

Case Study

Text Mining on Web Chat Rooms

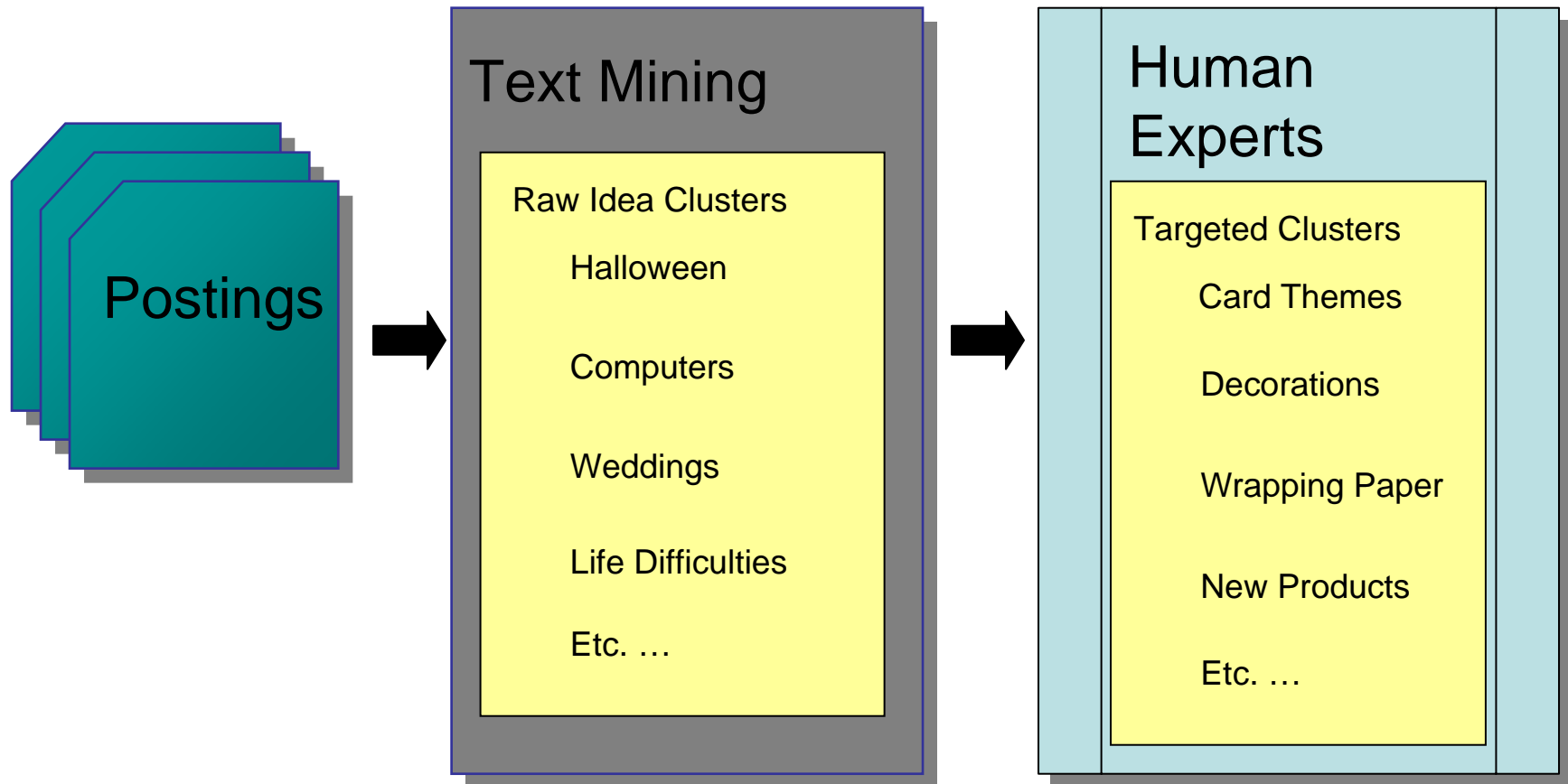
Greeting Cards Industry

Presentation Outline

- Introduction to Text Mining
- Contribution Analysis
- Clustering Results
- Predictive Model & Text Mining

Text Mining Overview

Text Mining Process



Some Vocabulary

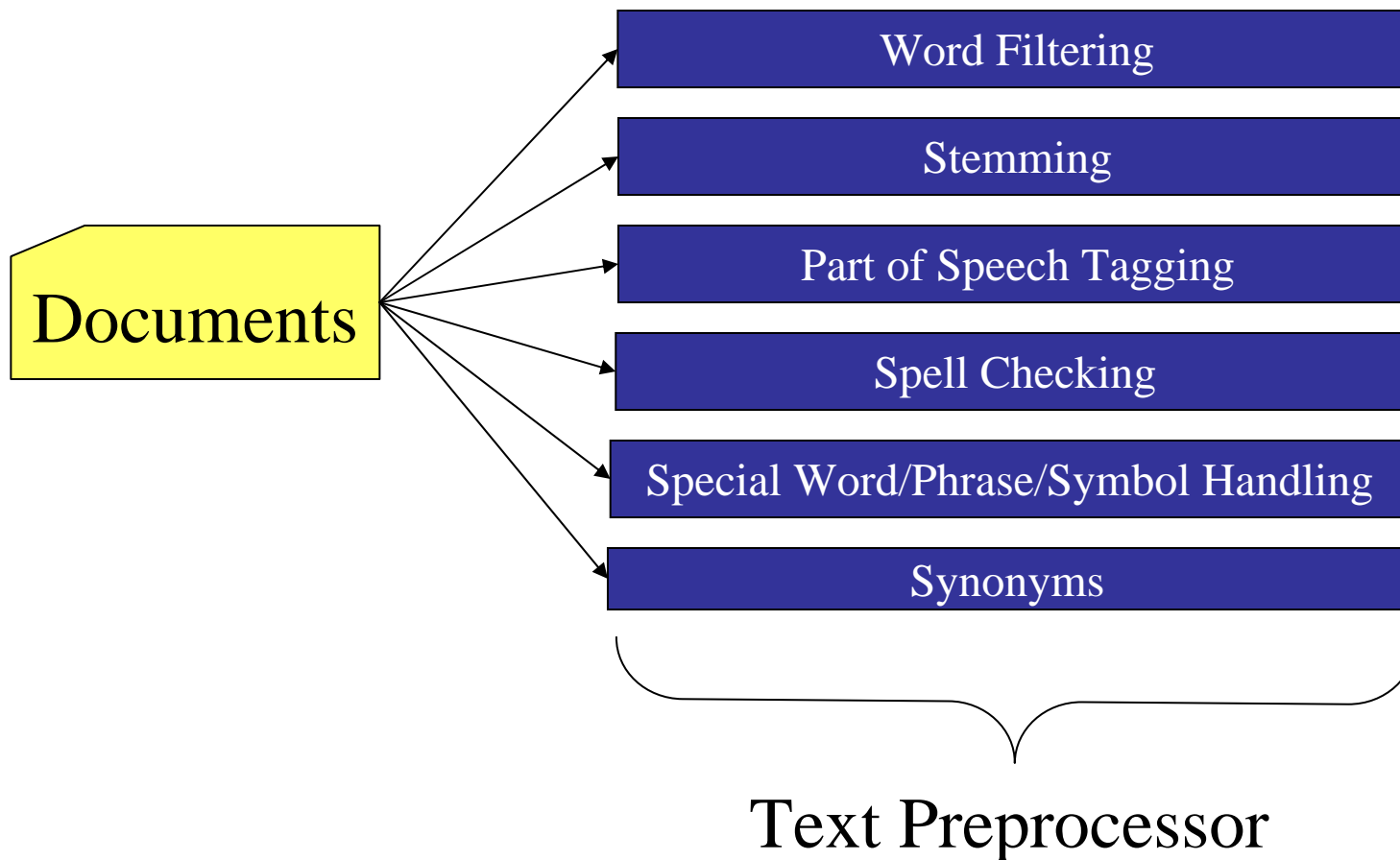
- **Text Relationship Models** - *numerical representations of text and ideas that are generated by mathematical algorithms applied to a large quantity of text data.*
- **Concept Clusters** - *clusters of words alone. The words contained in the cluster typically paint a picture of an idea or concept.*
- **Topic Clusters** - *clusters of entire contributions into groups. When analyzing the contributions together, a central topic often appears.*

Text Relationship Models

- Foundation of text mining
- Mathematical representation of words/ideas
- Allow for complex clustering of the data
- Allow for other traditional statistical techniques to be applied

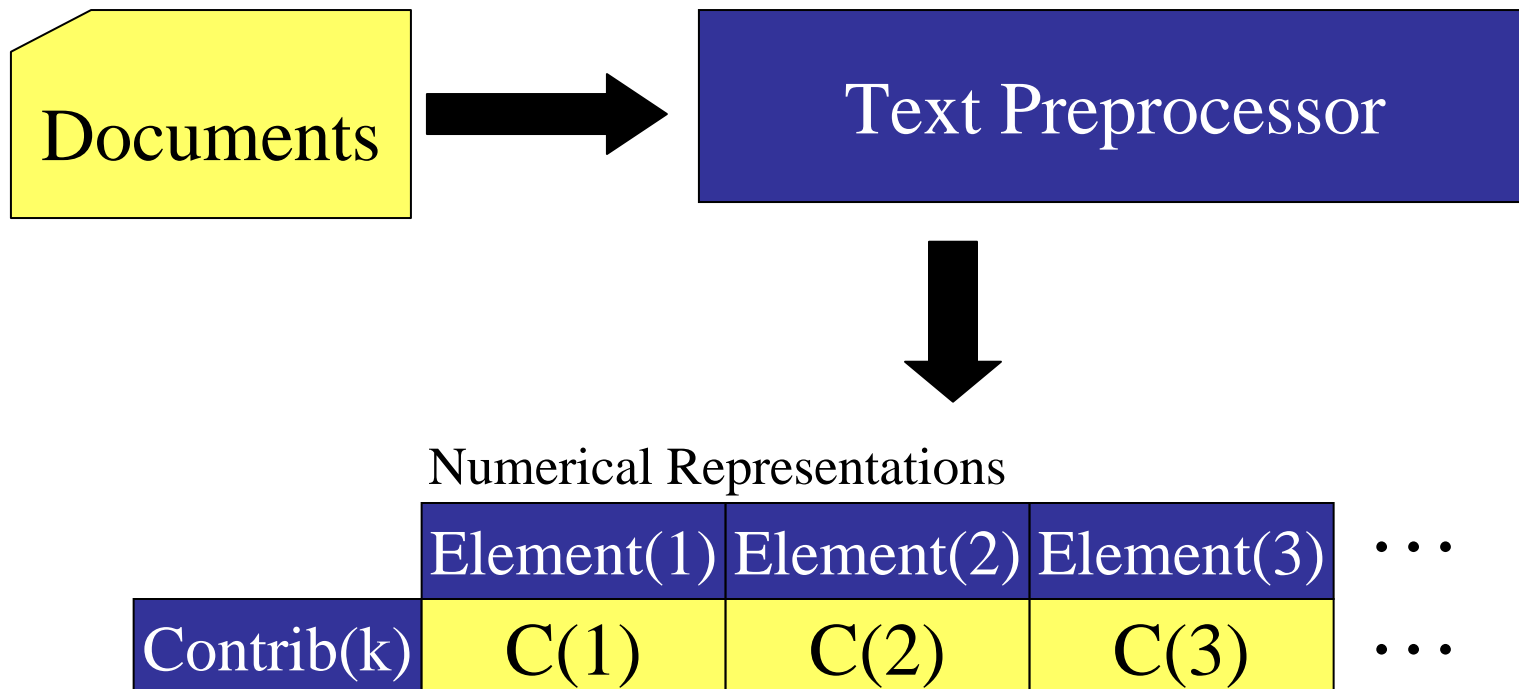
Text Mining Step-by-Step

Step 1 – Word information passed through text preprocessor



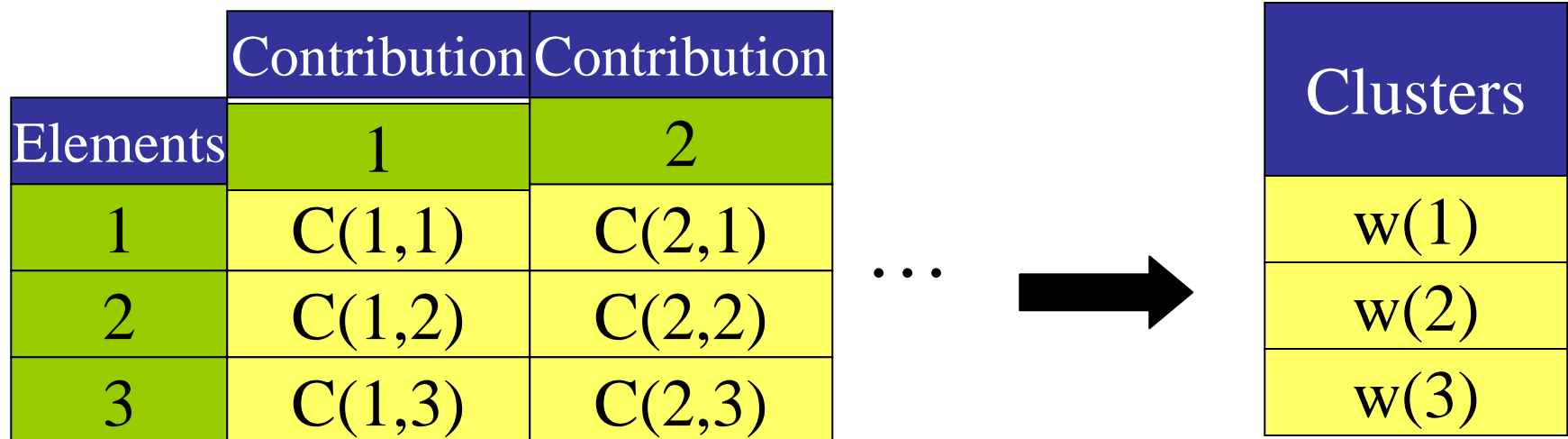
Text Mining Step-by-Step

Step 2 – Language translated into Numbers using Equity Decision Systems' text mining techniques to create text relationship models



Text Mining Step-by-Step

Step 3 - Mathematical and statistical techniques are applied to derive clusters



⋮

Clusters are derived
from the documents
and words



⋮

Generic Text Mining Software – Won't Succeed

- Text models tend to be very problem specific (NOT GENERIC!)
- Quality text model development requires many iterations and domain expertise
- Custom analysis will offer multiple ways to represent a contribution, commercial software will be limited
- Choice of metrics for determining document similarities are often too restrictive

Data Overview & Contribution

Analysis

Analysis Report Overview

- Several interesting trends identified in the postings
- Clusters
 - Identified 98 concept clusters (Sample of 29 in report)
 - Identified 45 topic clusters crossing multiple forums and communities (Sample of 17 in report)
- Predictive model overview provided

Overview of the Data

Community	Number of Participants (Excludes facilitators)	Number of Forums (Must have 1 or more non-facilitator)	Number of Contributions (Includes facilitators)	First Contribution Date (Excludes facilitators)
Parents	186	1,174	27,121	Nov. 9
Grandparents	97	502	12,649	June 7
La Comunidad de Latina	96	462	5,186	June 7
Removed from Community	36	N/A	1,634	N/A
Total	417	2,138	46,590	N/A

Overview of the Data

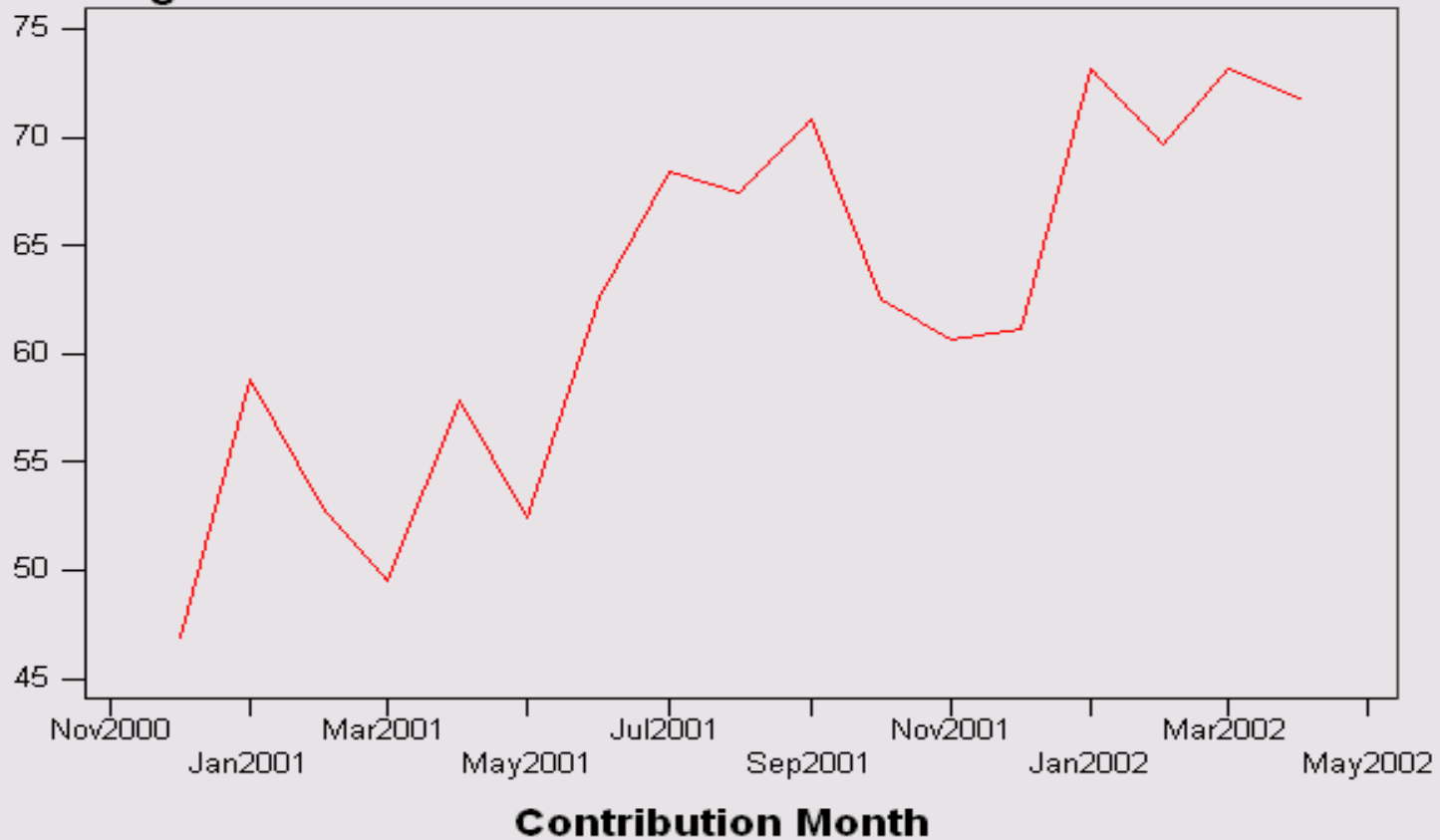
Community	Number of Participants	Average Number of Contributions per Participant per Month
Parents	186	15.3
Grandparents	97	18.0
La Comunidad de Latina	96	7.5
Overall	379	14.9

Frequently Used Proper Names

Name	Number of times Mentioned		Name	Number of times mentioned
God	1347		Walt Disney	13
Jesus	150		Bing Crosby	12
Harry Potter	141		Julia Alvarez	12
Teddy Bear	70		Sandra Cisneros	12
Martha Stewart	44		Stephen King	12
Santa Claus	42		Dr Seuss	11
Scooby Doo	34		Linda O'Dell	11
President Bush	33		Peter Pan	11
Lillian Vernon	19		Willie Nelson	11
Charlie Brown	17		Billy Joel	10
John Edwards	17		JC Penney	10
Vice President	15		Kenny Rogers	10
Jennifer Lopez	14		Mother Goose	10

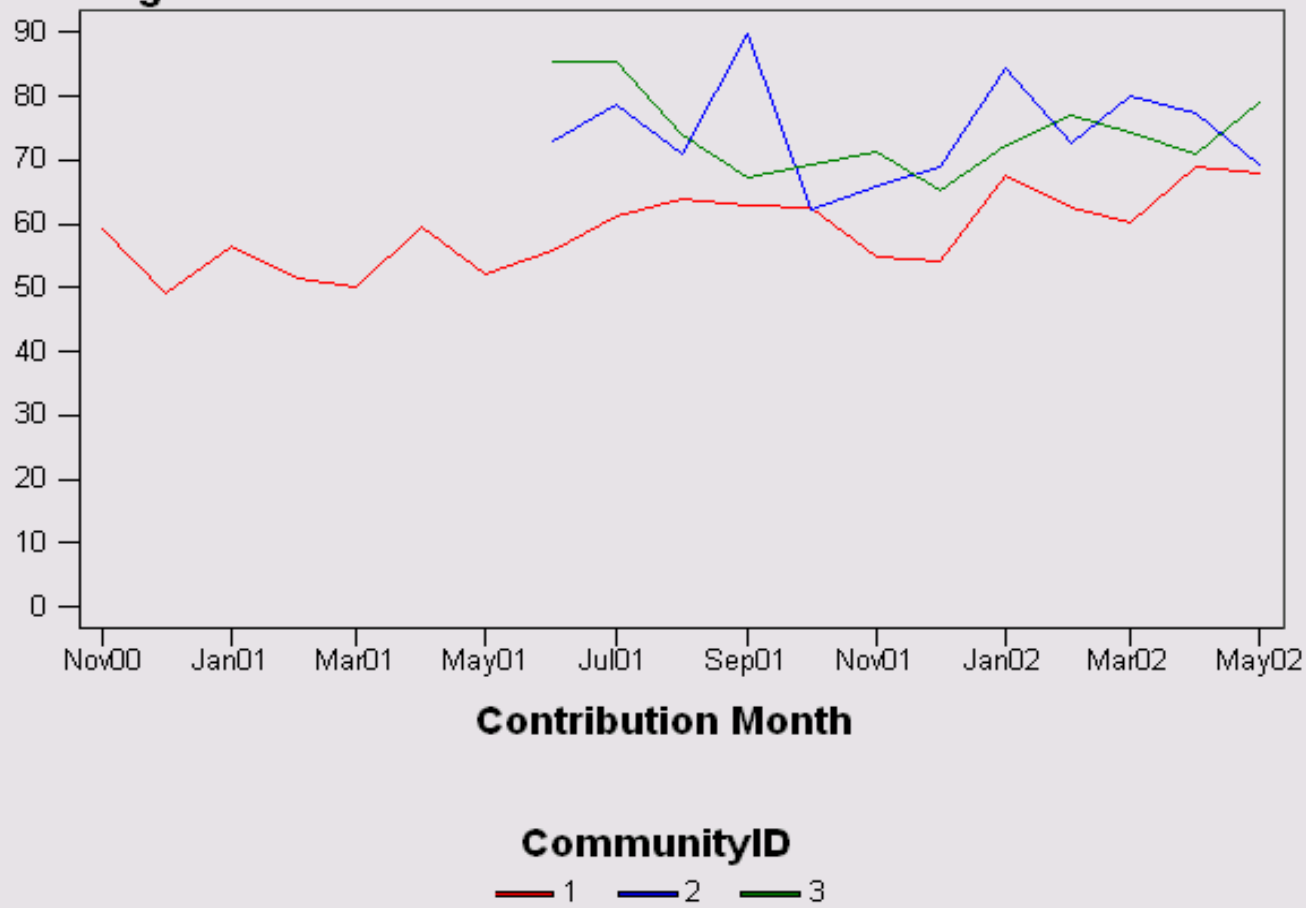
Contribution Size Trends

Average Contribution Word Count



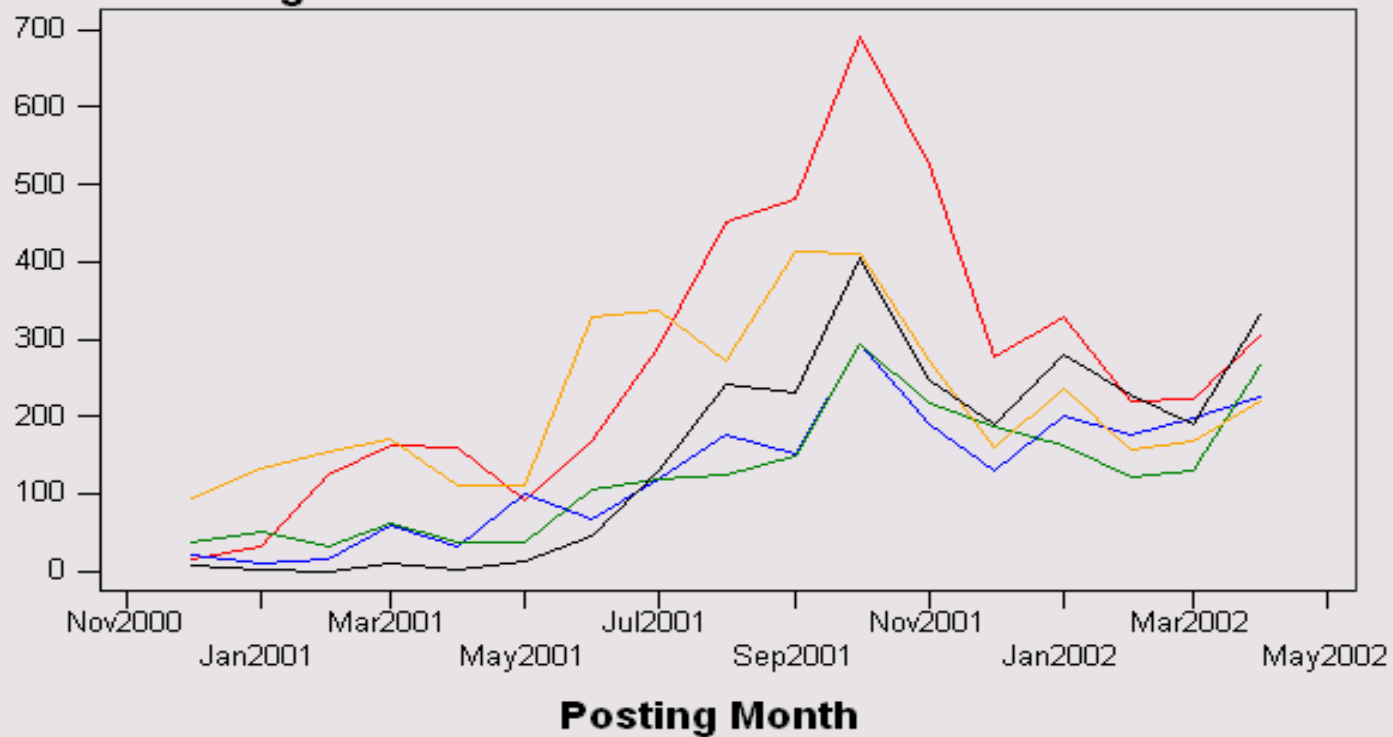
Contribution Size Trends by Community

Average Contribution Word Count



Top 5 States

Total Postings



StateCode

— CA — FL — IN — KS — VA

Clustering Results

Concept Clusters

- The word relationship models assign numerical quantities to words
- These numerical quantities are then clustered into abstract groups
- After human review, meaning is assigned to the clusters

Concept Cluster Examples

Cluster General Description	Common Word 1	Common Word 2	Common Word 3	Common Word 4	Common Word 5
Easter	Candy	Basket	Hunt	Rabbit	Fill
Halloween	Fall	Halloween	Pumpkin	Carve	Ghost
Weddings/showers	Baby	Wedding	Invite	Shower	Planner
Daily events	Dinner	Walk	Lunch	Midnight	Rest
Choice / Opinion	Select	Choose	Comment	Define	Type
Animals / Pets	Dog	Cat	Animal	Adopt	Puppy
Medical	Doctor	Test	Blood	Result	Allergy
Teaching/Education	Taught	Lesson	Instruction	Important	Genius
Weather/Seasons	Outside	Snow	Hot	Spring	Heat
Crafts	Tie	Ribbon	Attach	Glue	Bead
Idea Exchange	Response	Contribution	Reply	Click	Screen
Desserts	Cook	Cake	Chocolate	Favorite	Bake
Art/Crafts	Art	Craft	Fair	Project	Supply
Diet/exercise	Lose	Weight	Diets	Battle	Overweight
Customer service	Important	Service	Value	Trust	Employee
Collectables	Collect	Ornament	Keepsake	Baby	Handmade
Journal/Diary	Kept	Journal	Diary	Notebook	Bound
Immediate Family	Husband	Mom	Dad	Sister	Brother
Christmas	Christmas	Decorations	Tradition	Eve	Festive
Scrapbooks	Frame	Photo	Album	Scrapbook	Enlarge

Topic Cluster Examples

Topic = Arts/Crafts Around the House	Forum ID	Community
The family down the street from me has a lighthouse at the end of their walkway that is about 5' tall and lights up.	531	Parents
flowers around the pond and along the fence yes sounds great. we can each place what we like it will be pretty. Then we can put benches and tables to eat and look at all the color	2082	Grandparents
There's also a lot of birdhouse stuff. Not just for the outside--inside stuff like towels dishes and table decorations.	412	Parents
I feel the same way when I sit on the sofa in my livingroom in a quiet house and gaze at the angels on the shelf. It is so calming. When something is beautiful and special to you I guess it has that effect on you.	381	Parents
This is the new thing at the florist to buy a planter of grass to place on your coffee/end tables.	531	Parents

Topic Cluster Examples

Topic = Christmas	Forum ID	Community
<p>We are celebrating Christmas in a big way. My husband and son and I went to cut down our very own tree and we put it up right after Thanksgiving. We decorated our house together we did all our Christmas shopping together we even wrapped presents together. I would pick off the prices and fix everything in the box my husband wrapped and my son (he's 3) picked out the bows. My husband took off two weeks and we just ...</p>	1566	Latina
<p>Christmas is my favorite holiday just because of what it represents and because it is FAMILY TIME. This is really the only holiday that I 100% don't forget to send cards and love receiving them. I have an assortment of Xmas card holders and display them throughout my house.</p>	598	Latina
<p>and hugs....wow if you can find a copy I bet hubby would make a Xmas purchase for me....he knows how I loved that little book...well not little.....you are an angel!.....</p>	1108	Removed
<p>Besides my earlier comment I would like to see Hallmark bring out a "larger Easter Egg Tree" every year I buy Hallmark ornaments and go out in the woods and cut a branch of a tree down and spray it so as to make a tree for them .My collection has grown but you tree's you sell havent' . I am not the only one who dose this and my friends would appreciate it also--it's a nice pick me up after Xmas.</p>	1463	Grandparents
<p>I agree thanks for all your doing I think I let you know Will have to check did we decorate the new fence with xmas decorations? Are we going to do red white and blue anywhere?</p>	1373	Grandparents

Topic Cluster Examples

Topic = Trips/Vacations/Reunions	Forum ID	Community
I went to Disneyland and there was little bobble head mickeys in different sports gear. I bought my friend a golf one it was just too cute.	1944	Latina
I think I'd have gone crazy driving that far with kids! Our plans for the summer: one or two more camping trips and in early August our California trip (five days in San Diego four days in Huntington Beach with a jaunt to Disneyland). Other than that just relax and enjoy summer a little.	698	Parents
St. Louis is a great city to have a gathering in. Lots of historic sites to visit a great zoo a lovely riverfront baseball football good food and it's pretty centrally located. Kansas City is good too.	190	Parents
we went to alcatraz in 80 something but I would not take courtney there I think shes too young we did go to fishermans wharf and hey if beth reads this on vacation hubby id=s taking us to a rv dealer oh well lol but we want to hit the san diego zoo and the aquariam and of course Disneyland	1845	Grandparents
We always drive up to Anaheim to go to Disneyland when we're in San Diego (this year we'll do the new park too). What is Club Disney? Something we should take the time to do?	263	Parents

Topic Cluster Examples

Topic = Music	Forum ID	Community
I love the music at Christmas and all the musicians you named are wonderful. As i said in my email to you. I hope they have the CD cards for Christmas.	1940	Grandparents
I have to find those CD's too!	2094	Parents
I am with you all the way a song or music from the 40's 50's 60's or 70's is just fine with me too --thanks!	2183	Grandparents
I am covering all bases by posting favorites on both this board and your email.<p>Collections: Native American Owls<p>Music: Native American flute Enya jazz instrumentals classical<p>Books: since we will be moving into an RV eventually prefer paperbacks; topics could include Cooking for Two Inspiration Humor Mysteries.<p>Personal Passions: the colors green and purple candles fancy bookmarks fun coffee mugs.	1248	Grandparents
Mary Did You Know? by Kenny Rogers and then Grandma Got Run Over By a Reindeer. Also Go Tell It On The Mountain etc.	1525	Parents

Predictive Models & Text

Mining

Predictive Modeling in Text Data

- Singles out a particular topic of interest, like discussions about pricing
- More powerful than clustering since it is completely focused on one topic
- Determines the probability that a contribution contains a discussion of an isolated topic

Predictive Models for Contributions

1. The computer would read each contribution
2. A statistical model generates the probability that the contribution is discussing the topic of interest
3. A small portion of the contributions with the highest probabilities is returned
4. A human reviewer would sort through and select the relevant contributions

Possible Prediction Targets

- Father's Day topics
- Product pricing and affordability
- New product ideas
- Arts and crafts ideas

Example: Product Pricing

- Find contributions containing appropriate keywords: price, expensive, cheap, affordable, etc.
- Find other words commonly used along with these keywords to profile the topic
- Build models utilizing all discovered words
- Score new contributions (high score=high probability)

Other Words Associated with Pricing

cheaper	item
outrageous	catalog
ninety nine	cost
discount	afford
merchandise	ship
fundraiser	cheap
frugal	brand
range	shopper
fifty percent	profit
sale	checkout
budgets	consignment
bargain	customer
few	specialty
lower	search
convenient	retail
thrift	outlets
buyer	highest

Predictive Model Example

Top Scoring entries not containing the key words related to pricing

I love the 99 cent cards. When buying a bunch of cards the higher cost cards really add up. With the 99 cents cards you can get more for your money.

Florida does not have good garage sales for what I am looking for...mostly children's items. There are great finds for collectors at Estate Sales (which there are a lot of here...) But I never seem to do well at Garage Sales. I do find some of the items at local thrift store...

Yes I did know that some stores have a twin discount. As a matter of fact we registered at Burlington (Baby Depot) Coat Factory. They give a 20% discount on items that are the same. We got a great deal on car seats. They were on sale and we also got the 20% discount on one of them.

I usually shop for cost. First I stop at Aldi's for groceries foil baggies. Then to Wegman's for any thing Aldi's doesn't have like sale items and items that cost less like milk. Diapers I will get where ever they are on sale and I use a coupon.

Predictive Model Example

Low Scoring entries containing the key words related to pricing

I read somewhere once and I don't remember who said it or I would quote them "The best thing a man can do for his children is to love their mother." I think that is really true. If you are happy and get along well the children have a sense of security that is priceless.

I've always felt that the children never asked to be here in the first place and they shouldn't have to pay the price. And the grandparents. My thoughts are with you.

Oh my gosh! Neither has my husband. I really never thought of it but you hit it right on. I'm going to have to think of a nice way to bring that up to him just to see what answer he gives. Thanks for a wonderful laugh as well as a priceless insight. ;)

Every year my mom says that the only thing she is thankful for is her family. So if everyone comes home for Thanksgiving then she is extremely happy. So every year we are all there small price for happiness. Especially since 2 years ago we almost lost her to a heart attack! Family always be thankful for this blessing!

Conclusions

- Text Mining provides a method for analyzing unstructured text
- Text Mining allows unstructured text to be used in predictive modeling settings
- When applied to a chat board, text mining can be used to organize and categorize the entries efficiently into key topics of interest