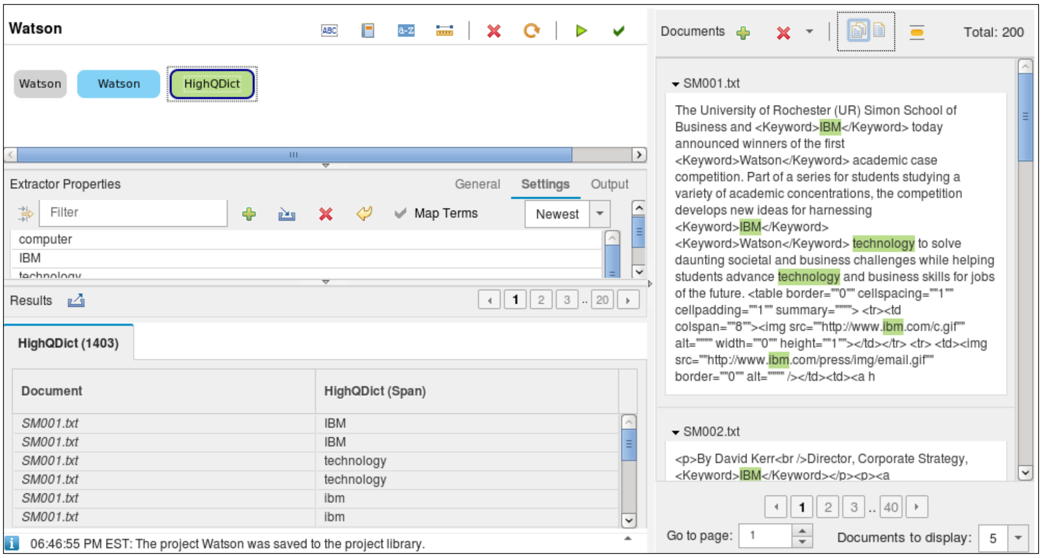
IBM Training

Demonstration 1

Creating dictionaries for your Watson project



computer
IBM
technology

HighQDict (1403)

Document	HighQDict (Span)
SM001.txt	IBM
SM001.txt	IBM
SM001.txt	technology
SM001.txt	technology
SM001.txt	ibm
SM001.txt	ibm

06:46:55 PM EST: The project Watson was saved to the project library.

SM001.txt

The University of Rochester (UR) Simon School of Business and <Keyword>IBM</Keyword> today announced winners of the first <Keyword>Watson</Keyword> academic case competition. Part of a series for students studying a variety of academic concentrations, the competition develops new ideas for harnessing <Keyword>IBM</Keyword> <Keyword>Watson</Keyword> technology to solve daunting societal and business challenges while helping students advance technology and business skills for jobs of the future. <table border=""0"" cellspacing=""1"" cellpadding=""1"" summary=""""> <tr><td colspan=""8""><img src=""http://www.ibm.com/c.gif"" alt="""" width=""0"" height=""1""></td></tr> <tr> <td><img src=""http://www.ibm.com/press/img/email.gif"" border=""0"" alt="""" /></td><td><a h

SM002.txt

<p>By David Kerr<br />Director, Corporate Strategy, <Keyword>IBM</Keyword></p><p><a

Annotation Query Language (AQL)

Demonstration 1: Creating dictionaries for your Watson project

Demonstration 1: Working with basic AQL features using the Web UI

Purpose:
The purpose of the demo is to get you started with the extractors using the Web UI. You will build dictionaries using the clues that you identified in the previous demo. Everything is done by dragging and dropping the extractors onto the canvas.

User ids / Passwords

OS:	**biadmin/biadmin**
Root:	**root/dalvm3**
Ambari:	**admin/admin**
BigInsights Home:	**guest/guest-password**

Ambari Services Required:

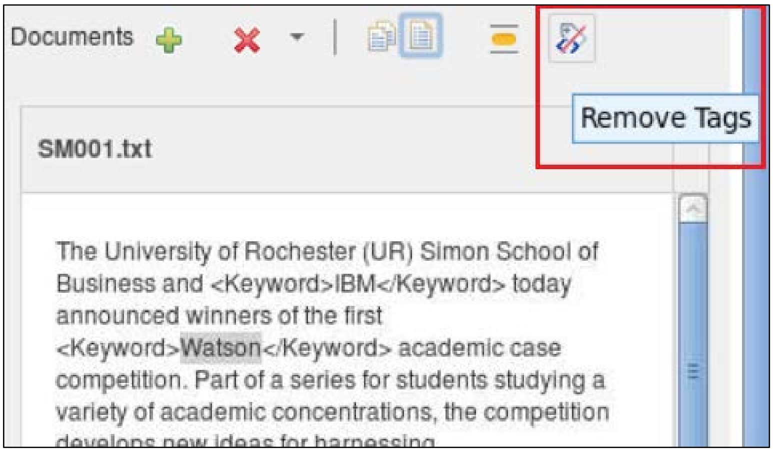
- HDFS
- MapReduce2
- YARN
- Knox (also start the Demo LDAP service)
- BigInsights - Text Analytics
- BigInsights - Home

Task 1. Creating and running a literal.

1. Continue on with the same project, **Watson**.
2. On the toolbar on top of the canvas, click the **New Literal** button.
3. Type the word **Watson** in the literal extractor**.**
4. With the literal selected, click the **Run** button (green arrow).

5. It should return 298 rows where the word Watson was found within the documents.

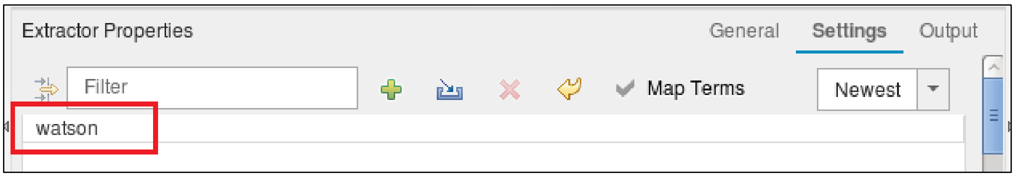
   If you do not get 298 rows, it's probably because you didn't remove the tags when you loaded the document. Remove the tags by clicking on the Remove tag button:



The University of Rochester (UR) Simon School of Business and <Keyword>IBM</Keyword> today announced winners of the first <Keyword>Watson</Keyword> academic case competition. Part of a series for students studying a variety of academic concentrations, the competition develops new ideas for harnessing

Task 2. Creating and running a dictionary.

1. Create a new dictionary. Click the **New Dictionary** button.
2. Name the dictionary: **WatsonDict**
3. Add the word Watson. Type **Watson** in the field under the **Extractor Properties.**

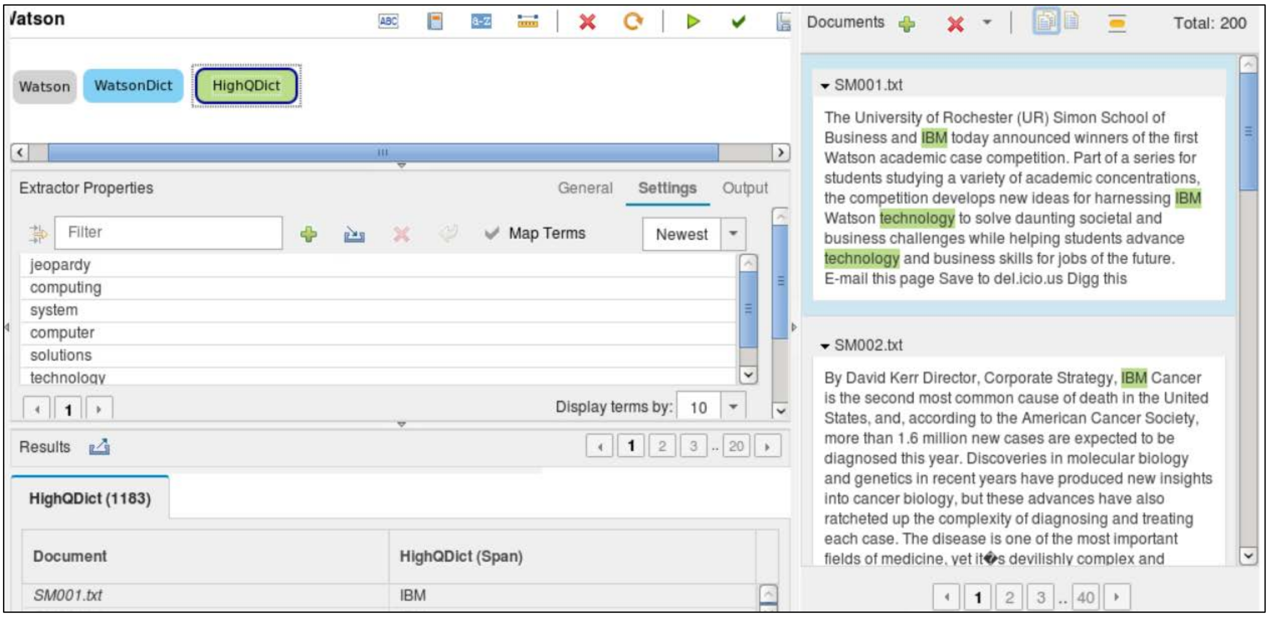


watson

4. Run the extractor by clicking on the **green arrow**.
5. The results returned should be 298 rows as well, since we searched on the exact same term.
6. Now create another dictionary called **HighQDict**, with all the terms you identified in the previous lab. Add the word **computing** to the list as well. Here they are:

   Positive clues: IBM, technology, solutions, computer, system, jeopardy, computing.
7. Run the HighQDict extractor.

8. You should get 1183 results returned:



jeopardy
computing
system
computer
solutions
technology

HighQDict (1183)

Document	HighQDict (Span)
SM001.txt	IBM

SM001.txt

The University of Rochester (UR) Simon School of Business and IBM today announced winners of the first Watson academic case competition. Part of a series for students studying a variety of academic concentrations, the competition develops new ideas for harnessing IBM Watson technology to solve daunting societal and business challenges while helping students advance technology and business skills for jobs of the future. E-mail this page Save to del.icio.us Digg this

SM002.txt

By David Kerr Director, Corporate Strategy, IBM Cancer is the second most common cause of death in the United States, and, according to the American Cancer Society, more than 1.6 million new cases are expected to be diagnosed this year. Discoveries in molecular biology and genetics in recent years have produced new insights into cancer biology, but these advances have also ratcheted up the complexity of diagnosing and treating each case. The disease is one of the most important fields of medicine, yet it�s devilishly complex and

Results:
you should be able to build dictionaries using the clues that you identified in the previous example.