XML vs. SQL Bake-off - November 4, 2013

Welcome to the XML vs. SQL bake-off, where you and your group will race other groups in answering a set of queries!

The data set is given to you in JSON, and you will have to choose to convert it to either XML or SQL for querying. (Hint: look at the queries before deciding whether to use XML or SQL!)

We've provided starter code in python to help you create an XML file or a CSV file to load into SQLite. Note that this code is only a starting point; you will still have to write the code that actually transforms the data. Everything is in place for you to run your final python program and queries on one of the cluster machines (Myth or Corn). However, you are also free to run everything on your local machine instead. If you do this, make sure you have python, plus SQLite (for SQL) or Saxon or Kernow (for XML).

Start by copying the files below into a directory of your choice (or your own machine):

> Data set: `/usr/class/cs145/bakeoff/bakeoff_cookbook.json`
> Prettier version: `/usr/class/cs145/bakeoff/bakeoff_cookbook_pretty.json`
> Starter code: `/usr/class/cs145/bakeoff/bakeoff_starter.py`

If you are developing on Myth or Corn **AND** you have chosen XML, you should also copy this file:

> Saxon JAR file: `/usr/class/cs145/bakeoff/saxon9he.jar`

To run your data transformation program, type:

```
python bakeoff_starter.py
```

Once you have your data in a SQL database (e.g. final.db) or XML file (e.g. final.xml), use the commands below to run your queries:

> For SQL:
> ```
> sqlite3 final.db < query1.sql
> ```
>
> For XML (make sure you run this command from the directory where final.xml lives):
> ```
> java -cp saxon9he.jar net.sf.saxon.Query query1.xq
> ```

Queries:
1. Find the titles of all recipes in the breakfast food section.
2. Find the number of recipes with more than 5 ingredients.
3. Find the average number of ingredients needed for a recipe.
4. Find the recipe(s) with the fewest number of ingredients.
5. Find the average number of ingredients needed for recipes requiring butter.
6. For recipes with more than 4 ingredients, find the 4th ingredient.

Once you're finished, bring your solutions up to the TAs to have them checked!