

# **BOLT IR Query Development Guidelines**

Version 1.0, May 3, 2013

**Linguistic Data Consortium**

**Changes in this version**

- Added target language description

## Table of Contents

<b>1. Introduction</b> .....	<b>1</b>
<b>2. Searching and beginning query development</b> .....	<b>1</b>
2.1 Initial search/exploration.....	1
2.2 Using the search tool .....	1
<b>3. Developing a full query</b> .....	<b>2</b>
3.1 Writing a natural language query.....	2
3.2 Filling in the query information .....	2
3.2.1. Development time.....	2
3.2.2. Description.....	2
3.2.3. Language target.....	2
3.2.4. Asks about.....	3
3.2.5. Asks for.....	3
3.2.6. Languages: arz, cmn, eng.....	3
3.2.7. Rules of Interpretation.....	3
3.3 Adding human answers to the query .....	3
3.3.1. Answer Text.....	4
3.3.2. Answer doc ID.....	4
<b>4. Constraints on queries</b> .....	<b>4</b>
4.1 Avoid queries with strong lexical triggers .....	4
4.2 Balancing simple and complex parts of queries .....	4
4.3 Variety in wording.....	5
4.4 No “How do I..” queries .....	5
<b>5. Appendix: Some example natural language queries</b> .....	<b>5</b>

## 1. Introduction

BOLT Information Retrieval (BOLT IR) is a task in which computer systems are confronted by a large corpus of discussion forum threads, and are given a **natural language query** (e.g. “What do people say about Facebook?”, or “What caused the Euro crisis?”), and then asked to look for **answers** to these queries in a set of discussion forum threads harvested from the web. These discussion forum threads are either in Arabic, Chinese, or English.

To support this task, we will develop natural language queries in the form of a single English query sentence, a set of query information, and a set of human answers to the query. The query information consists of a query’s description, its target language, its properties, and the **rules** for assessing answers to the query (these rules will be used later to assess system answers to the query).

These guidelines describe how to search the discussion forum corpus and create queries for the BOLT IR task.

## 2. Searching and beginning query development

### 2.1 Initial search/exploration

Begin by searching the corpus for a topic or question of interest to you, and reading through some of the threads that come up as a result of your search to see if the topic or question you are thinking of has some answers in the source data.

**Note:** Preferably, queries ask about the following: persons, organizations, locations, facilities, events, movements, practices-or-customs, products, publications, laws, awards, diseases, or abstract entities.

### 2.2 Using the search tool

The corpus search tool can be accessed by logging into webann at the following URL: <https://webann ldc.upenn.edu/tools/off/line/search>

To search the corpus, select the index that you want to search, and enter your search terms in the search bar at the top of the interface. A list of results should appear in the form of document IDs. Click on a document ID to view one of the threads returned by the search. You should see the search string entered highlighted in yellow. You can also use CTRL-F in the browser to look for other words and phrases in the thread you are viewing.

**Note:** You can use quotes to search for an exact phrase. For instance, entering the following in the search box:

**“Barack Obama”**

Will only return threads that contain the exact string **Barack Obama**

In contrast, searching without quotes:

Barack Obama

Will return threads that contain the exact string **Barack Obama** as well as partial matches, such as just **Obama** or just **Barack**.

### 3. Developing a full query

#### 3.1 Writing a natural language query

Once you have found a topic or question with at least two different answers, write it out as a natural language query in English.

**Note:** Natural language queries should be a sentence in length, and should be fairly straightforward to read. Natural language queries can be in statement or question form, as long as it is clear that they are asking for information.

#### 3.2 Filling in the query information

##### 3.2.1. Development time

Development time is the amount of time it takes us to develop a query, in minutes. We measure development time from the beginning of the search process to filling in the human answers to a query.

**Note:** Query development time should not exceed 2 hours. See section 2.3 on human answers for more information about the amount of time that should be devoted to looking for human answers.

##### 3.2.2. Description

The Description is short statement of the information we are looking for as an answer to your query. The Description is the basis for the rules of interpretation.

**Note:** We should not include in the Description anything that cannot reasonably be inferred from the natural language query.

##### 3.2.3. Language target

Language target is used for indicating whether we would like query answers to only come from a single language.

If we select “arz” as the Language target, then systems will only look for answers to the query in the Arabic source data. If we select “cmn” as the Language target, then systems will only look for answers to the query in the Chinese source data. If we select “eng” as the Language target, then systems will only look for answers to the query in the English source data.

**Note:** Unlike Phase 1, try not to use a country modifier to force systems to return results from a particular language, instead, use Language target

- P1: What do people say about drinking coffee in China?
- Better: What do people say about drinking coffee? [target language: cmn]

**Note:** If we select “none” as the Language target, then systems will look for answers to our query in all three sets of source data (Arabic, Chinese, and English).

#### **3.2.4. Asks about**

Asks about indicates the topic or type of entity our query asks about. Select this from the drop-down list: persons, organizations, locations, facilities, events, movements, practices-or-customs, products, publications, laws, awards, diseases, abstract entities, or other.

#### **3.2.5. Asks for**

Asks for indicates the type of answer or information your query is asking for. Select this from the drop-down list: statements-or-opinions, causes of, effects of, relationship between, or other.

#### **3.2.6. Languages: arz, cmn, eng**

The Languages columns indicate the languages where we expect answers to the query to be found. These are not a definitive statement that query answers are only found in those languages. Rather, they indicate that that based on our search for human answers, these are the languages represented.

Select either “T” (for True=answers are expected to be found in this language) or “F” (for False=answers are not expected to be found in this language).

**Note:** This is different from Language target, which indicates whether that we are only interested in answers from a particular language.

#### **3.2.7. Rules of Interpretation**

The Rules of interpretation are a set of basic, common-sense rules that will be used to assess system answers to the query. Essentially they are the basic requirements information has to meet to be an answer to the query.

Note: As with the Description, the rules of interpretation should not contain anything that is not reasonably be inferable from the natural language query.

### **3.3 Adding human answers to the query**

Search for answers to your query in the corpus. These can include the answers we saw during the initial search and exploration of our topic or question, but we should also take time to look for additional answers to our query.

**Note:** On average, try not to spend more than an hour looking for answers to a query. If after 15 minutes it is clear that a query is not very productive (i.e. does not have a lot of answers), then feel free to finish it up and move on to the next one. If a query is very productive (i.e. has many answers), you can spend extra time finding and entering answers to it, but do not spend more than 2 hours on a query.

**Note:** Enter each human answer we add should be unique (i.e. we should not add the same answer more than once).

### **3.3.1. Answer Text**

To enter a human answer, copy and paste the text from the discussion forum thread that you think answers the query. Answers should be 1-2 sentences long, as opposed to paragraph length.

### **3.3.2. Answer doc ID**

To enter an answer doc ID, carefully highlight in blue the thread ID in which the answer text occurs and paste it into the spreadsheet. (Be sure to double check that you have copied the correct doc ID).

## **4. Constraints on queries**

### **4.1 Avoid queries with strong lexical triggers**

As a rule, we should avoid writing simple queries that are strongly lexically triggered. A typical example of a strongly lexically triggered query is one that asks about a named entity (e.g. "What are people saying about the Affordable Care Act?").

Strongly lexically triggered queries are less desirable because they are too easy for systems; they can just search the source data for the named entity (e.g. "Affordable Care Act") and come up with lots of answers to the query.

### **4.2 Balancing simple and complex parts of queries**

We should balance simple and complex parts of our queries, so that if we ask about a simple entity, we ask for more complex information as a response (e.g. "Why are people leaving the church of Scientology?").

In contrast, if we ask about a complicated entity, we can ask for more simple information as a response (e.g. "What are people's opinions about Obama's attempt to reform health care?")

**Note:** In general, we should avoid asking simple queries about simple entities (e.g. "What are people saying about the Affordable Care Act").

Note: One way to make a simple entity more complex is to find a way of paraphrasing it (e.g. asking about “Obama’s attempts to reform health care” rather than “the Affordable Care Act”).

### **4.3 Variety in wording**

In general, it is desirable that we vary the language we use to write natural language queries, rather than using the sentence “template” over and over. For example, if we write several queries that asks for statements-or-opinions, we should vary the language we use to ask for statements or opinions (e.g. “What are people saying about ..”, “What do people think about ..” etc.)

**Note:** Unlike Phase 1, we can write queries that use Yes/No question form (e.g. “Is it safe to travel to Mexico?”) and systems will understand that these queries are actually asking for richer answers than a simple “yes” or “no”.

### **4.4 No “How do I ..” queries ..**

Unlike Phase 1, we should not write queries this year that ask how one does something (e.g. “How does one become a dentist?”) as these are too difficult for systems.

## **5. Appendix: Some example natural language queries**

Are teenage boys dangerous drivers?

What did Anders Breivik say about the Norway massacre?

What were the effects on New York City of the smoking ban in restaurants?

What were the effects of the New York City smoking ban on the restaurant business?

Describe Nicolas Sarkozy's relations with Turkey.

What did the New York Times say about open-source journals?

Which problems arose due to Newt Gingrich's welfare reforms?

Why is Netanyahu attacking Iran?

What is the conflict between Azerbaijan and Georgia about?

How did the U.S. government react to the Deepwater horizon oil spill?