MICHIGAN STATE
UNIVERSITY

Alpha Presentation

Text Classification of Seller Forums Content

The Capstone Experience

Team Amazon

Maxime Goovaerts
Carl Johnson
Luke Pritchett
Benjamin Taylor
Johnny Zheng

Department of Computer Science and Engineering
Michigan State University

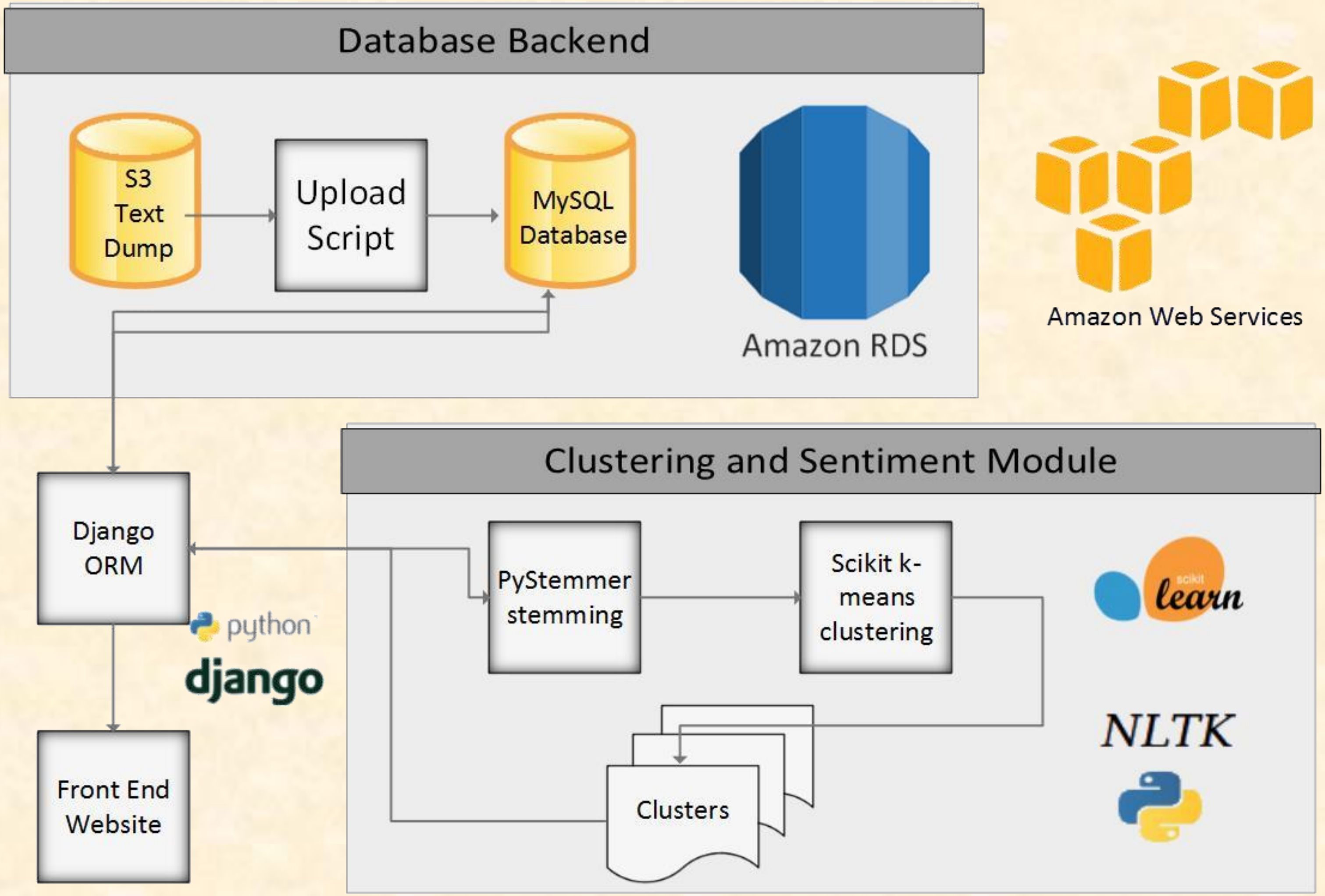
Spring 2015

From Students…
…to Professionals

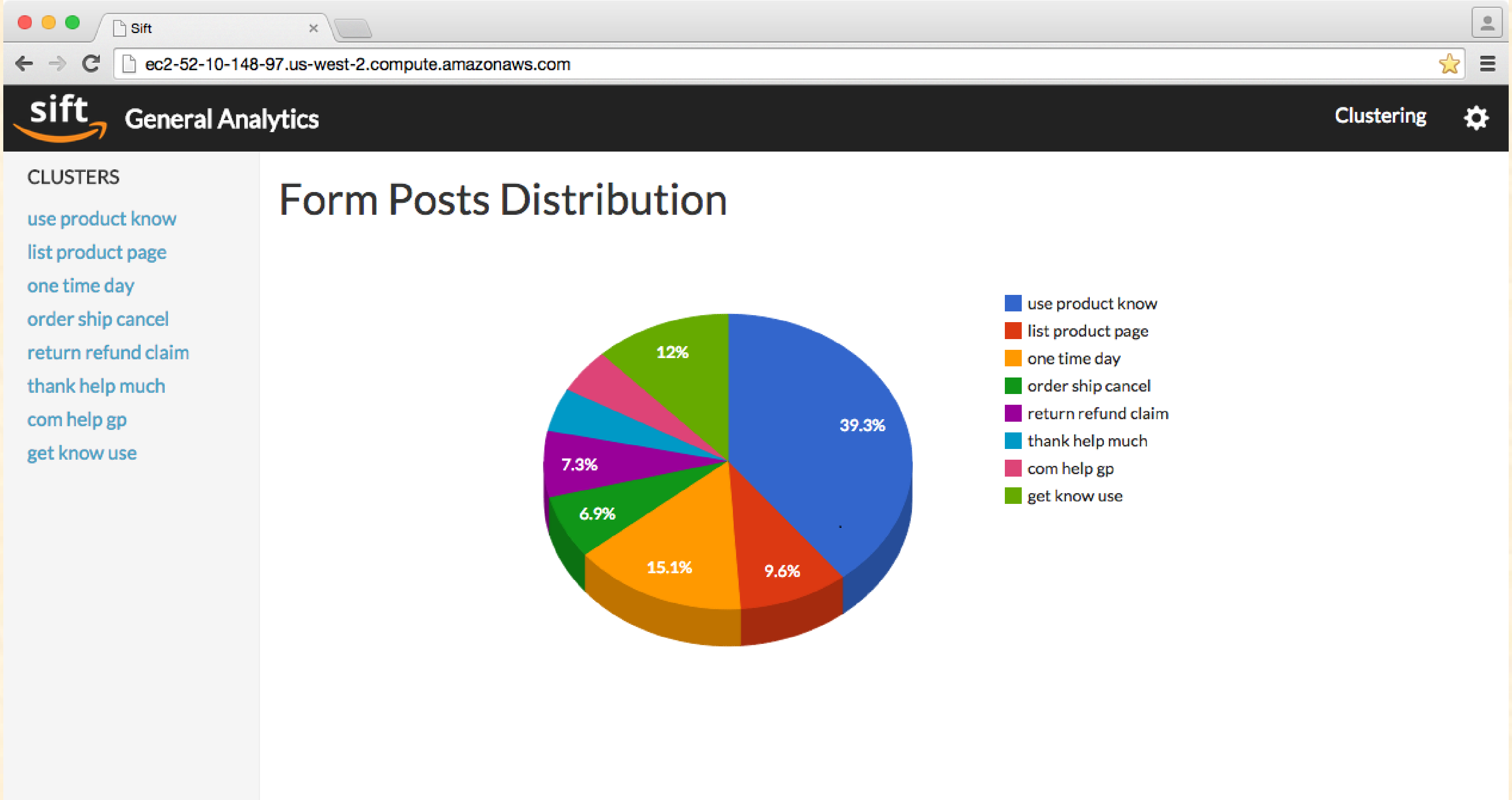
Project Overview

- 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

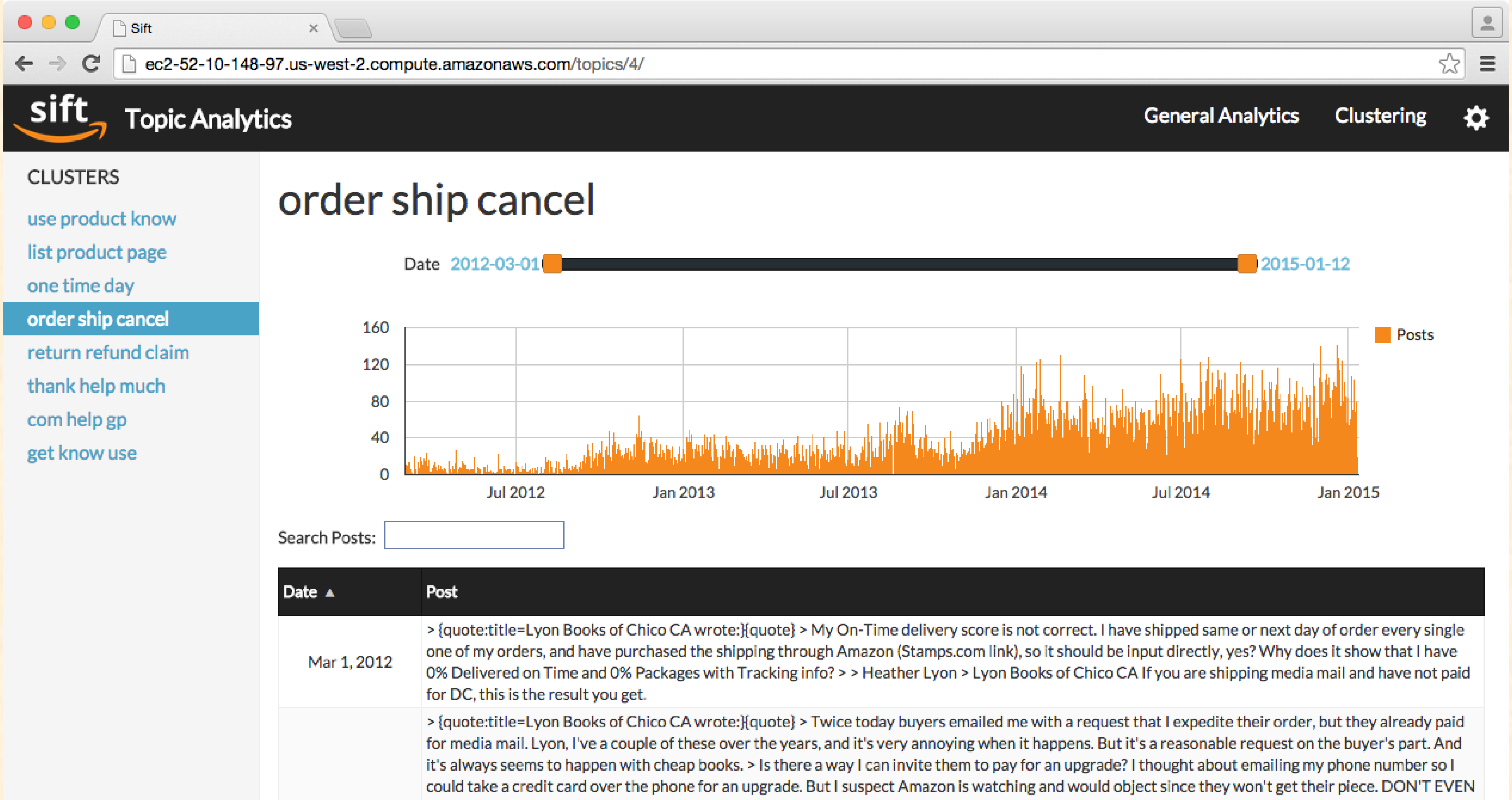
System Architecture



General Analytics



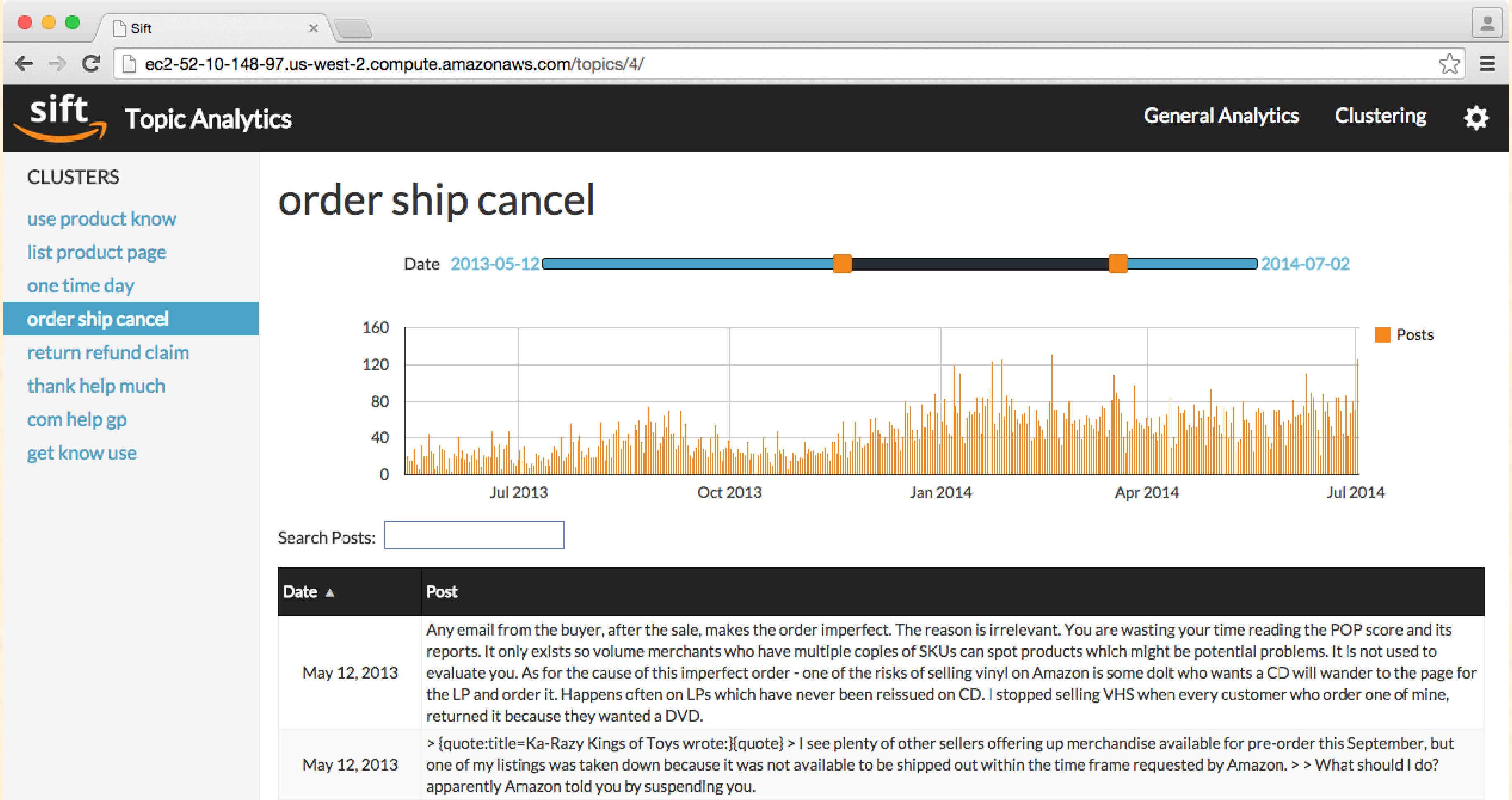
Cluster Details



order ship cancel

Date ▲	Post
Mar 1, 2012	> {quote:title=Lyon Books of Chico CA wrote:}{quote} > My On-Time delivery score is not correct. I have shipped same or next day of order every single one of my orders, and have purchased the shipping through Amazon (Stamps.com link), so it should be input directly, yes? Why does it show that I have 0% Delivered on Time and 0% Packages with Tracking info? > > Heather Lyon > Lyon Books of Chico CA If you are shipping media mail and have not paid for DC, this is the result you get.
	> {quote:title=Lyon Books of Chico CA wrote:}{quote} > Twice today buyers emailed me with a request that I expedite their order, but they already paid for media mail. Lyon, I've a couple of these over the years, and it's very annoying when it happens. But it's a reasonable request on the buyer's part. And it's always seems to happen with cheap books. > Is there a way I can invite them to pay for an upgrade? I thought about emailing my phone number so I could take a credit card over the phone for an upgrade. But I suspect Amazon is watching and would object since they won't get their piece. DON'T EVEN

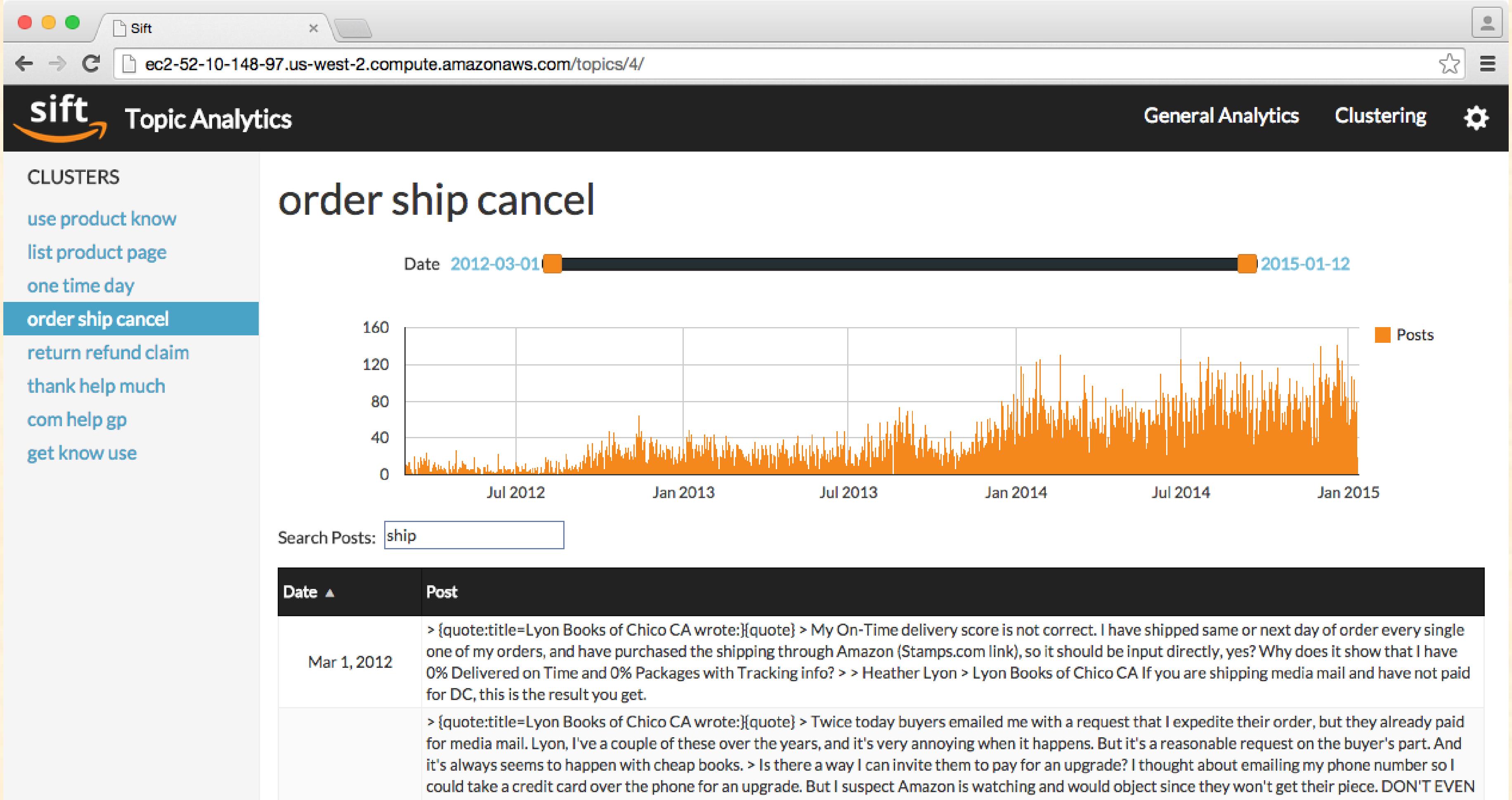
Cluster Details – *Date Selection*



order ship cancel

Date ▲	Post
May 12, 2013	Any email from the buyer, after the sale, makes the order imperfect. The reason is irrelevant. You are wasting your time reading the POP score and its reports. It only exists so volume merchants who have multiple copies of SKUs can spot products which might be potential problems. It is not used to evaluate you. As for the cause of this imperfect order - one of the risks of selling vinyl on Amazon is some dolt who wants a CD will wander to the page for the LP and order it. Happens often on LPs which have never been reissued on CD. I stopped selling VHS when every customer who order one of mine, returned it because they wanted a DVD.
May 12, 2013	> {quote:title=Ka-Razy Kings of Toys wrote:}{quote} > I see plenty of other sellers offering up merchandise available for pre-order this September, but one of my listings was taken down because it was not available to be shipped out within the time frame requested by Amazon. > > What should I do? apparently Amazon told you by suspending you.

Cluster Details – *Searching*



CLUSTERS

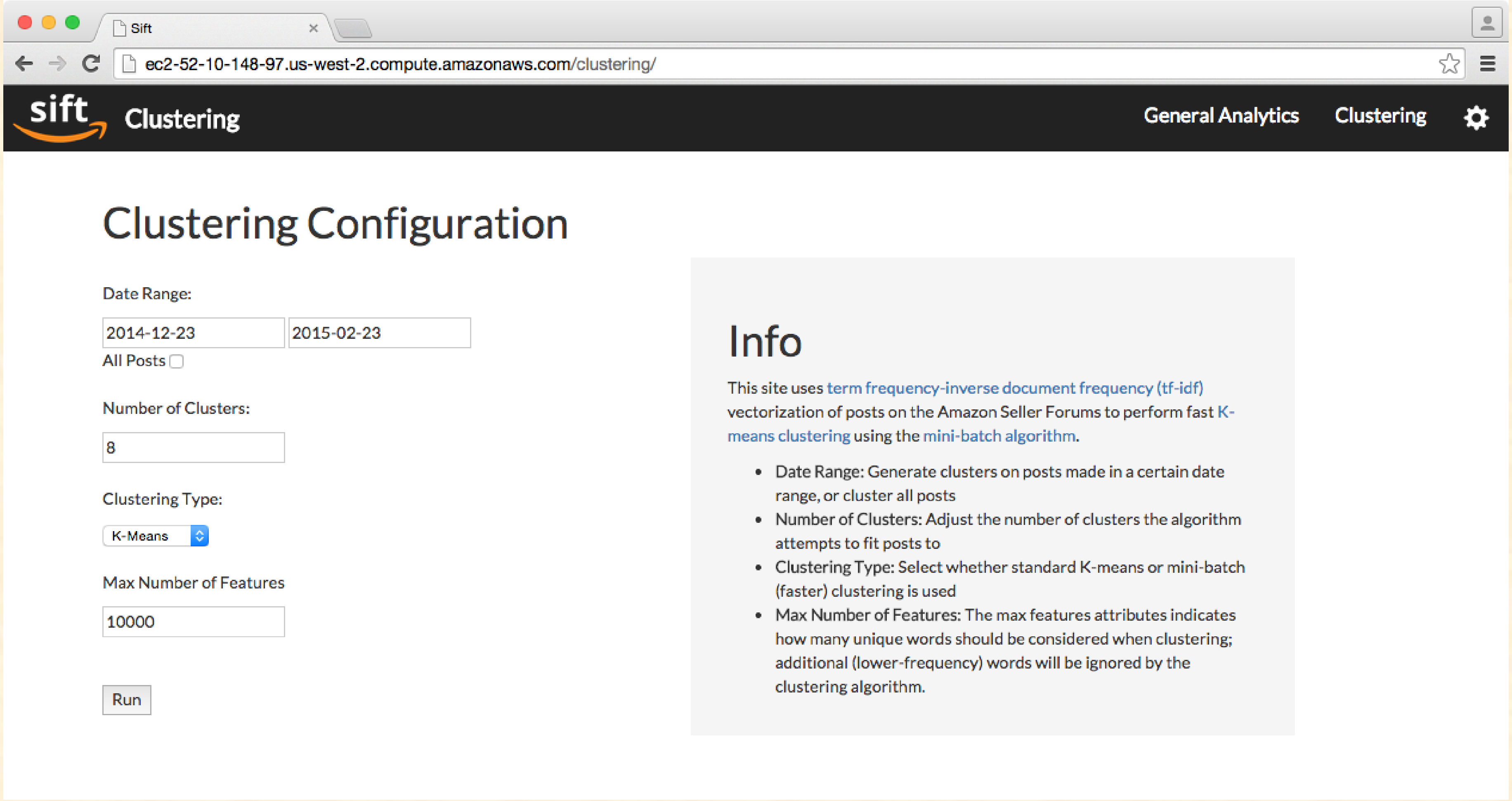
- use product know
- list product page
- one time day
- order ship cancel
- return refund claim
- thank help much
- com help gp
- get know use

order ship cancel

Search Posts: ship

Date ▲	Post
Mar 1, 2012	> {quote:title=Lyon Books of Chico CA wrote:}{quote} > My On-Time delivery score is not correct. I have shipped same or next day of order every single one of my orders, and have purchased the shipping through Amazon (Stamps.com link), so it should be input directly, yes? Why does it show that I have 0% Delivered on Time and 0% Packages with Tracking info? > > Heather Lyon > Lyon Books of Chico CA If you are shipping media mail and have not paid for DC, this is the result you get.
	> {quote:title=Lyon Books of Chico CA wrote:}{quote} > Twice today buyers emailed me with a request that I expedite their order, but they already paid for media mail. Lyon, I've a couple of these over the years, and it's very annoying when it happens. But it's a reasonable request on the buyer's part. And it's always seems to happen with cheap books. > Is there a way I can invite them to pay for an upgrade? I thought about emailing my phone number so I could take a credit card over the phone for an upgrade. But I suspect Amazon is watching and would object since they won't get their piece. DON'T EVEN

Clustering Configuration



Sift
ec2-52-10-148-97.us-west-2.compute.amazonaws.com/clustering/
sift
Clustering
General Analytics
Clustering

Clustering Configuration

Date Range:

2014-12-23 2015-02-23

All Posts □

Number of Clusters:

8

Clustering Type:

K-Means

Max Number of Features

10000

Run

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.

What’s left to do?

- Optimize Clustering
- Implement Dynamic Loading
- Generate Sci-Kit Cluster Charts
- Track Trending Clusters
- Sentiment Analysis
- Notifications
- Develop more Visualization and Analysis Tools