

ProjectNomNom



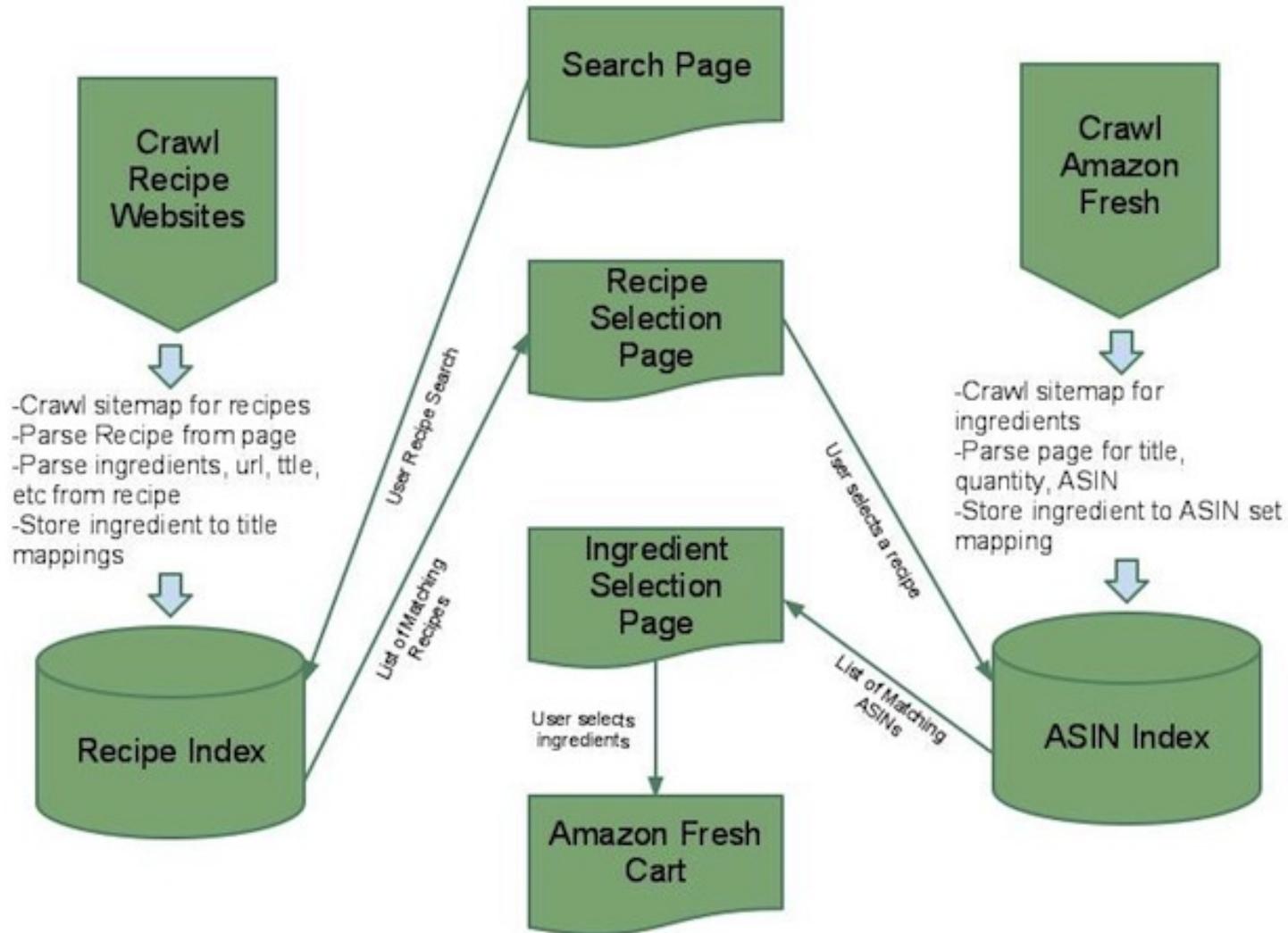
CSE 454: Advanced Internet and Web Services
Autumn 2010

Noé Khalfa · Roy McElmurry · Josh Mottaz · Aryan Naraghi · Ryan Oman

Proposed Features

- A search engine for recipes from select recipe sites
- Ingredient recognition for each recipe
- Ingredient-matching to AmazonFresh's catalogue
- The ability to automatically build an AmazonFresh cart from a given recipe while allowing user intervention
- The ability to continue browsing more recipes or be directed to AmazonFresh's checkout page

System Overview



Proposed Tasks

- Crawl and store recipes found on select sites into a database indexed by Solr (an information-retrieval system)
- Crawl and store AmazonFresh's catalogue into a Solr index
- Extract ingredients from the recipes
- Build a search interface and connect it to Solr
- Provide a method for the user to choose from a selection of product hits for every ingredient in a given recipe

Surprises and Realities

- Recipes sites did not store their recipes in a standard format
 - We ended up only parsing through a Wikia dump of about 53,000 recipes and were only able to pull out about 8,800 "clean" recipes
- AmazonFresh does not have a public API and furthermore they use RefIDs (similar to a nonce) on every session
 - We couldn't use AmazonFresh without embedding their site into ours
- AmazonFresh carries inedible items!
 - Needed to semi-manually remove categories of items
- Heritrix has poor documentation when it comes to learning how to crawl and process crawled data

Demo

ProjectNomNom website interface showing search results for "salad". The page includes a search bar with "salad" entered, a "Find Recipes!" button, and a list of search results. The first result is "Salad Seasoning" with a "Get Ingredients" link. The second result is "Corn Salad" with a "Get Ingredients" link. Below the "Corn Salad" result, there is a link to "See more details at: http://recipes.wikia.com/wiki/CORN_SALAD" and a list of ingredients: 2 cans white or yellow corn (or 1 of each), 1/2 spanish onion, 1 red pepper, 1/4 pound snow peas, 1/4 cup oil, 1 1/2 - 2 tablespoon lemon juice, salt, and pepper. A "Directions" section is also visible, starting with "Drain corn and pour into a bowl. Mix with Onion, red pepper, and snow Peas. Combine oil, lemon juice and salt with the corn mixture. Pour over corn mixture." The page is framed by a browser window with the URL "projectnomnom.com".

AmazonFresh website interface showing a product page for "Chicken Breast With Drumette, Boneless, Fresh, 6 oz each (2 pc) 12 oz total". The page features the AmazonFresh logo, navigation links (home, shop, past purchases, my list, lists, bestsellers, deals, blog, automatic delivery), and a search bar. The product is displayed with an image of two chicken breasts, a star rating of 4.5 (12 reviews), and a price of \$5.69. There is an "Add to Cart" button. Below the product, there is a note: "(Please sign in to begin shopping)" and a promotional message: "Buy with Automatic Delivery, SAVE 15%". The page also includes a "Cart" section with a "Sign in" button and a "Where We Deliver" section listing various zip codes. The page is framed by a browser window with the URL "projectnomnom.com".

ProjectNomNom

What We Learned

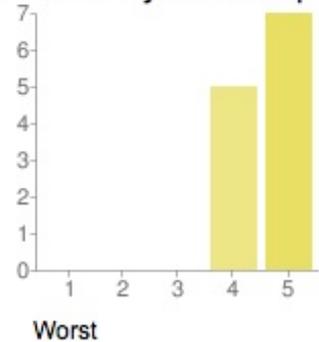
- The MVC framework methodology (Ruby on Rails)
- Solr for allowing us to quickly search our recipes database and for storing and searching the AmazonFresh data
- Git for version control
- Heritrix for crawling AmazonFresh
- Elastic Cloud Computing on Amazon Web Services for hosting our project and running our AmazonFresh crawl
- Google Docs for creating our evaluation form and this presentation :)

Self Evaluation

Recipe Search Term	Relevant Search Result Ranking	Ingredient Extraction Errors	Ingredient Matching Errors
Spaghetti	2	1	3
Meatloaf	1	0	3
Mashed Potatoes	1	0	1
Hummus	1	0	2
Sourdough	Not Found	N/A	N/A
Lemon Drop	1	0	1
Borscht	2	0	7
Turdunken	Not Found	N/A	N/A
Tabouli	Not Found	N/A	N/A

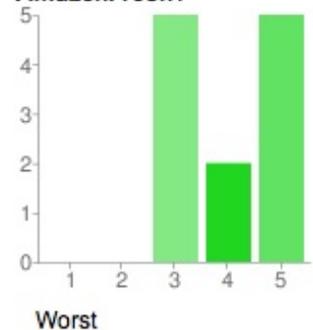
Peer Evaluation

How would you rate the quality of the recipe search results?



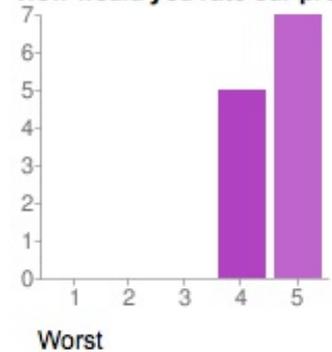
1 - Worst	0	0%
2	0	0%
3	0	0%
4	5	38%
5 - Best	7	54%

How would you rate the quality of the ingredients from the recipes as they were matched to AmazonFresh?



1 - Worst	0	0%
2	0	0%
3	5	38%
4	2	15%
5 - Best	5	38%

How would you rate our project overall?



1 - Worst	0	0%
2	0	0%
3	0	0%
4	5	38%
5 - Best	7	54%

Division of Labor

- Roy
 - Recipe parsing/data cleaning
 - Ingredient conflict page UI
- Noé
 - UI design
 - Searching infrastructure
- Ryan
 - Ruby on Rails infrastructure
 - Server maintenance
- Aryan
 - AmazonFresh data processing and indexing
 - Search auto-suggest backend
- Josh
 - AmazonFresh crawling

Questions?



(P.S.: Lunchtime is almost here!)

ProjectNomNom