Create Field Tag Design Document

# Overview

1. We have an unstructured MongoDB data set containing 6400 documents. Within each of those 6400 documents is a field named “Party1” which contains short string values.
2. We also have a .CSV file containing ~10,000 *Keywords*. In each row we are concerned with Column A titled “*Keyword”*. This “*Keyword*” field represents a name. This .CSV is basically a list of 10,000 names.

# Objective

Create a python script, that goes through each of the 6400 document's "*Party1*" fields and when there is a match for the any of the 10,000 keywords (within the .CSV master list) the program needs to generate a "tag" object and insert it into the "***Party1Details***" array within that document of the DB.

Within each of our 6400 documents in our MongoDB we have an array called "***Party1Details***". This array contains a list of "tag" objects. Each tag is made up of 4 string fields. An example of one of these “tag” object is in the screenshot below.

**\*NOTE\***

**\*\*\***The final product should be modular in the sense that the target field (Party1), the tag destination field (Party1Details), the field names (keyword, firstName, lastName, and gender), and the csv (master.csv) can be customized and set within a simply python settings file.**\*\*\***

## Logic

Each time we search through a document's "Party1" value for each of the 10,000 keywords and successfully find one, we must create an object like the one seen below and insert it into the "***Party1Details***" array.



Example of a “tag” object

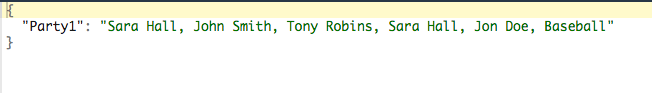
Figure : tag object

As we can see in the screenshot above, there are 4 fields within each of these “tag” objects. Please see the attached spreadsheet ***“Field Tag Description.xls”*** for details regarding each field.

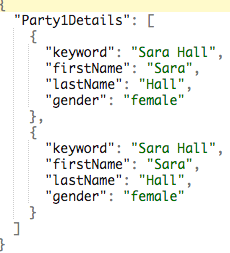
# Important Note

If you have multiple instances of the same keyword within a single document’s “Party1” field, you must create a new, independent "tag".  
  
For example. If the keyword is "Sara Hall" and it appears twice within the same "Party1" value of a single document, then you will generate two "tags" as seen in the screenshot below.

**IF:**

****

**THEN:**



# Conclusion

That is all the information you should need to complete this project. There are 3 working documents you will need to get started.

1. Copy of **Dev MongoDB**
   * I will share this via dropbox on UpWork.
2. **master.CSV**
   * A copy of the csv file containing a subset of the 10,000 keywords. This will also be shared via Dropbox on UpWork chat.
3. **Tag Object Description.xls**
   * This spreadsheet shows a breakdown of a single keyword “tag”. It gives you a detailed explanation of what each line in the .JSON object represents.