

Chapter 11: Retrieve And Rank

Learning Bluemix & Cognitive

Bob Dill, IBM Distinguished Engineer, CTO Global Technical Sales

Chapter 11: Using Retrieve and Rank

- Retrieve:
 - What will we retrieve?
 - Using pdf documents from the IBM Institute for Business Value
 - Document conversion
 - What's in a document?
 - Review, getting rid of the 'less than useful' content
 - Indexing and setting up the 'Retrieve' service
 - Testing Retrieve
- Ranking
 - Questions ... we need 5X as many questions as we have documents
 - Associating questions with document sections
 - Training
 - Getting Ranked results.



Documents

- We can read, index, and search almost anything. Notably:
 - PDF, HTML, Word
- For this exercise, we're using the pdf documents located here:
 - <http://www-935.ibm.com/services/us/gbs/thoughtleadership/>
- There are 11 documents (as of November 2016) at this site and we'll use all of them.
- Our first step is to 'convert' the documents.
- Then we'll need to add them to a collection.



Watson Retrieve Collections

- First, create a Retrieve and Rank Service instance and get your credentials
- Based on Apache Solr
 - Create a Solr cluster:
 - `curl -X POST -u "{username}":"{password}" "https://gateway.watsonplatform.net/retrieve-and-rank/api/v1/solr_clusters" -d`
 - Define the configuration you'll use for the collection you're about to create. We'll use "config_1.zip"
 - `curl -X POST -H "Content-Type: application/zip" -u "{username}":"{password}" "https://gateway.watsonplatform.net/retrieve-and-rank/api/v1/solr_clusters/YOUR-CLUSTER-ID-GOES-HERE/config/config_1" --data-binary @HTML/IBV_Conversion/config_1.zip`
 - Create a Solr Collection, ours will be called Z2C_IBV_Articles:
 - `curl -X POST -u "{username}":"{password}" "https://gateway.watsonplatform.net/retrieve-and-rank/api/v1/solr_clusters/YOUR-CLUSTER-ID-GOES-HERE/solr/admin/collections" -d "action=CREATE&name=Z2C_IBV_Articles&collection.configName=config_1&wt=json"`
- Convert documents and make a combined document list
- Review the combined list and remove trivial sections
- Index the resulting documents into the collection
 - `curl -X POST -H "Content-Type: application/json" -u "{username}":"{password}" "https://gateway.watsonplatform.net/retrieve-and-rank/api/v1/solr_clusters/YOUR-CLUSTER-ID-GOES-HERE/solr/Z2C_IBV_Articles/update" --data-binary @HTML/IBV_Conversion/docList.json`

Each curl command is executed from the Chapter11 folder



The Plan: 30 minute Chapters with an hour or two of practice

- | | |
|--|---|
| 1. The Story, Architecture for this app | |
| 2. Setting up Bluemix | |
| 3. Building your first Watson App | (Watson Speech to Text) |
| 4. Getting Watson to talk back | (Watson Text to Speech) |
| 5. Understanding Classifiers | (Watson NLC) |
| 6. Creating a custom dialog with Watson | (custom Q&A, session management) |
| 7. Authentication | (puts C2 thru 6 together) |
| 8. Alchemy News | (Watson Alchemy) |
| 9. Visual Recognition and Images | (Watson Visual Recognition) |
| 10. Watson Conversations | (Watson Conversations) |
| 11. Rank & Retrieve | (Retrieve & Rank) |
| 12. Getting started on my first client prototype | Design Thinking, Stories, Architecture, Keeping it simple |



Chapter 12: Getting started on my first client prototype

Learning Bluemix & Cognitive

Bob Dill, IBM Distinguished Engineer, CTO Global Technical Sales