FooDIS

Food.com Data, Information and Search

Project Milestone 3
PRI@FEUP - 2021/2022

António Bezerra
Gonçalo Alves
Pedro Seixas

Our data

Collection

- Using a single **recipes** collection with **506k** documents

- Each document is a recipe:
  - Recipe information
  - Reviews

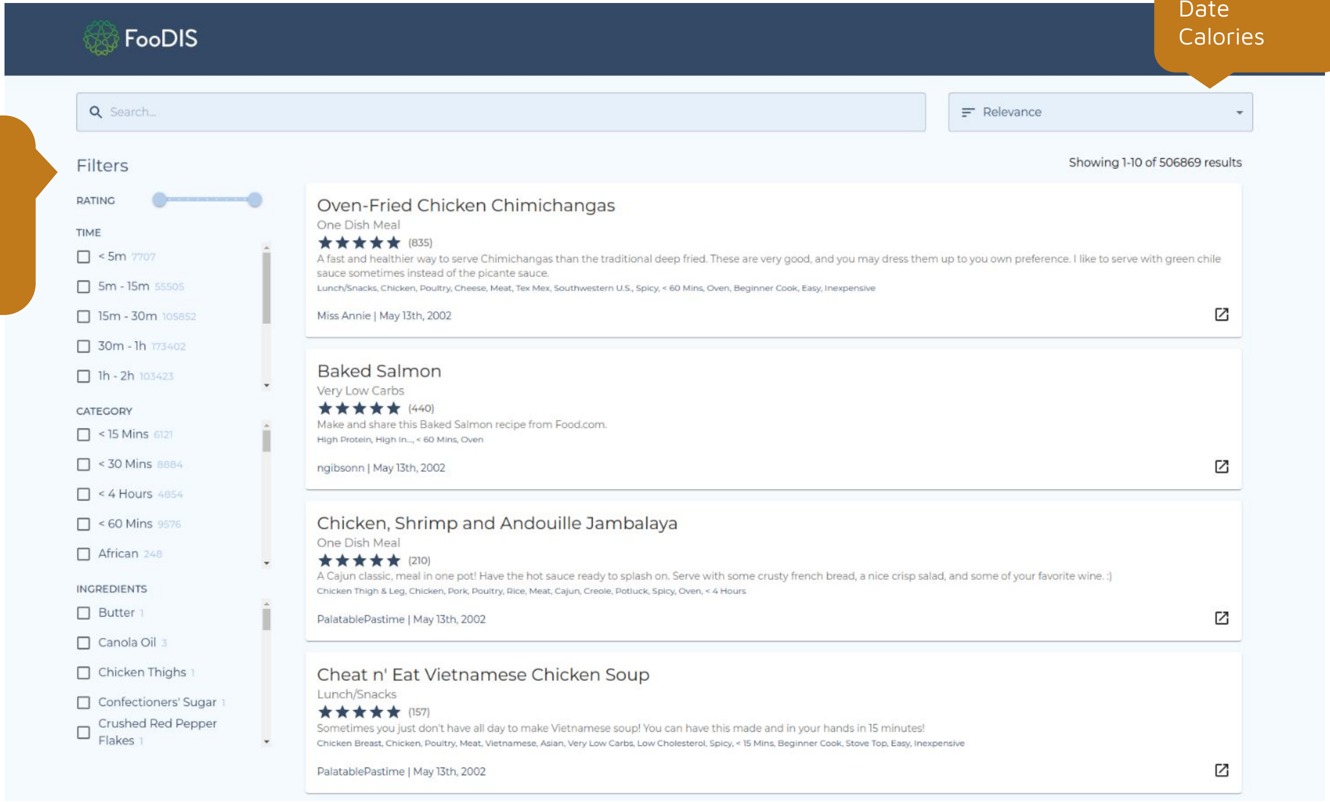
- **Most relevant fields:**
  - Name
  - Description
  - Category
  - Keywords
  - Ingredients
  - Author Name
  - Nutritional Content
  - Time
  - Date
  - Reviews

User Interface

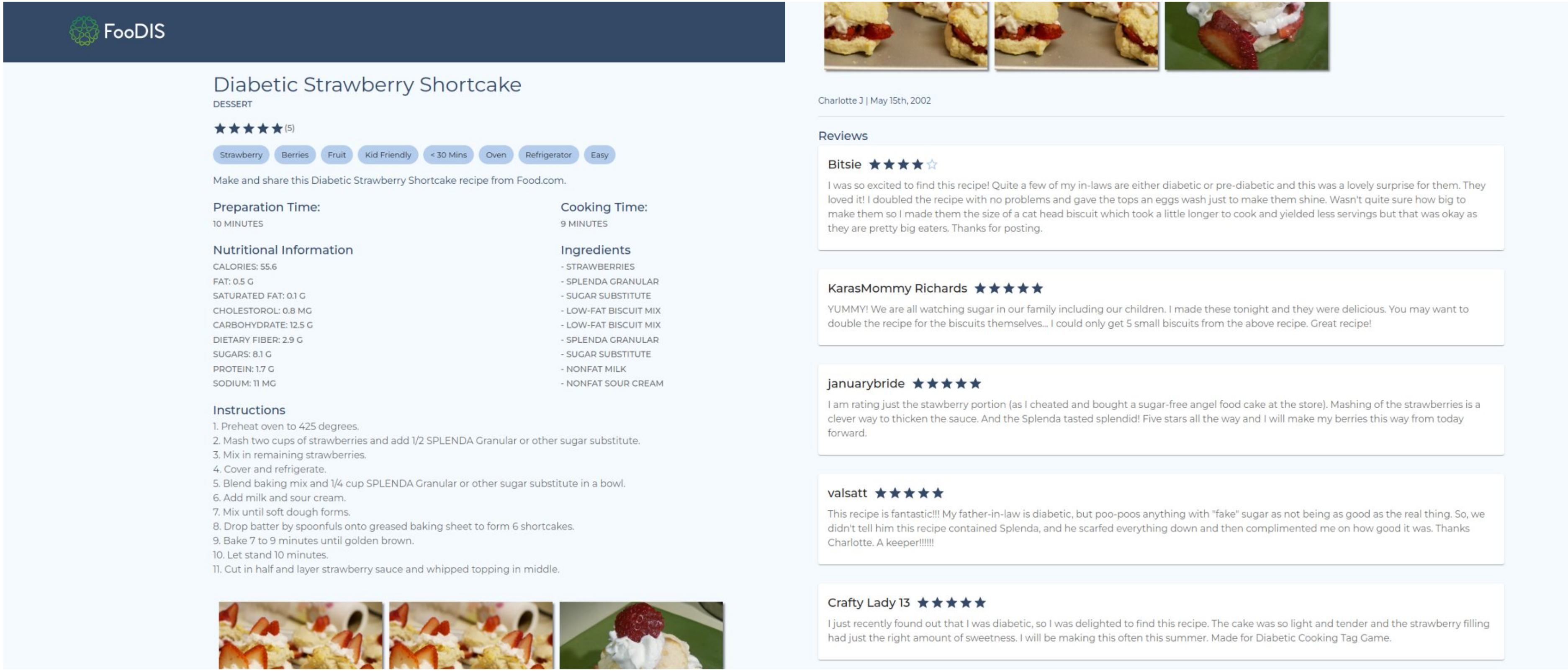
Searching for recipes

Rating
Time
Category
Ingredients

Relevance
Date
Calories



Recipe page



FooDIS
Diabetic Strawberry Shortcake
DESSERT
(5)
Strawberry
Berries
Fruit
Kid Friendly
< 30 Mins
Oven
Refrigerator
Easy
Make and share this Diabetic Strawberry Shortcake recipe from Food.com.
Preparation Time:
10 MINUTES
Cooking Time:
9 MINUTES
Nutritional Information
CALORIES: 55.6
FAT: 0.5 G
SATURATED FAT: 0.1 G
CHOLESTOROL: 0.8 MG
CARBOHYDRATE: 12.5 G
DIETARY FIBER: 2.9 G
SUGARS: 8.1 G
PROTEIN: 1.7 G
SODIUM: 11 MG
Ingredients
- STRAWBERRIES
- SPLENDA GRANULAR
- SUGAR SUBSTITUTE
- LOW-FAT BISCUIT MIX
- LOW-FAT BISCUIT MIX
- SPLENDA GRANULAR
- SUGAR SUBSTITUTE
- NONFAT MILK
- NONFAT SOUR CREAM
Instructions
1. Preheat oven to 425 degrees.
2. Mash two cups of strawberries and add 1/2 SPLENDA Granular or other sugar substitute.
3. Mix in remaining strawberries.
4. Cover and refrigerate.
5. Blend baking mix and 1/4 cup SPLENDA Granular or other sugar substitute in a bowl.
6. Add milk and sour cream.
7. Mix until soft dough forms.
8. Drop batter by spoonfuls onto greased baking sheet to form 6 shortcakes.
9. Bake 7 to 9 minutes until golden brown.
10. Let stand 10 minutes.
11. Cut in half and layer strawberry sauce and whipped topping in middle.
Charlotte J | May 15th, 2002
Reviews
Bitsie
I was so excited to find this recipe! Quite a few of my in-laws are either diabetic or pre-diabetic and this was a lovely surprise for them. They loved it! I doubled the recipe with no problems and gave the tops an eggs wash just to make them shine. Wasn't quite sure how big to make them so I made them the size of a cat head biscuit which took a little longer to cook and yielded less servings but that was okay as they are pretty big eaters. Thanks for posting.
KarasMommy Richards
YUMMY! We are all watching sugar in our family including our children. I made these tonight and they were delicious. You may want to double the recipe for the biscuits themselves... I could only get 5 small biscuits from the above recipe. Great recipe!
januarybride
I am rating just the stawberry portion (as I cheated and bought a sugar-free angel food cake at the store). Mashing of the strawberries is a clever way to thicken the sauce. And the Splenda tasted splendid! Five stars all the way and I will make my berries this way from today forward.
valsatt
This recipe is fantastic!!! My father-in-law is diabetic, but poo-poos anything with "fake" sugar as not being as good as the real thing. So, we didn't tell him this recipe contained Splenda, and he scarfed everything down and then complimented me on how good it was. Thanks Charlotte. A keeper!!!!!!
Crafty Lady 13
I just recently found out that I was diabetic, so I was delighted to find this recipe. The cake was so light and tender and the strawberry filling had just the right amount of sweetness. I will be making this often this summer. Made for Diabetic Cooking Tag Game.

System Improvements

Boosting

Milestone 2

Query Level Boosting:
- **Pros:** Better results
- **Cons:** How to get the best combination of boosts for each possible query? And is it possible to do it?

Milestone 3

System Level Boosting:
- **Pros:** Actually feasible
- **Cons:** Not optimal for the majority of queries

Boosting

Milestone 2

Query Level Boosting:

- **Pros:** Better results
- **Cons:** How to get the best combination of boosts for each possible query? And is it possible to do it?

Milestone 3

System Level Boosting:

- **Pros:** Actually feasible
- **Cons:** Not optimal for the majority of queries

Name^5, Category^2, Ingredients^2, Keywords^2, Reviews^0.5, AuthorName^0.2

Boosting

Milestone 2

Query Level Boosting:
- **Pros:** Better results
- **Cons:** How to get the best combination of boosts for each possible query? And is it possible to do it?

Milestone 3

System Level Boosting:
- **Pros:** Actually feasible
- **Cons:** Not optimal for the majority of queries

But can we make adjustments on a query level, without having to manually boost the system?

Yes! With LTR

Learning to Rank: LTR

- **Learning to Rank** is an application of **machine learning** to **ranking models.**
- Solr has a plugin that enables the use of these models for **reranking** the top N results obtained by the search engine.
- **Features** are extracted from the results to be reranked, are fed to a model and a new ordering is made based on the model's output.
- It is also possible to feed **external information** (i.e. not in the document) to the model for computing certain features (for example: user age can be taken as a feature to target results to a certain demographic).
- **In our application** we will be testing out the application of a **simple LTR model**.

LTR: Training data

- **Machine learning** models need **training data**.
- **Training data** must provide information about a **document's relevance** to a given query, and also the **document's features.**
- **How do we calculate a document's relevance?**
  - **Using humans:** have judges look at query results and manually classify them.

    Pros: Generates good quality and clean data.

    Cons: Is expensive to do at scale (solution: crowd-sourcing).
  - **Using analytics:** infer relevance from clicks, retention time, etc.

    Pros: Generates a lot of data for cheap, less bias.

    Cons: The data is noisy and may not be very informative.

LTR: Training data

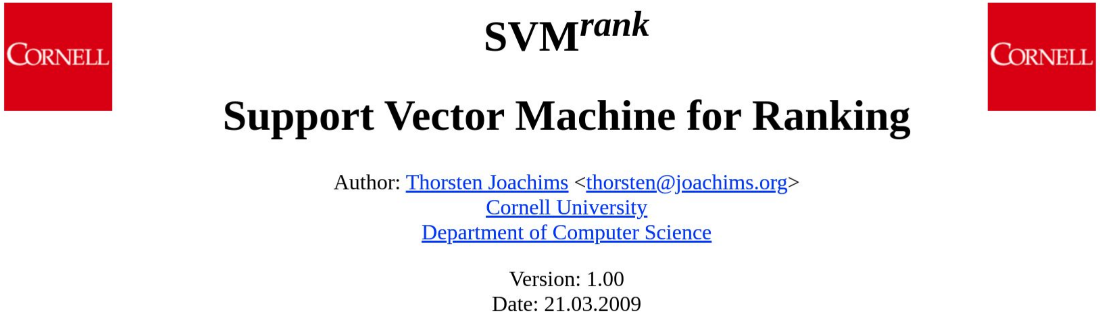
- We relied on **manual classification** of query results for a set of 5 queries.
- Each result was ranked on a scale from 0 to 5 according to established criteria.
- The queries chosen were:
  - "Apple pie"
  - "Chicken" in the African category
  - "Easy bread"
  - "Pasta bolognese"
  - "Oatmeal"
- For each query, the top 100 results were classified, generating **462 data points**.
  (One of the queries only returned 62 results)

LTR: Features

Feature	Description
queryMatch<field>	Set of features that are the score of only matching the query with the Name, Description, Category, Ingredients, Keywords, Instructions, Reviews and AuthorName fields, individually
recency	Measure of how recent the recipe is
rating	Recipe's Rating
<field>Count	Set of features that counts the number of reviews, keywords, ingredients, instructions and images
originalScore	The original score returned by Solr's boosted ranking

LTR: Model

- Solr’s LTR implementation supports three kinds of model: Linear, Multiple Additive Tree and Neural Network.
- We opted to work with **linear models** for their simplicity and higher availability of resources.



SVM^{rank}

Support Vector Machine for Ranking

Author: Thorsten Joachims <thorsten@joachims.org>
Cornell University
Department of Computer Science

Version: 1.00
Date: 21.03.2009

LTR: Model

Feature	Description	Weight
queryMatchName	Score of only matching the query with the Name field	-0.61264533
queryMatchDescription	Score of only matching the query with the Description field	-0.016012382
queryMatchCategory	Score of only matching the query with the Category field	0.40750793
queryMatchIngredients	Score of only matching the query with the Ingredients field	0.10536964
queryMatchKeywords	Score of only matching the query with the Keywords field	0.23444816
queryMatchInstructions	Score of only matching the query with the Instructions field	-0.030611534
queryMatchReviews	Score of only matching the query with the Reviews field	0.29199001
queryMatchAuthorName	Score of only matching the query with the AuthorName field	-0.88715649
recency	Measure of how recent the recipe is	-0.43097624
rating	Recipe's Rating	0.84916848
reviewCount	Number of reviews	1.3468372
keywordCount	Number of keywords	1.6347141
ingredientCount	Number of ingredients	1.6268826
instructionCount	Number of instructions	0.045574259
imageCount	Number of images	-5.1624022
originalScore	The original score returned by Solr's boosted ranking	0.29625478

LTR: Evaluation

- Evaluation Measures
  - **Average Precision**
  - **Recall**
  - **Expected Reciprocal Rank (ERR)**
    This measure is based on the cascade user model.
    It predicts the value 1/s, where s is the position at which a result that satisfies the user's query will be found.
- To evaluate the LTR model, we evaluated a new set of **testing queries**.
- These are queries on which the **model was not trained on**.
- The **top 10 results** returned by each system were classified.

LTR: Evaluation

Query	System	Ranks										AP	Recall	ERR
		1	2	3	4	5	6	7	8	9	10			
pumpkin pie	**Regular**	3	4	3	2	2	2	2	2	3	2	1.0	1.0	0.46
	Enhanced	4	3	3	3	4	5	5	3	3	5	1.0	1.0	0.62 ($\Delta = 0.16$)
vegetable curry	**Regular**	4	2	3	3	4	3	4	2	3	1	1.0	1.0	0.59
	Enhanced	4	3	3	4	3	3	3	3	3	4	1.0	1.0	0.62 ($\Delta = 0.03$)
mojito	**Regular**	4	4	5	2	2	1	2	2	1	2	1.0	1.0	0.69
	Enhanced	5	4	5	3	4	4	4	4	2	5	1.0	1.0	0.98 ($\Delta = 0.29$)
noodles	**Regular**	4	2	2	2	2	3	4	5	3	2	1.0	1.0	0.58
	Enhanced	5	3	4	3	5	4	3	4	4	4	1.0	1.0	0.98 ($\Delta = 0.40$)
beef stew	**Regular**	3	2	2	3	4	1	2	2	2	2	1.0	1.0	0.37
	Enhanced	5	3	5	3	4	3	3	3	5	4	1.0	1.0	0.98 ($\Delta = 0.61$)

LTR Evaluation

System	MAP	ERR
Regular (w/ Boosts)	1.0	0.54
Enhanced (w/ Boosts & LTR)	1.0	0.84

Questions?