

Table of Contents

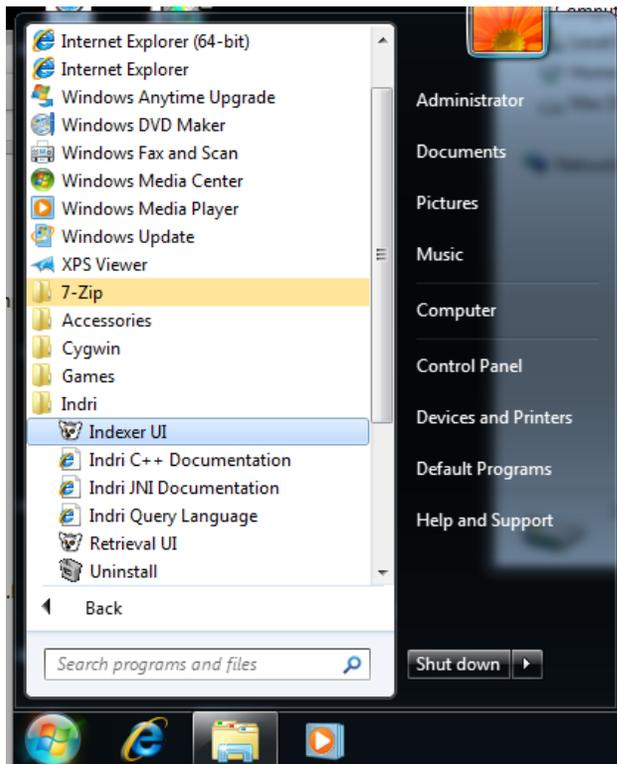
Running Indri	2
Indexing	2
Indexing - Alternate Approach.....	4
Searching with Indri	5
Running OpenEphyra	8
Allowing OpenEphyra to support French questions.....	9
Running MEAD	10

Running Indri

Using Indri is composed of two parts. First indexing a corpus that will be used for searching. Then searching that corpus.

Indexing

1. In a web browser open <http://scaiano.com/edu>
2. Download and run AutoSetup.bat
3. To run the Indexer UI, from the start menu click Programs > Indri > Indexer IU



4. In the first screen set the following settings:

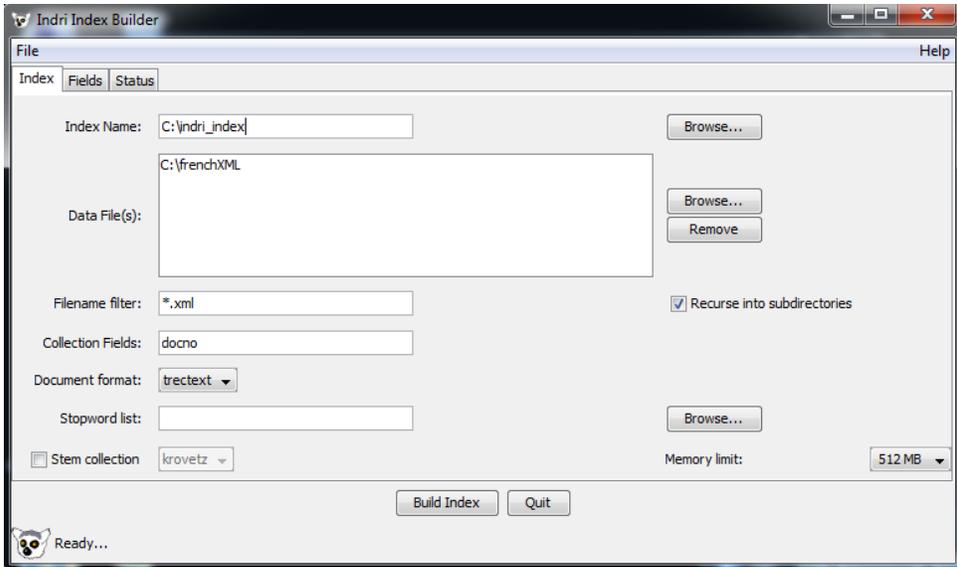
Index Name – this is where the index will be store. Both the *Retrieval UI* and *OpenEphyra* will require this path.

Data File(s) – this is the location of the corpus to be indexed. Do not add each file, just the folder that contains all the files.

Recurse into subdirectories – check this box.

Document Format: trext – the corpus has been prepared in this format.

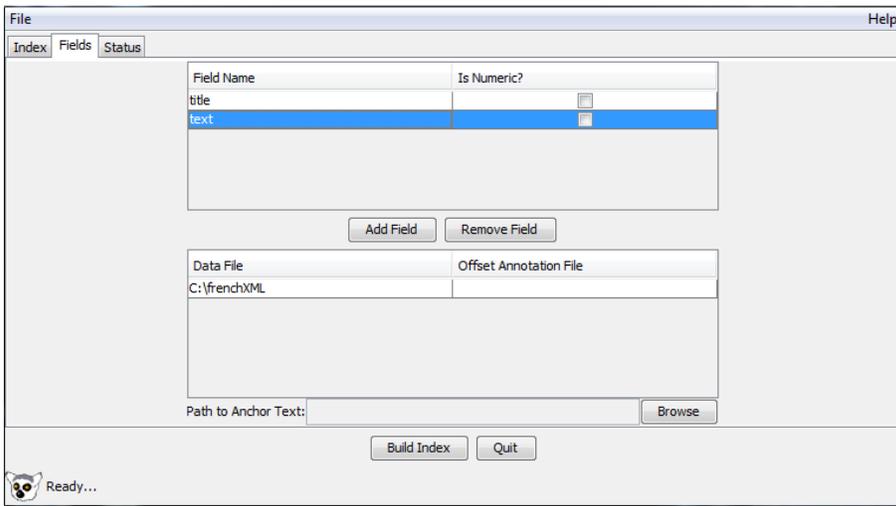
Stemmer: none or uncheck the box– The stemmer provided are only intended for English.



5. Switch to the field tab and set the following settings

Add Field > title

Add Field > text



6. Click Build Index

Indexing – Alternate Approach

These are the steps use to build the index that OpenEphyra uses.

1. Open a Command Line
 - a. Click the start menu
 - b. Type: **cmd** in the search box
 - c. Hit Enter



2. Move into the frenchXML directory
cd \frenchXML
3. Download scaiano.com/download/french.indri.config.xml file and place it frenchXML directory.
4. Open french.indri.config.xml in notepad
5. Change the corpus and index paths to match those required for your system

```
<corpus>
```

```
  <path>/Users/martinscaiano/dev/data/french_corpus/corpus/Trec  
format/frenchXML</path>
```

```
  <class>trectext</class>
```

```
</corpus>
```

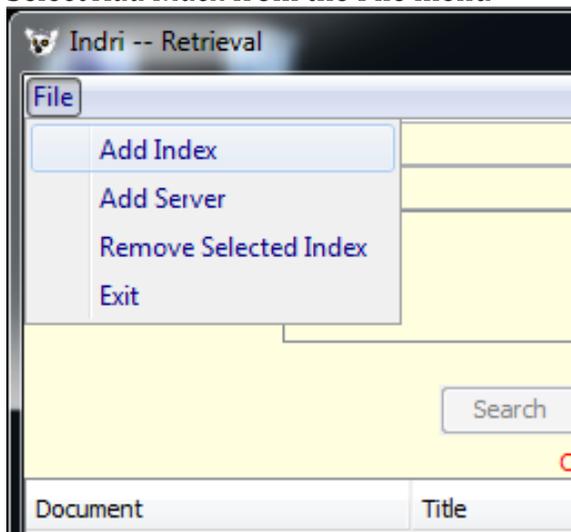
```
<memory>1g</memory>
```

```
<index>/Users/martinscaiano/dev/data/french_corpus/corpus/Trec
format/indri_index</index>
```

6. Type: **IndriBuildIndex** french.indri.config.xml
This will build the index

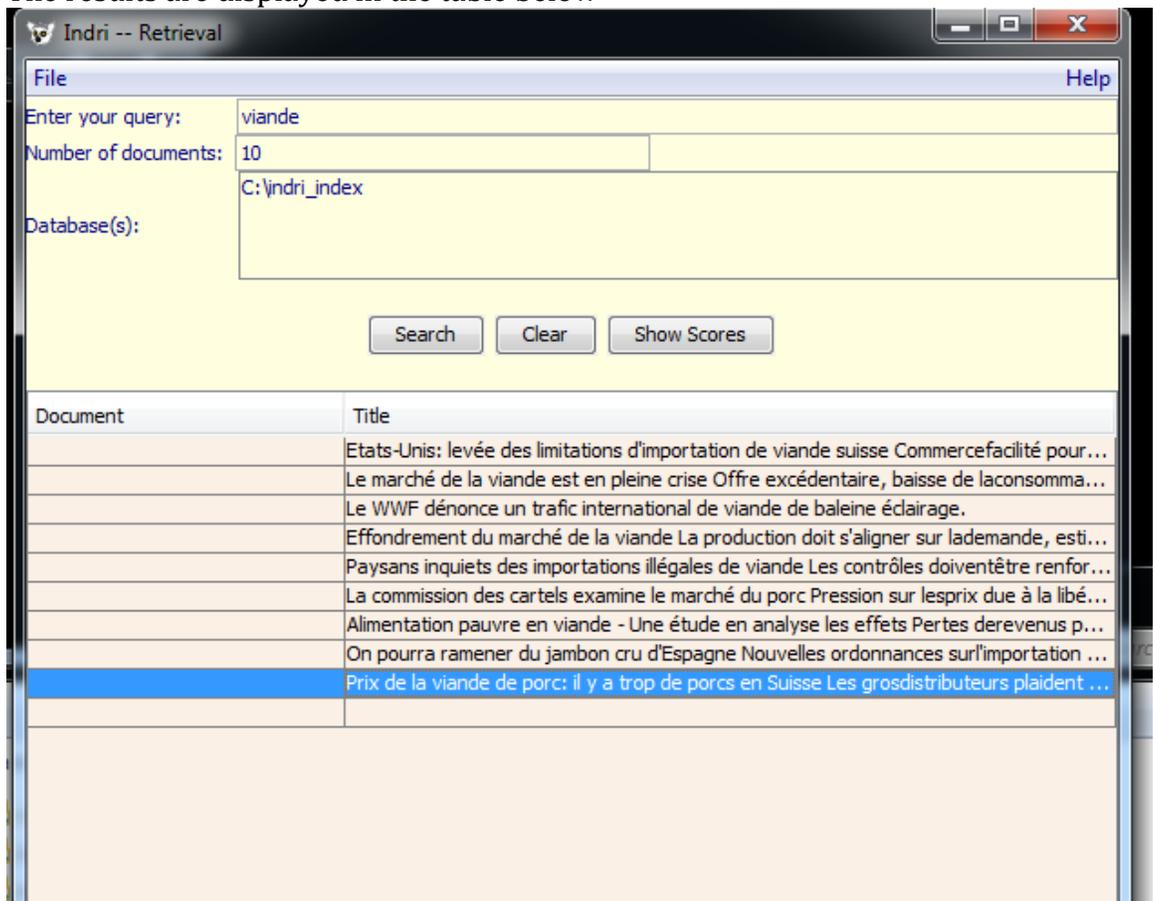
Searching with Indri

1. Start the *Retrieval UI*
From the start menu Programs > Indri > Retrieval UI
2. Select *Add Index* from the File menu



3. Select where you created the Indri index.
4. Type search terms into the *Enter your query* field:
Either hit *enter* or click *Search*

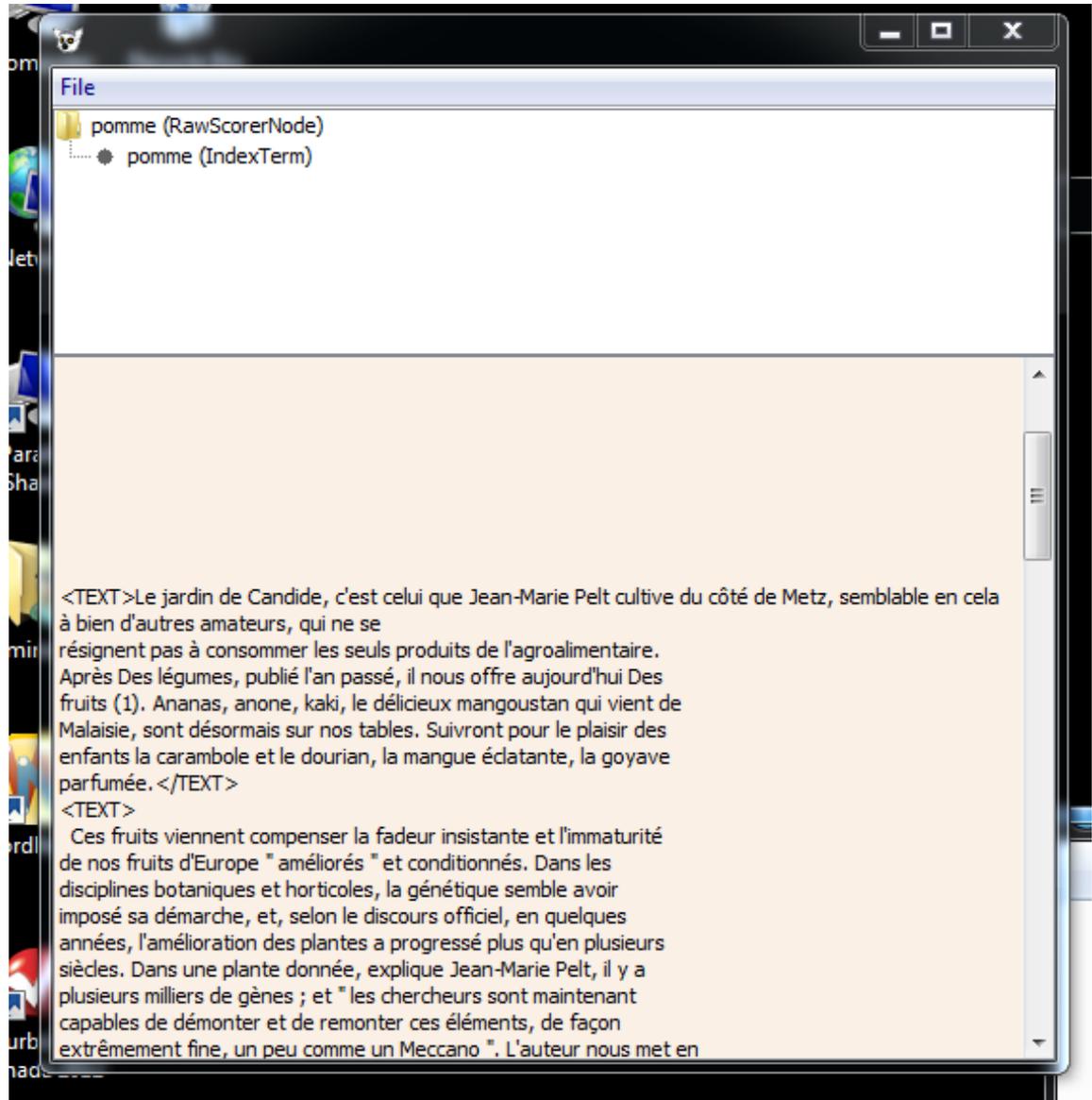
5. The results are displayed in the table below



The screenshot shows the 'Indri -- Retrieval' application window. The interface includes a menu bar with 'File' and 'Help', a search input field containing 'viande', a 'Number of documents' field set to '10', and a 'Database(s)' field containing 'C:\indri_index'. Below these fields are 'Search', 'Clear', and 'Show Scores' buttons. The results are displayed in a table with two columns: 'Document' and 'Title'. The table contains ten rows of search results, with the last row highlighted in blue.

Document	Title
	Etats-Unis: levée des limitations d'importation de viande suisse Commercefacilité pour ...
	Le marché de la viande est en pleine crise Offre excédentaire, baisse de laconsomma...
	Le WWF dénonce un trafic international de viande de baleine éclairage.
	Effondrement du marché de la viande La production doit s'aligner sur lademande, esti...
	Paysans inquiets des importations illégales de viande Les contrôles doiventêtre renfor...
	La commission des cartels examine le marché du porc Pression sur lesprix due à la libé...
	Alimentation pauvre en viande - Une étude en analyse les effets Pertes derevenus p...
	On pourra ramener du jambon cru d'Espagne Nouvelles ordonnances surl'importation ...
	Prix de la viande de porc: il y a trop de porcs en Suisse Les grosdistributeurs plaident ...

6. Click on a result to see the full document
A dialog with the original document will pop up.
The document will highlight where the search terms appeared. Note this does not work in French because of accent characters.



Running OpenEphyra

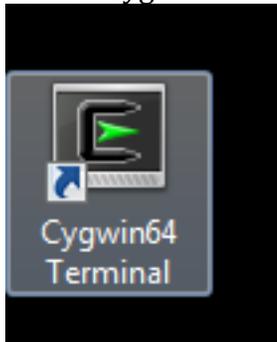
Then

1. Open an explorer window (⌘ + E) to Downloads
2. Double click **launchOE.bat**
3. Type: **c:\runOE.bat**
4. Wait for: **Question:**
5. Type your Question: **Who is Colin Powel?**
Note the start of the question must be in English "Who is" because OpenEphyra was only designed for use in English.

Running MEAD

Mead must be run from the Cygwin command line.

1. Launch Cygwin



2. Type: `cp /cygdrive/c/meadrc /cygdrive/q/Mead-3.12/mead/data/mlp2011/meadrc`
3. Move to the mead directory. Type: `cd /cygdrive/q/Mead-3.12/mead`
4. Run Mead. Type
`bin/mead.pl -rc data/mlp2011/meadrc -o test.txt`
`data/mlp2011/mlp000`

`bin/mead.pl` - call the mead summarizer

`-rc data/mlp2011/meadrc` - indicates to use the meadrc configuration file.
You can edit this file to change the output of mead

`-o test.txt` - stores the summary - you can view the summary by typing: `more test.txt`

`data/mlp2011/mlp000` - indicates to use the mlp2011 data set (or cluster).
There are clusters ranging from `mlp000` to `mlp009`

5. The summary can be view by typing: `more test.txt`