MICHIGAN STATE UNIVERSITY

Beta Presentation

SIFT: *Seller-Forums Information Filtering Tool*

The Capstone Experience

Team Amazon

Maxime Goovaerts
Carl Johnson
Luke Pritchett
Benjamin Taylor
Johnny Zheng

Department of Computer Science and Engineering
Michigan State University

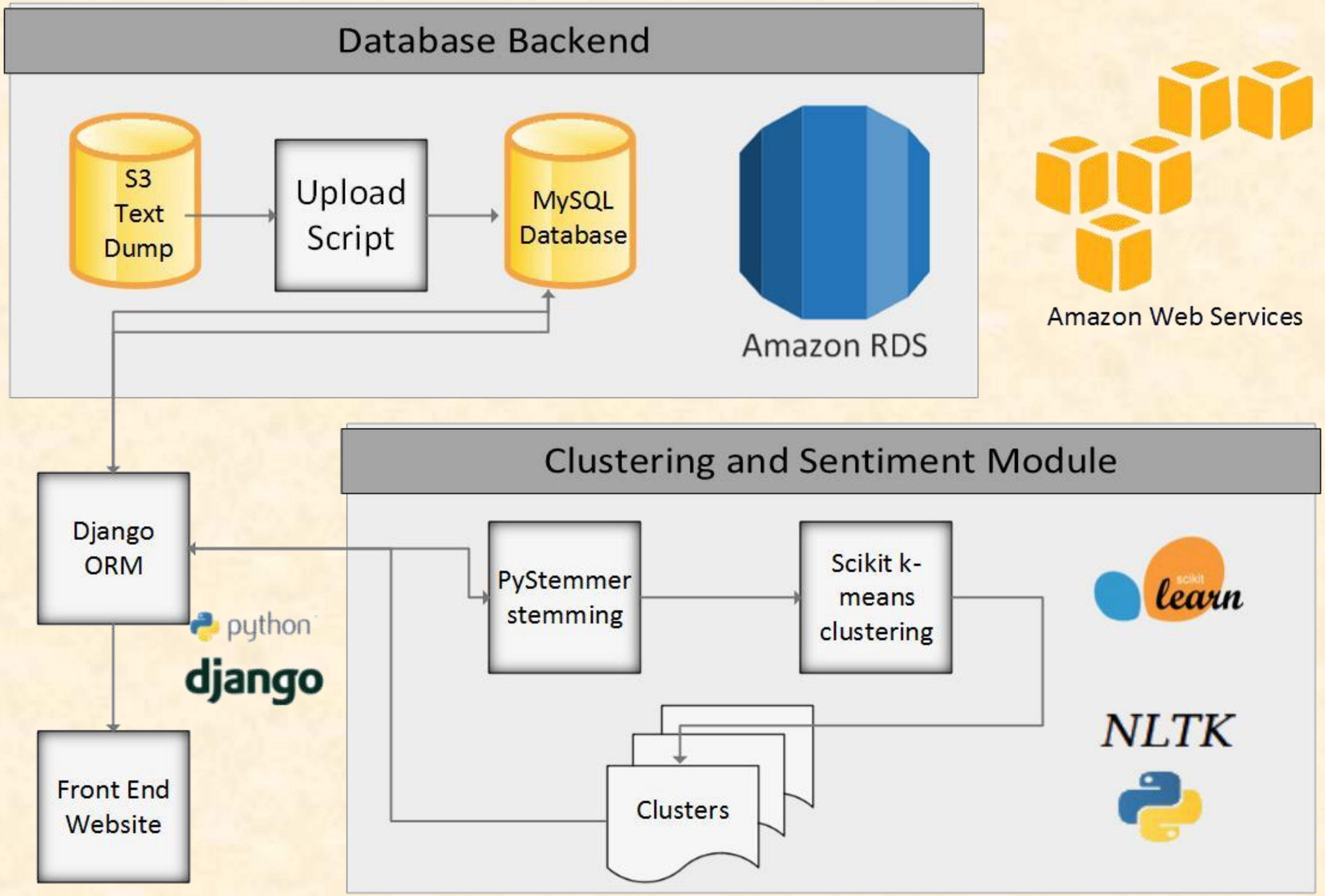
Spring 2015

From Students…
…to Professionals

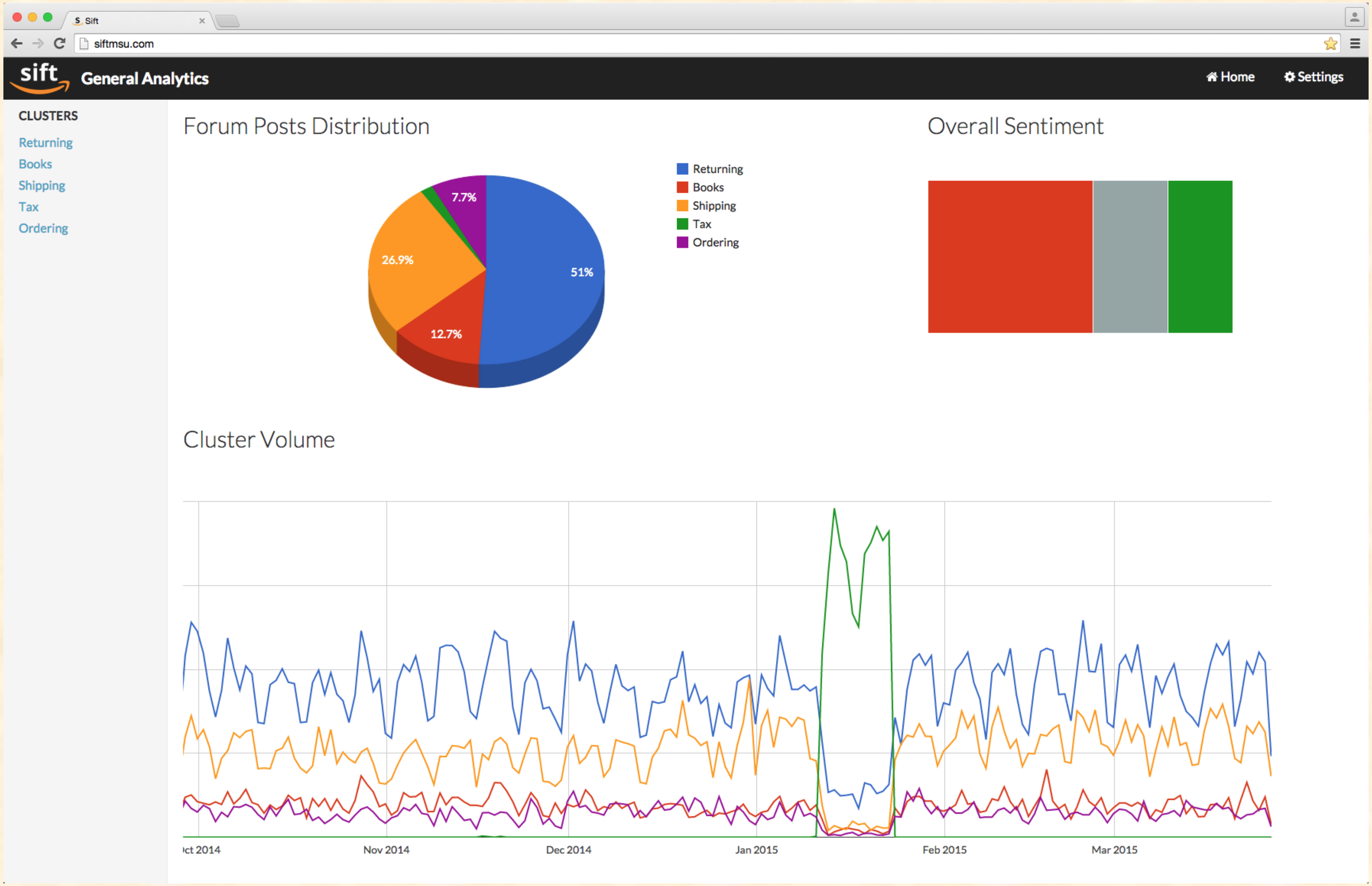
Project Overview

- Amazon is the largest internet-based retailer
- Unlock the value of 3rd party Seller Forums
- Data Organization and Analysis
  - Clustering
  - Sentiment
- Dashboard
  - Graphs and tables
  - Notifications

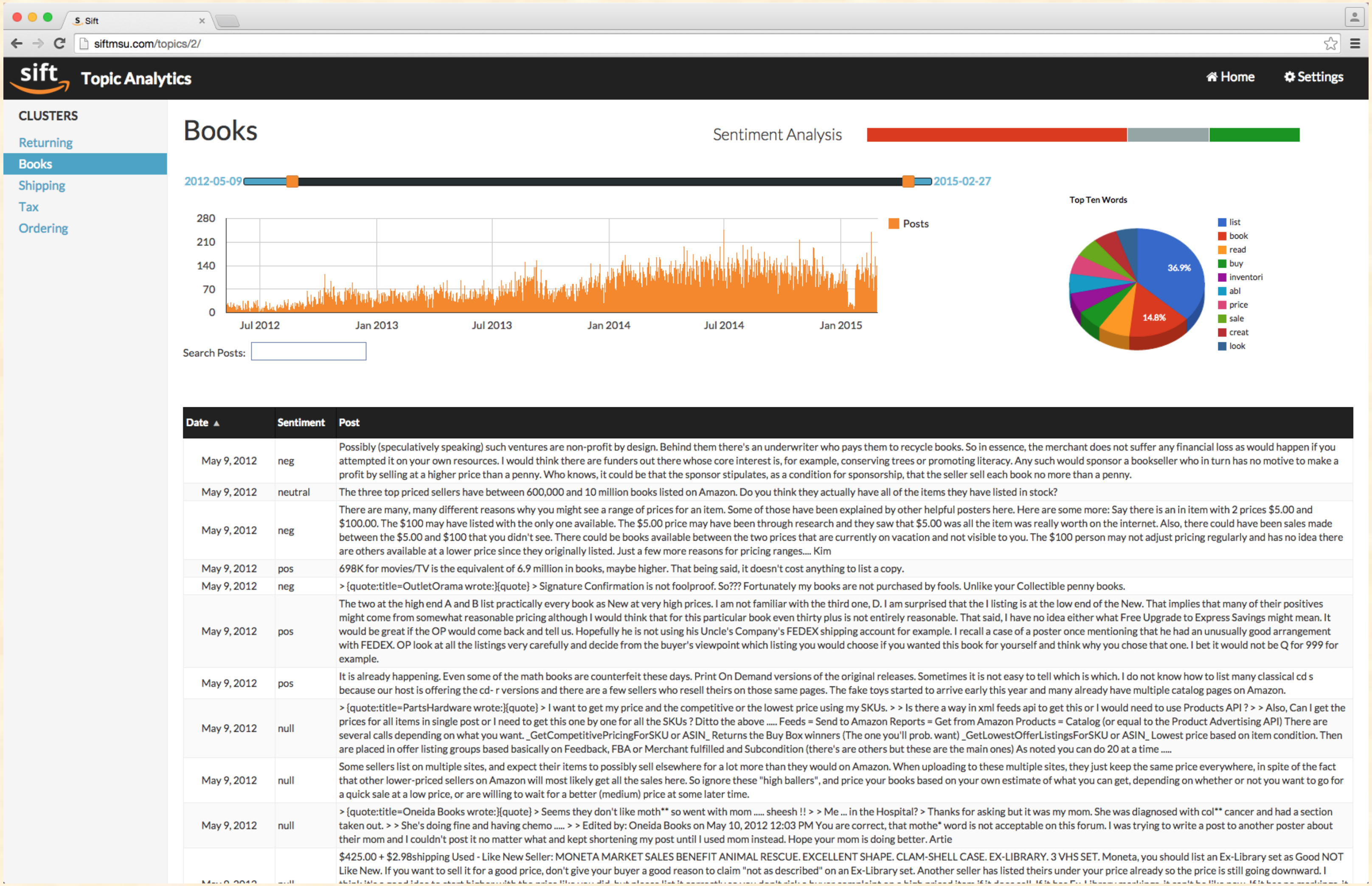
System Architecture



General Analytics

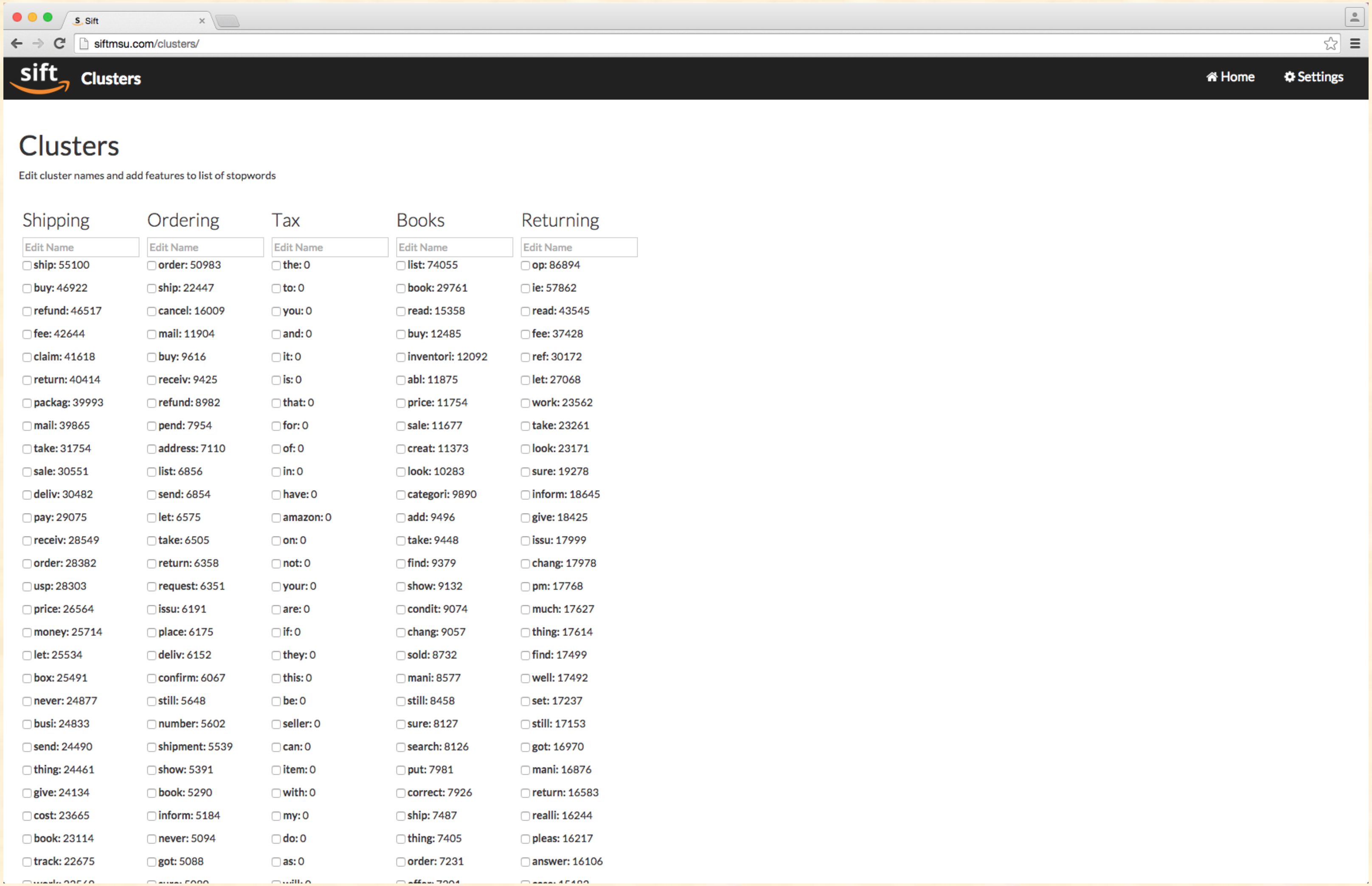


Cluster Details



Date ▲	Sentiment	Post
May 9, 2012	neg	Possibly (speculatively speaking) such ventures are non-profit by design. Behind them there's an underwriter who pays them to recycle books. So in essence, the merchant does not suffer any financial loss as would happen if you attempted it on your own resources. I would think there are funders out there whose core interest is, for example, conserving trees or promoting literacy. Any such would sponsor a bookseller who in turn has no motive to make a profit by selling at a higher price than a penny. Who knows, it could be that the sponsor stipulates, as a condition for sponsorship, that the seller sell each book no more than a penny.
May 9, 2012	neutral	The three top priced sellers have between 600,000 and 10 million books listed on Amazon. Do you think they actually have all of the items they have listed in stock?
May 9, 2012	neg	There are many, many different reasons why you might see a range of prices for an item. Some of those have been explained by other helpful posters here. Here are some more: Say there is an in item with 2 prices $5.00 and $100.00. The $100 may have listed with the only one available. The $5.00 price may have been through research and they saw that $5.00 was all the item was really worth on the internet. Also, there could have been sales made between the $5.00 and $100 that you didn't see. There could be books available between the two prices that are currently on vacation and not visible to you. The $100 person may not adjust pricing regularly and has no idea there are others available at a lower price since they originally listed. Just a few more reasons for pricing ranges.... Kim
May 9, 2012	pos	698K for movies/TV is the equivalent of 6.9 million in books, maybe higher. That being said, it doesn't cost anything to list a copy.
May 9, 2012	neg	> {quote:title=OutletOrama wrote:}{quote} > Signature Confirmation is not foolproof. So??? Fortunately my books are not purchased by fools. Unlike your Collectible penny books.
May 9, 2012	pos	The two at the high end A and B list practically every book as New at very high prices. I am not familiar with the third one, D. I am surprised that the I listing is at the low end of the New. That implies that many of their positives might come from somewhat reasonable pricing although I would think that for this particular book even thirty plus is not entirely reasonable. That said, I have no idea either what Free Upgrade to Express Savings might mean. It would be great if the OP would come back and tell us. Hopefully he is not using his Uncle's Company's FEDEX shipping account for example. I recall a case of a poster once mentioning that he had an unusually good arrangement with FEDEX. OP look at all the listings very carefully and decide from the buyer's viewpoint which listing you would choose if you wanted this book for yourself and think why you chose that one. I bet it would not be Q for 999 for example.
May 9, 2012	pos	It is already happening. Even some of the math books are counterfeit these days. Print On Demand versions of the original releases. Sometimes it is not easy to tell which is which. I do not know how to list many classical cd s because our host is offering the cd- r versions and there are a few sellers who resell theirs on those same pages. The fake toys started to arrive early this year and many already have multiple catalog pages on Amazon.
May 9, 2012	null	> {quote:title=PartsHardware wrote:}{quote} > I want to get my price and the competitive or the lowest price using my SKUs. > > Is there a way in xml feeds api to get this or I would need to use Products API ? > > Also, Can I get the prices for all items in single post or I need to get this one by one for all the SKUs ? Ditto the above ..... Feeds = Send to Amazon Reports = Get from Amazon Products = Catalog (or equal to the Product Advertising API) There are several calls depending on what you want. _GetCompetitivePricingForSKU or ASIN_ Returns the Buy Box winners (The one you'll prob. want) _GetLowestOfferListingsForSKU or ASIN_ Lowest price based on item condition. Then are placed in offer listing groups based basically on Feedback, FBA or Merchant fulfilled and Subcondition (there's are others but these are the main ones) As noted you can do 20 at a time .....
May 9, 2012	null	Some sellers list on multiple sites, and expect their items to possibly sell elsewhere for a lot more than they would on Amazon. When uploading to these multiple sites, they just keep the same price everywhere, in spite of the fact that other lower-priced sellers on Amazon will most likely get all the sales here. So ignore these "high ballers", and price your books based on your own estimate of what you can get, depending on whether or not you want to go for a quick sale at a low price, or are willing to wait for a better (medium) price at some later time.
May 9, 2012	null	> {quote:title=Oneida Books wrote:}{quote} > Seems they don't like moth** so went with mom ..... sheesh !! > > Me ... in the Hospital? > Thanks for asking but it was my mom. She was diagnosed with col** cancer and had a section taken out. > > She's doing fine and having chemo ..... > > Edited by: Oneida Books on May 10, 2012 12:03 PM You are correct, that mothe* word is not acceptable on this forum. I was trying to write a post to another poster about their mom and I couldn't post it no matter what and kept shortening my post until I used mom instead. Hope your mom is doing better. Artie
May 9, 2012	null	$425.00 + $2.98shipping Used - Like New Seller: MONETA MARKET SALES BENEFIT ANIMAL RESCUE. EXCELLENT SHAPE. CLAM-SHELL CASE. EX-LIBRARY. 3 VHS SET. Moneta, you should list an Ex-Library set as Good NOT Like New. If you want to sell it for a good price, don't give your buyer a good reason to claim "not as described" on an Ex-Library set. Another seller has listed theirs under your price already so the price is still going downward. I

Settings – *Clusters*

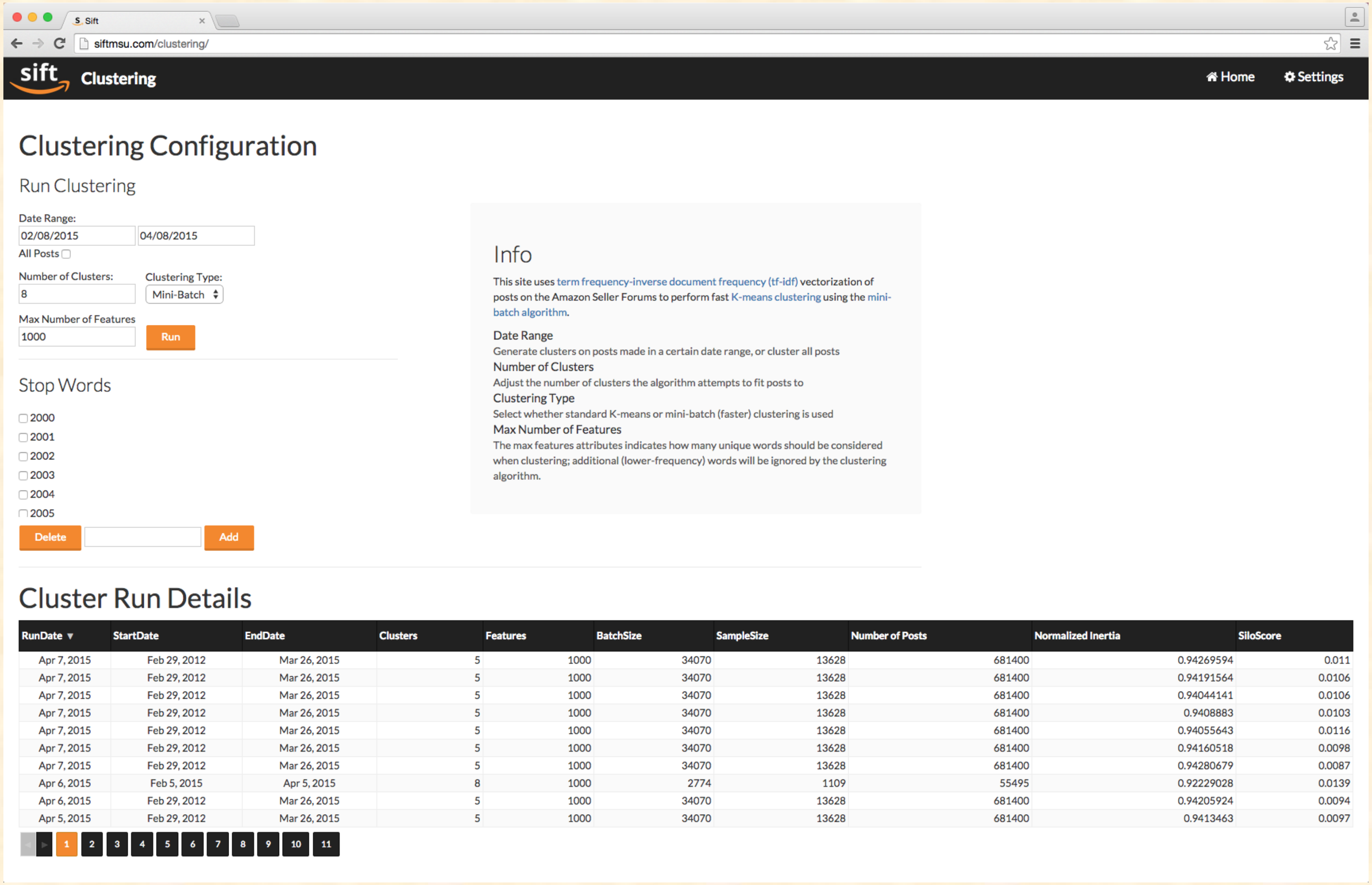


Clusters

Edit cluster names and add features to list of stopwords

Shipping	Ordering	Tax	Books	Returning
Edit Name	Edit Name	Edit Name	Edit Name	Edit Name
□ ship: 55100	□ order: 50983	□ the: 0	□ list: 74055	□ op: 86894
□ buy: 46922	□ ship: 22447	□ to: 0	□ book: 29761	□ ie: 57862
□ refund: 46517	□ cancel: 16009	□ you: 0	□ read: 15358	□ read: 43545
□ fee: 42644	□ mail: 11904	□ and: 0	□ buy: 12485	□ fee: 37428
□ claim: 41618	□ buy: 9616	□ it: 0	□ inventori: 12092	□ ref: 30172
□ return: 40414	□ receiv: 9425	□ is: 0	□ abl: 11875	□ let: 27068
□ packag: 39993	□ refund: 8982	□ that: 0	□ price: 11754	□ work: 23562
□ mail: 39865	□ pend: 7954	□ for: 0	□ sale: 11677	□ take: 23261
□ take: 31754	□ address: 7110	□ of: 0	□ creat: 11373	□ look: 23171
□ sale: 30551	□ list: 6856	□ in: 0	□ look: 10283	□ sure: 19278
□ deliv: 30482	□ send: 6854	□ have: 0	□ categori: 9890	□ inform: 18645
□ pay: 29075	□ let: 6575	□ amazon: 0	□ add: 9496	□ give: 18425
□ receiv: 28549	□ take: 6505	□ on: 0	□ take: 9448	□ issu: 17999
□ order: 28382	□ return: 6358	□ not: 0	□ find: 9379	□ chang: 17978
□ usp: 28303	□ request: 6351	□ your: 0	□ show: 9132	□ pm: 17768
□ price: 26564	□ issu: 6191	□ are: 0	□ condit: 9074	□ much: 17627
□ money: 25714	□ place: 6175	□ if: 0	□ chang: 9057	□ thing: 17614
□ let: 25534	□ deliv: 6152	□ they: 0	□ sold: 8732	□ find: 17499
□ box: 25491	□ confirm: 6067	□ this: 0	□ mani: 8577	□ well: 17492
□ never: 24877	□ still: 5648	□ be: 0	□ still: 8458	□ set: 17237
□ busi: 24833	□ number: 5602	□ seller: 0	□ sure: 8127	□ still: 17153
□ send: 24490	□ shipment: 5539	□ can: 0	□ search: 8126	□ got: 16970
□ thing: 24461	□ show: 5391	□ item: 0	□ put: 7981	□ mani: 16876
□ give: 24134	□ book: 5290	□ with: 0	□ correct: 7926	□ return: 16583
□ cost: 23665	□ inform: 5184	□ my: 0	□ ship: 7487	□ realli: 16244
□ book: 23114	□ never: 5094	□ do: 0	□ thing: 7405	□ pleas: 16217
□ track: 22675	□ got: 5088	□ as: 0	□ order: 7231	□ answer: 16106

Settings – *Cluster Configuration*



Clustering Configuration

Run Clustering

Date Range: 02/08/2015 04/08/2015

All Posts □

Number of Clusters: 8

Clustering Type: Mini-Batch

Max Number of Features: 1000

Run

Stop Words

□ 2000

□ 2001

□ 2002

□ 2003

□ 2004

□ 2005

Delete Add

Info

This site uses term frequency-inverse document frequency (tf-idf) vectorization of posts on the Amazon Seller Forums to perform fast K-means clustering using the mini-batch algorithm.

Date Range
Generate clusters on posts made in a certain date range, or cluster all posts

Number of Clusters
Adjust the number of clusters the algorithm attempts to fit posts to

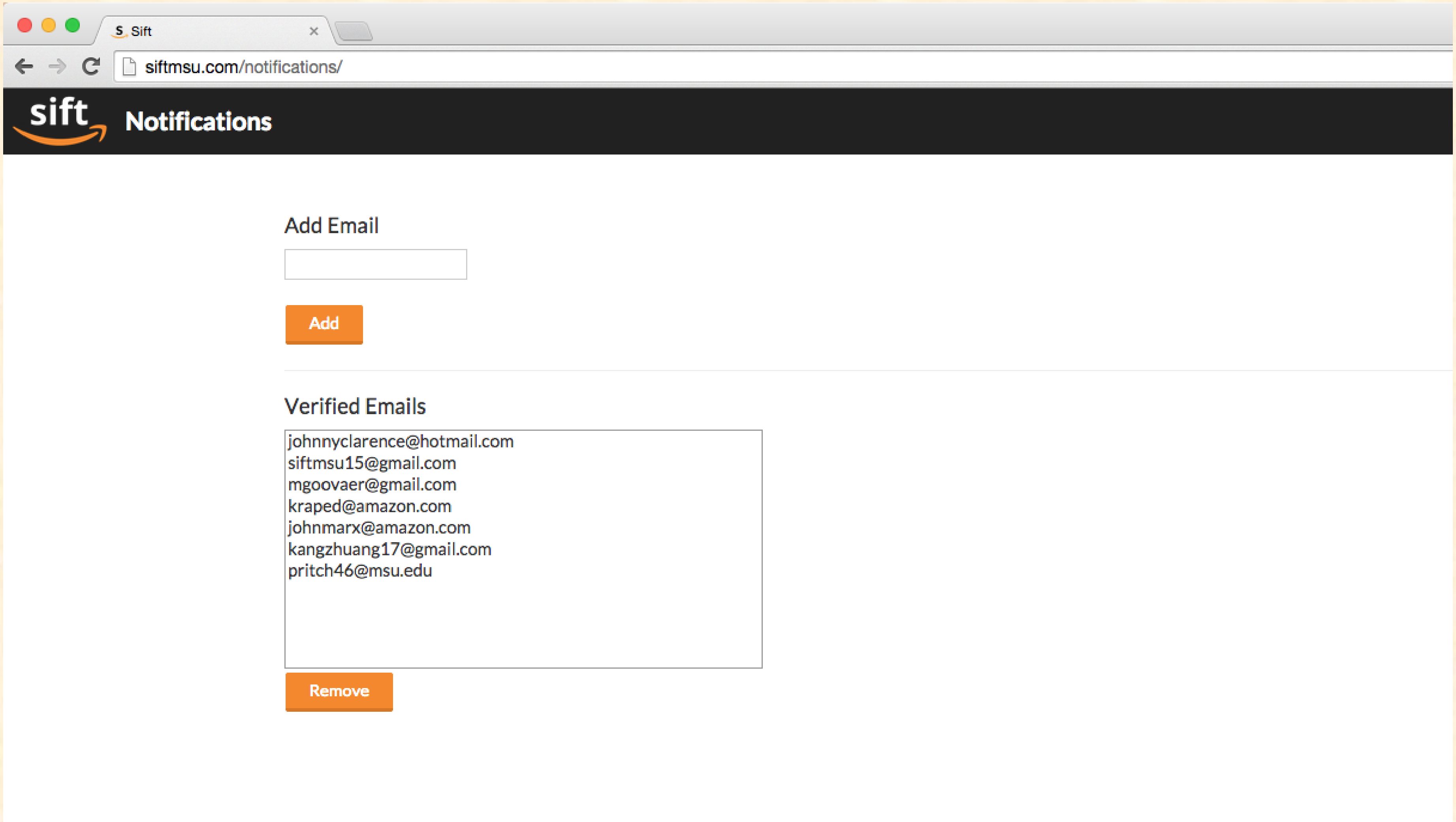
Clustering Type
Select whether standard K-means or mini-batch (faster) clustering is used

Max Number of Features
The max features attributes indicates how many unique words should be considered when clustering; additional (lower-frequency) words will be ignored by the clustering algorithm.

Cluster Run Details

RunDate ▼	StartDate	EndDate	Clusters	Features	BatchSize	SampleSize	Number of Posts	Normalized Inertia	SiloScore
Apr 7, 2015	Feb 29, 2012	Mar 26, 2015	5	1000	34070	13628	681400	0.94269594	0.011
Apr 7, 2015	Feb 29, 2012	Mar 26, 2015	5	1000	34070	13628	681400	0.94191564	0.0106
Apr 7, 2015	Feb 29, 2012	Mar 26, 2015	5	1000	34070	13628	681400	0.94044141	0.0106
Apr 7, 2015	Feb 29, 2012	Mar 26, 2015	5	1000	34070	13628	681400	0.9408883	0.0103
Apr 7, 2015	Feb 29, 2012	Mar 26, 2015	5	1000	34070	13628	681400	0.94055643	0.0116
Apr 7, 2015	Feb 29, 2012	Mar 26, 2015	5	1000	34070	13628	681400	0.94160518	0.0098
Apr 7, 2015	Feb 29, 2012	Mar 26, 2015	5	1000	34070	13628	681400	0.94280679	0.0087
Apr 6, 2015	Feb 5, 2015	Apr 5, 2015	8	1000	2774	1109	55495	0.92229028	0.0139
Apr 6, 2015	Feb 29, 2012	Mar 26, 2015	5	1000	34070	13628	681400	0.94205924	0.0094
Apr 5, 2015	Feb 29, 2012	Mar 26, 2015	5	1000	34070	13628	681400	0.9413463	0.0097

Settings – *Notifications*



Add Email

Verified Emails

johnnyclarence@hotmail.com
siftmsu15@gmail.com
mgoovaer@gmail.com
kraped@amazon.com
johnmarx@amazon.com
kangzhuang17@gmail.com
pritch46@msu.edu

What’s left to do?

- Diagnostic Clustering Progress Indicator
- Add detail to “scheduled” email
- Testing
- Code Refactoring
  - Speed up page loading by pagination
  - Make sure all functions are properly commented
- Create Project Documentation
- Project Video