Section 2 for the URL.

After submitting a topic you will be asked to fill in an online questionnaire (this should take no longer than 1-2 minutes). This is part of an effort to collect more context for INEX topics, thereby increasing the reusability of the test collection (see [1]).

## **4 Topic Selection**

There is no topic pruning or selection by the organizers. All created topics will be distributed to all participants and subsequently included in the test set to be used in the search phase.

## **References**

[1] Kamps, J. and Larsen, B. (2006). Understanding Differences between Search Requests in XML Element Retrieval. In *Proceedings of the SIGIR 2006 Workshop on XML Element Retrieval Methodology*, p. 13-19. [<http://www.cs.otago.ac.nz/sigirmw/>]

## Appendix: Topic Submission Template

**\* Your Topic**

Factual statement (type in here the fact you found in a book available at [www.booksearch.org.uk](http://www.booksearch.org.uk)):

Query (enter the query you would use in a web search engine to find information on the above fact):

Book URL (copy and paste the URL of the book that contains the fact):

Book page number (enter the page number **shown next to the blue arrows**, below the book, that contains the fact):

Wikipedia URL (copy and paste the URL of the Wikipedia article that contains the same fact as above – note that the fact does not have to match word by word):

Wikipedia version of the fact (copy the sentence(s) from Wikipedia that contains your fact):

Narrative (please detail your task and what aspects would make a text snippet relevant or irrelevant to your fact):

<task>

</task>

<infneed>

</infneed>

Comment (please enter any comments you may have):