Tweeter Analysis on University of Chicago

Presented By

Joan Zhang

THE UNIVERSITY OF CHICAGO

Outline

- Executive Summary
- Data Source and Exploration
- Methodology & Model Output
- Recommendations

Executive Summary

- Twitter is a favorite source of text data for analysis.
- The project is to get deeper insights into tweeter followers' opinions towards university of Chicago by using IBM SPSS modeler.
- From a cluster of tweets, we'll find Key differentiators between university of Chicago and other top universities nationwide such as Harvard and Massachusetts Institute of Technology (MIT).
- Get sentiment analysis on University of Chicago and find what distinguishes University of Chicago from other universities such as northwestern and University of Illinois (UIC) in the Chicago metropolitan area.
- Provides recommendation how to boost University of Chicago's name recognition
- Keywords: Tweeter, IBM SPSS modeler, university of chicago(uchicago)

THE UNIVERSITY OF CHICAGO

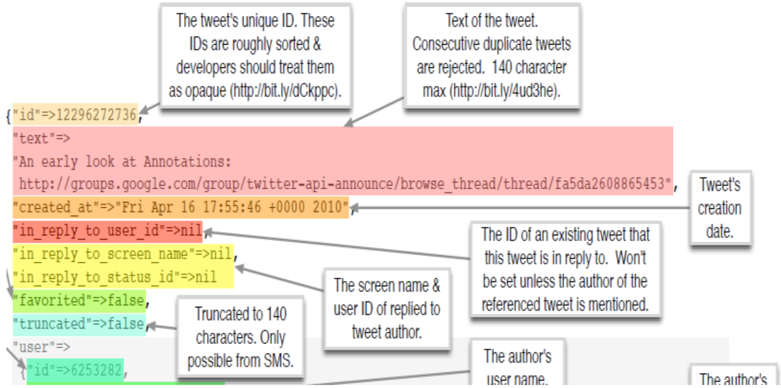
Data Source and Exploration

Collect Initial Data & Explore data

- A collection of ~7K tweets containing university of Chicago , and other universities from tweeter

- Tweeter data comes in JSON format. After combining all file, it is roughly 21 M records

- Below chart is an example of tweeter



The tweet's unique ID. These IDs are roughly sorted & developers should treat them as opaque (http://bit.ly/dCkppc).

Text of the tweet. Consecutive duplicate tweets are rejected. 140 character max (http://bit.ly/4ud3he).

```
{"id"=>12296272736,
 "text"=>
 "An early look at Annotations:
  http://groups.google.com/group/twitter-api-announce/browse_thread/thread/fa5da2608865453",
 "created_at"=>"Fri Apr 16 17:55:46 +0000 2010",
 "in_reply_to_user_id"=>nil,
 "in_reply_to_screen_name"=>nil,
 "in_reply_to_status_id"=>nil
 "favorited"=>false,
 "truncated"=>false,
 "user"=>
  {"id"=>6253282,
```

Tweet's creation date.

The ID of an existing tweet that this tweet is in reply to. Won't be set unless the author of the referenced tweet is mentioned.

The screen name & user ID of replied to tweet author.

Truncated to 140 characters. Only possible from SMS.

The author's user name.

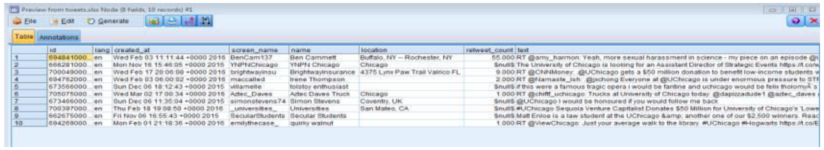
The author's

Data Preparation

- Below chart represents the variables we selected for this project

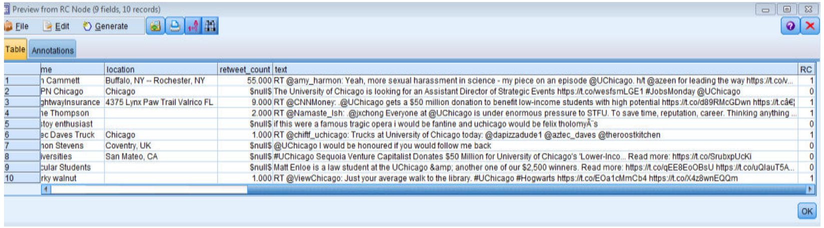
id	lang	created_at	screen_name	name	location	retweet_count	text

- Elephant bird package was used to eliminate the messages not related to the University of Chicago. The final records come to 15,886 in a text file



	id	lang	created_at	screen_name	name	location	retweet_count	text
1	694841000...	en	Wed Feb 03 11:11:44 +0000 2016	BenCam137	Ben Cammett	Buffalo, NY -- Rochester, NY	55.000	RT @amy_harmon: Yeah, more sexual harassment in science - my piece on an episode @U
2	666281000...	en	Mon Nov 16 15:46:05 +0000 2015	YNPNChicago	YNPN Chicago	Chicago	$null$	The University of Chicago is looking for an Assistant Director of Strategic Events https://t.co/w
3	700049000...	en	Wed Feb 17 20:06:08 +0000 2016	brightwayinsu	BrightwayInsurance	4375 Lynx Paw Trail Valrico FL	9.000	RT @CNNMoney: .@UChicago gets a $50 million donation to benefit low-income students w
4	694762000...	en	Wed Feb 03 06:00:02 +0000 2016	maccalled	Irene Thompson		2.000	RT @Namaste_Ish: .@jxchong Everyone at @UChicago is under enormous pressure to STF
5	673566000...	en	Sun Dec 06 18:12:43 +0000 2015	villamelle	tolstoy enthusiast		$null$	if this were a famous tragic opera i would be fantine and uchicago would be felix tholomyÃ¨s
6	705075000...	en	Wed Mar 02 17:00:34 +0000 2016	Aztec_Daves	Aztec Daves Truck	Chicago	1.000	RT @chifff_uchicago: Trucks at University of Chicago today: @dapizzadude1 @aztec_daves (
7	673466000...	en	Sun Dec 06 11:35:04 +0000 2015	simonstevens74	Simon Stevens	Coventry, UK	$null$	@UChicago I would be honoured if you would follow me back
8	700397000...	en	Thu Feb 18 19:08:50 +0000 2016	_universities_	Universities	San Mateo, CA	$null$	#UChicago Sequoia Venture Capitalist Donates $50 Million for University of Chicago's 'Lowe
9	662675000...	en	Fri Nov 06 16:55:43 +0000 2015	SecularStudents	Secular Students		$null$	Matt Enloe is a law student at the UChicago &amp; another one of our $2,500 winners. Reac
10	694269000...	en	Mon Feb 01 21:18:36 +0000 2016	emilythecase_	quirky walnut		1.000	RT @ViewChicago: Just your average walk to the library. #UChicago #Hogwarts https://t.co/E

- Retweet count (RT) is added as a new column to separate the original messages. RT is given



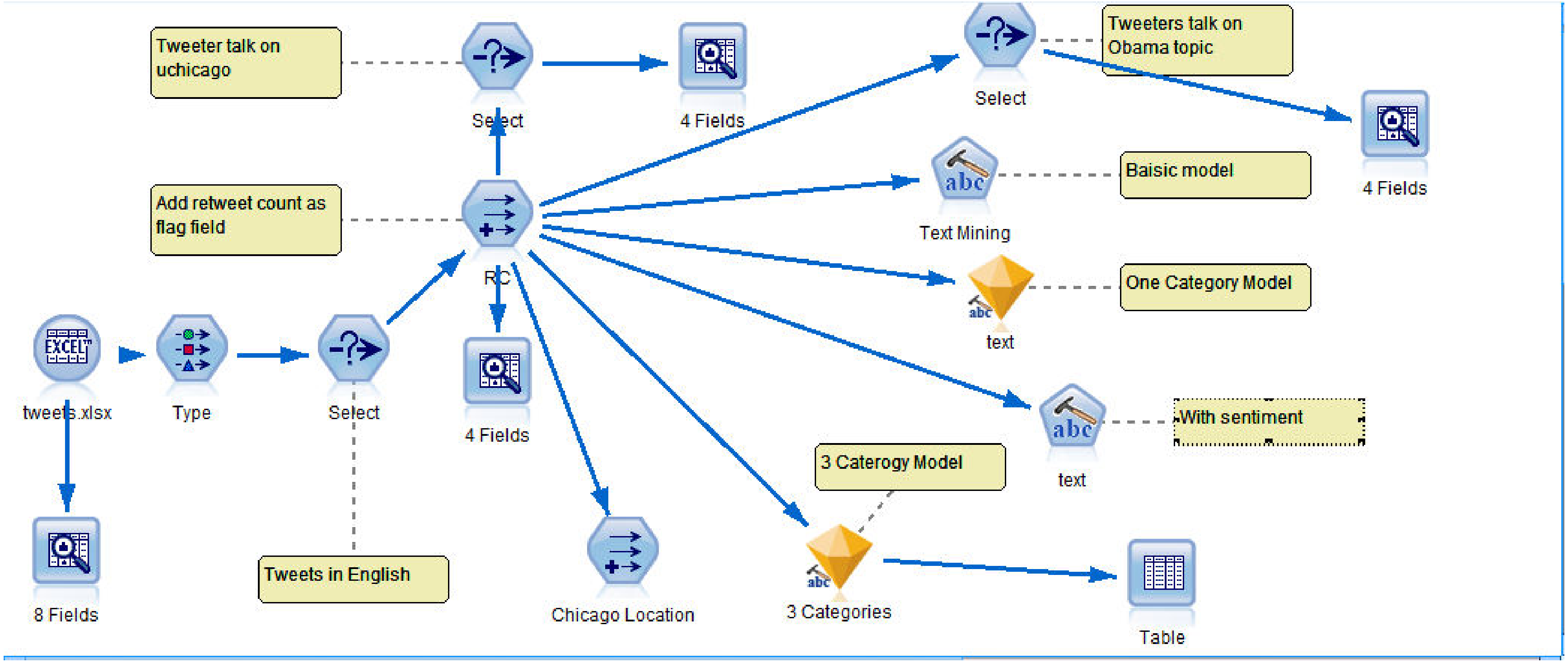
	me	location	retweet_count	text	RC
1	n Cammett	Buffalo, NY -- Rochester, NY	55.000	RT @amy_harmon: Yeah, more sexual harassment in science - my piece on an episode @UChicago. h/t @azeen for leading the way https://t.co/v...	1
2	PN Chicago	Chicago	$null$	The University of Chicago is looking for an Assistant Director of Strategic Events https://t.co/wesfsmLGE1 #JobsMonday @UChicago	0
3	ghtwayInsurance	4375 Lynx Paw Trail Valrico FL	9.000	RT @CNNMoney: .@UChicago gets a $50 million donation to benefit low-income students with high potential https://t.co/d89RMcGDwn https://t.câ€¦	1
4	ne Thompson		2.000	RT @Namaste_Ish: .@jxchong Everyone at @UChicago is under enormous pressure to STFU. To save time, reputation, career. Thinking anything ...	1
5	stoy enthusiast		$null$	if this were a famous tragic opera i would be fantine and uchicago would be felix tholomyÃ¨s	0
6	ec Daves Truck	Chicago	1.000	RT @chifff_uchicago: Trucks at University of Chicago today: @dapizzadude1 @aztec_daves @theroostkitchen	1
7	non Stevens	Coventry, UK	$null$	@UChicago I would be honoured if you would follow me back	0
8	iversities	San Mateo, CA	$null$	#UChicago Sequoia Venture Capitalist Donates $50 Million for University of Chicago's 'Lower-Inco... Read more: https://t.co/SrubxpUcKi	0
9	cular Students		$null$	Matt Enloe is a law student at the UChicago &amp; another one of our $2,500 winners. Read more: https://t.co/qEE8EoOBsU https://t.co/uQlauT5A...	0
10	rky walnut		1.000	RT @ViewChicago: Just your average walk to the library. #UChicago #Hogwarts https://t.co/EOa1cMmCb4 https://t.co/X4z8wnEQQm	1

Methodology

- IBM SPSS is used to analyze tweeter contents and create two types of models. Basic model [1] and sentimental analysis model
- Basic model with opinion loaded of resource template to compare University of Chicago with top universities nationwide and other universities in metropolitan area
- Model with sentimental analysis package loaded to analyze retweeters' sentiment towards University of Chicago
- Cluster tweets in terms of users' retweet count and users' location

[1] please see appendix on basic model chart

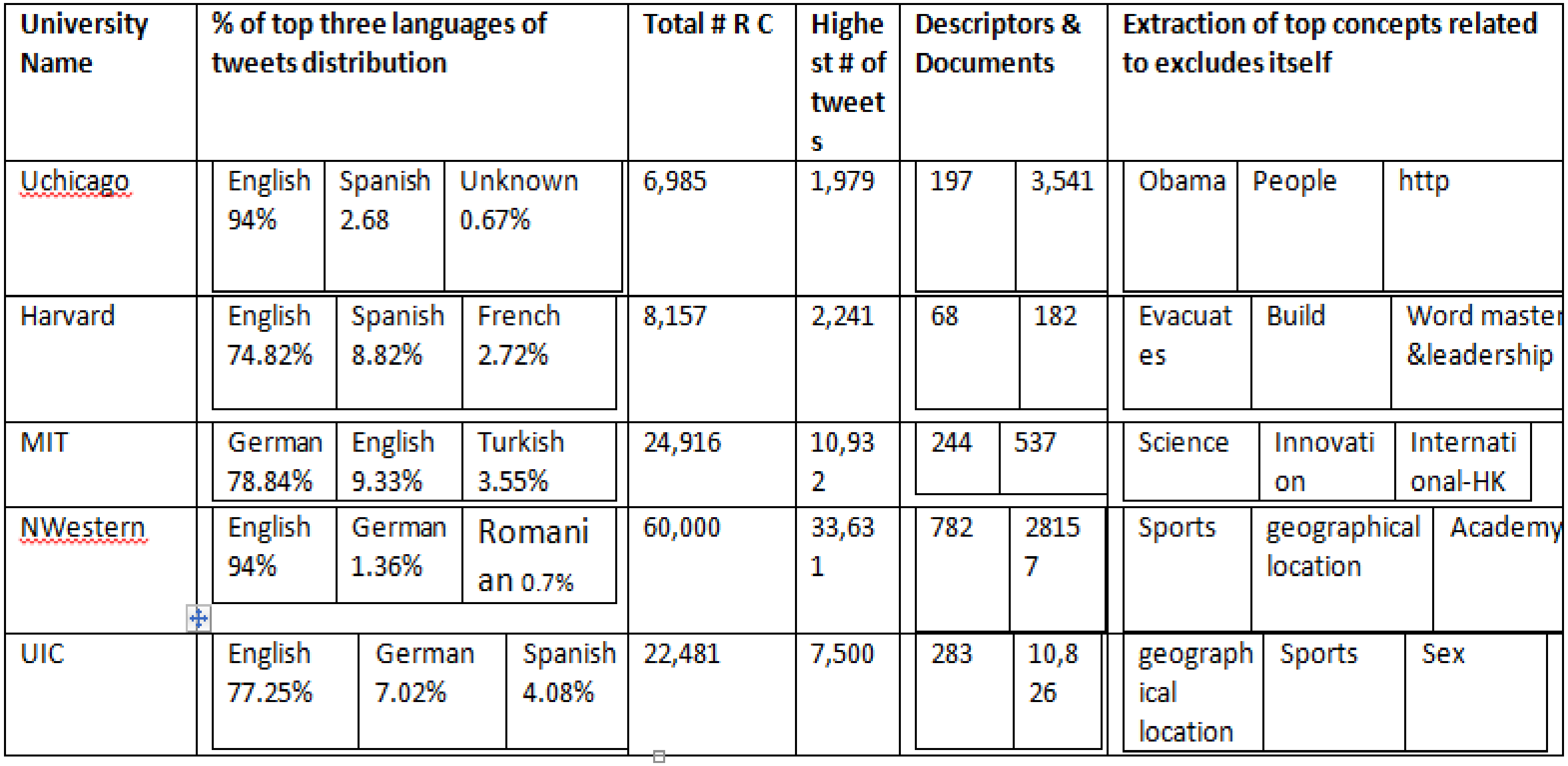
Model output on UChicago



4 Add categories as new fields

Key Difference Between University of Chicago and Other Top Universities Nationwide and Other Ones in Chicago Metropolitan area[2]

- UChicago has few tweets in foreign languages and less number of tweets compare with top university nationwide
- Uchicago more related to well-known individual Vs Northwestern and UIC



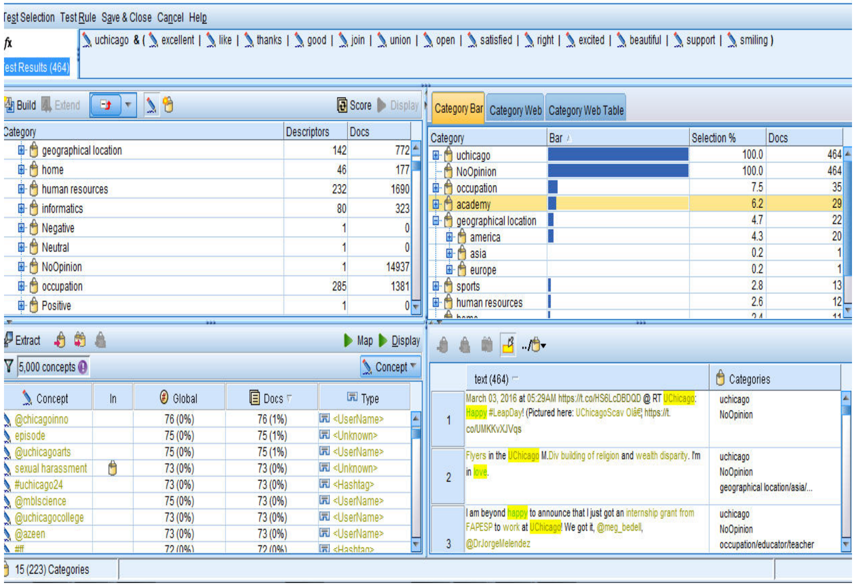
University Name	% of top three languages of tweets distribution			Total # R C	Highest # of tweets	Descriptors & Documents		Extraction of top concepts related to excludes itself		
Uchicago	English 94%	Spanish 2.68	Unknown 0.67%	6,985	1,979	197	3,541	Obama	People	http
Harvard	English 74.82%	Spanish 8.82%	French 2.72%	8,157	2,241	68	182	Evacuates	Build	Word master &leadership
MIT	German 78.84%	English 9.33%	Turkish 3.55%	24,916	10,932	244	537	Science	Innovation	International-HK
NWestern	English 94%	German 1.36%	Romanian 0.7%	60,000	33,631	782	28157	Sports	geographical location	Academy
UIC	English 77.25%	German 7.02%	Spanish 4.08%	22,481	7,500	283	10,826	geographical location	Sports	Sex

[2] Please see appendix for detail charts

Sentiment Analysis towards UChicago

- There are 646 records coming out from tweeter data **POSITIVE**[3] side. We see that people saying happy, in love and express good feeling about Uchicago.
- 7.5% response in the category occupation, 6.2% response in the category academy and 4.7 % in geographical location, 4.3% our 4.7% is from America.

- We see that there are 438 records from tweeter data on **NEGATIVE** side. People express their negative feeling.
- 10.5% response in the category of academy, 7.3% in human resource and 6.2 is in occupation.



uchicago & (excellent | like | thanks | good | join | union | open | satisfied | right | excited | beautiful | support | smiling)

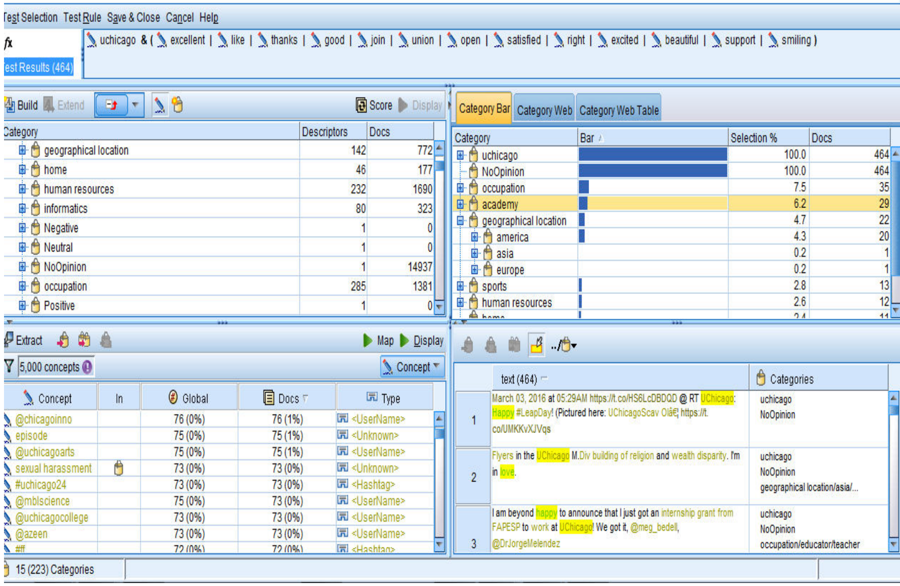
Category	Descriptors	Docs
geographical location	142	772
home	46	177
human resources	232	1690
informatics	80	323
Negative	1	0
Neutral	1	0
NoOpinion	1	14937
occupation	285	1381
Positive	1	0

Category	Bar	Selection %	Docs
uchicago		100.0	464
NoOpinion		100.0	464
occupation		7.5	35
academy		6.2	29
geographical location		4.7	22
america		4.3	20
asia		0.2	1
europe		0.2	1
sports		2.8	13
human resources		2.6	12
home		2.4	11

Concept	In	Global	Docs	Type
@chicagoinno		76 (0%)	76 (1%)	<UserName>
episode		75 (0%)	75 (1%)	<Unknown>
@uchicagoarts		75 (0%)	75 (1%)	<UserName>
sexual harassment		73 (0%)	73 (0%)	<Unknown>
#uchicago24		73 (0%)	73 (0%)	<Hashtag>
@mblscience		75 (0%)	73 (0%)	<UserName>
@uchicagocollege		73 (0%)	73 (0%)	<UserName>
@azeen		73 (0%)	73 (0%)	<UserName>
##f		72 (0%)	72 (0%)	<Hashtag>

15 (223) Categories

	text (464)	Categories
1	March 03, 2016 at 05:29AM https://t.co/HS6LcDBDQD @ RT UChicago: Happy #LeapDay! (Pictured here: UChicagoScav Olâ€¦ https://t.co/UMKKvXJVqs	uchicago NoOpinion
2	Flyers in the UChicago M.Div building of religion and wealth disparity. I'm in love.	uchicago NoOpinion geographical location/asia/...
3	I am beyond happy to announce that I just got an internship grant from FAPESP to work at UChicago! We got it, @meg_bedell, @DrJorgeMelendez	uchicago NoOpinion occupation/educator/teacher



uchicago & (excellent | like | thanks | good | join | union | open | satisfied | right | excited | beautiful | support | smiling)

Category	Descriptors	Docs
geographical location	142	772
home	46	177
human resources	232	1690
informatics	80	323
Negative	1	0
Neutral	1	0
NoOpinion	1	14937
occupation	285	1381
Positive	1	0

Category	Bar	Selection %	Docs
uchicago		100.0	464
NoOpinion		100.0	464
occupation		7.5	35
academy		6.2	29
geographical location		4.7	22
america		4.3	20
asia		0.2	1
europe		0.2	1
sports		2.8	13
human resources		2.6	12
home		2.4	11

Concept	In	Global	Docs	Type
@chicagoinno		76 (0%)	76 (1%)	<UserName>
episode		75 (0%)	75 (1%)	<Unknown>
@uchicagoarts		75 (0%)	75 (1%)	<UserName>
sexual harassment		73 (0%)	73 (0%)	<Unknown>
#uchicago24		73 (0%)	73 (0%)	<Hashtag>
@mblscience		75 (0%)	73 (0%)	<UserName>
@uchicagocollege		73 (0%)	73 (0%)	<UserName>
@azeen		73 (0%)	73 (0%)	<UserName>
##f		72 (0%)	72 (0%)	<Hashtag>

15 (223) Categories

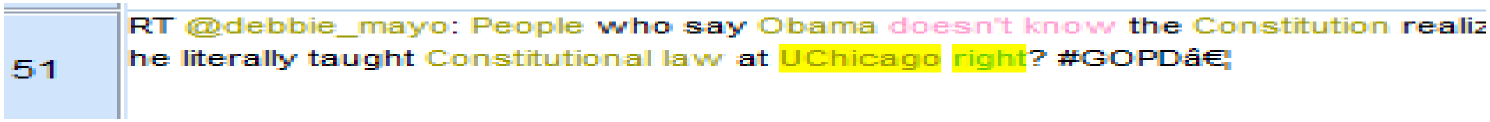
	text (464)	Categories
1	March 03, 2016 at 05:29AM https://t.co/HS6LcDBDQD @ RT UChicago: Happy #LeapDay! (Pictured here: UChicagoScav Olâ€¦ https://t.co/UMKKvXJVqs	uchicago NoOpinion
2	Flyers in the UChicago M.Div building of religion and wealth disparity. I'm in love.	uchicago NoOpinion geographical location/asia/...
3	I am beyond happy to announce that I just got an internship grant from FAPESP to work at UChicago! We got it, @meg_bedell, @DrJorgeMelendez	uchicago NoOpinion occupation/educator/teacher

[3] See tweeter data in Appendix

THE UNIVERSITY OF CHICAGO

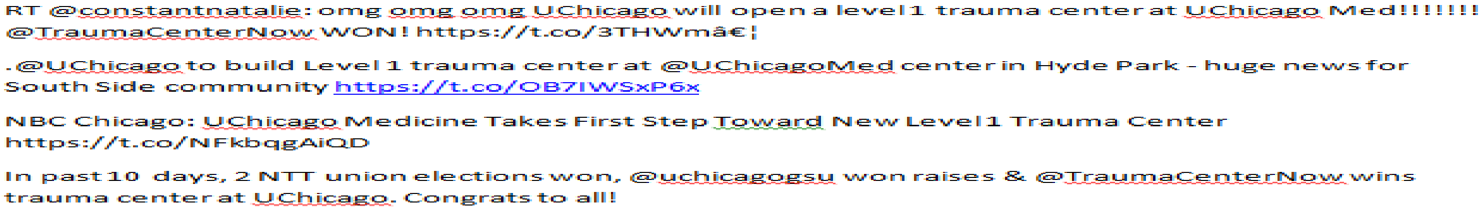
Recommendation to Boost Uchicago’s Name Recognition

- Invite President Obama giving speech as people did not know Obama taught in Uchicago. Below is what people said from tweeter.



51	RT @debbie_mayo: People who say Obama doesn't know the Constitution realiz he literally taught Constitutional law at UChicago right? #GOPDâ€¦

- Organize a tour via trauma center. People are excited about newly build trauma center and want to learn. See tweeters data.



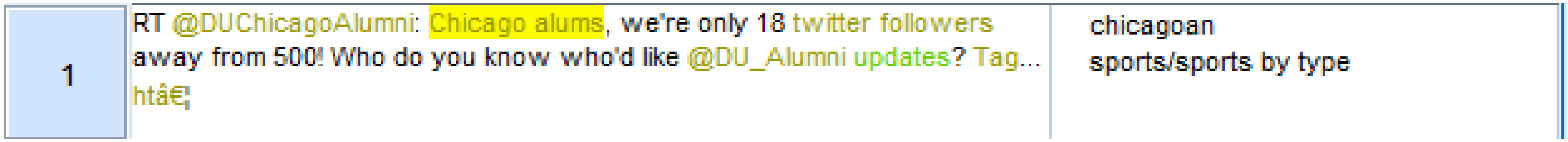
RT @constantnatalie: omg omg omg UChicago will open a level 1 trauma center at UChicago Med!!!!!!! @TraumaCenterNow WON! https://t.co/3THWmâ€¦

.@UChicago to build Level 1 trauma center at @UChicagoMed center in Hyde Park - huge news for South Side community https://t.co/OB7IWSxP6x

NBC Chicago: UChicago Medicine Takes First Step Toward New Level 1 Trauma Center https://t.co/NFkbqgAiQD

In past 10 days, 2 NTT union elections won, @uchicagogsu won raises & @TraumaCenterNow wins trauma center at UChicago. Congrats to all!

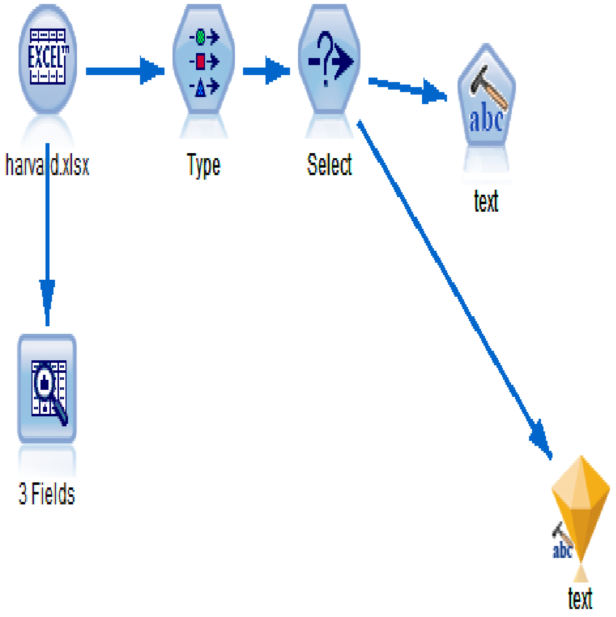
- Hold network events on popular academic topics through alumni from different universities as larger alumni tweeters are active

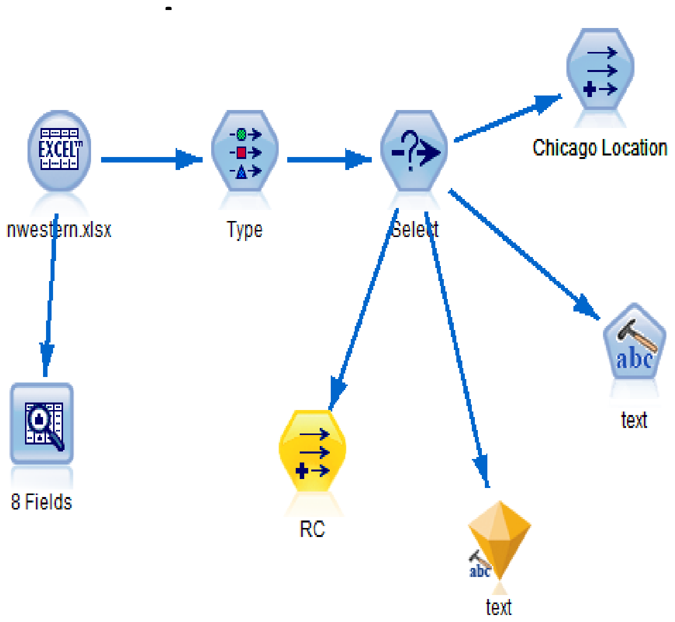


1	RT @DUChicagoAlumni: Chicago alums, we're only 18 twitter followers away from 500! Who do you know who'd like @DU_Alumni updates? Tag... htâ€¦	chicagoan sports/sports by type

Appendix

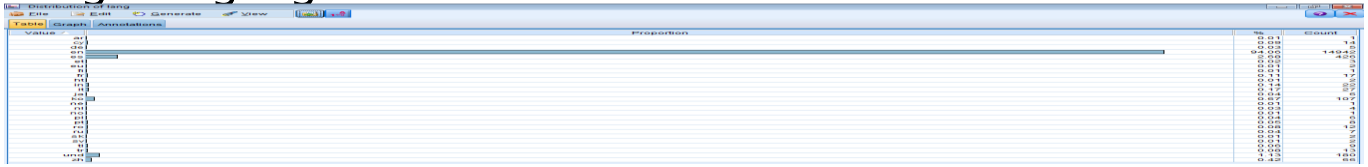
- Basic Models (select Harvard and Nworthern as examples 1





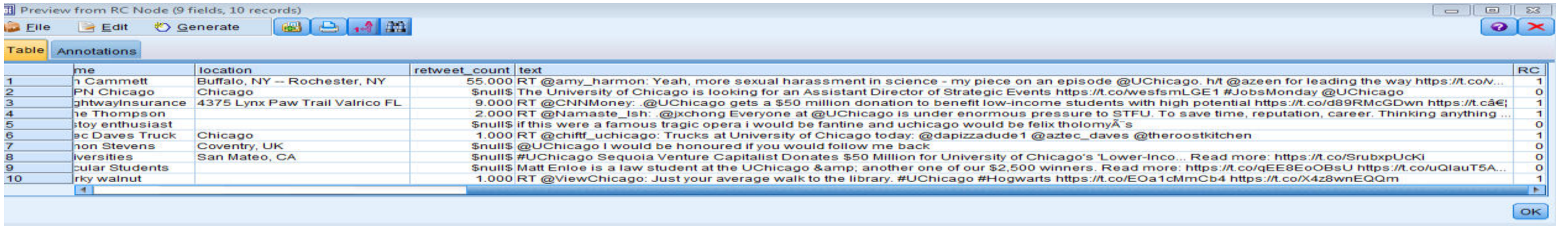
Chicago detail^2

- Chicago language distribution chart



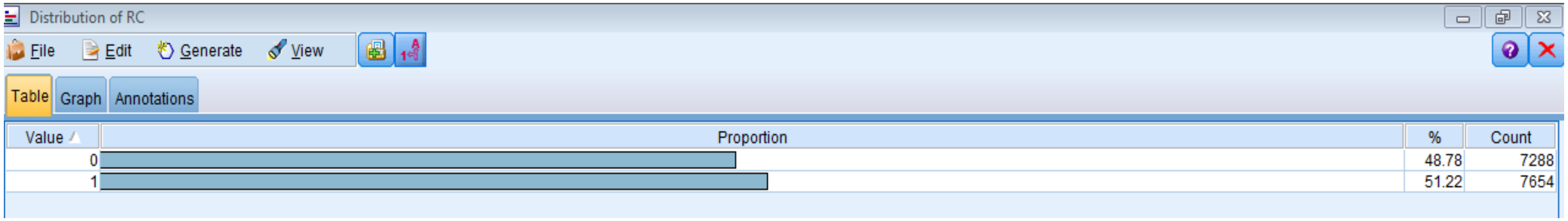
Value	Proportion	%	Count
ar		0.01	1
cy		0.09	14
de		0.03	5
en		94.06	14942
es		2.68	426
et		0.02	3
eu		0.01	2
fi		0.01	1
fr		0.11	17
ht		0.01	2
in		0.14	22
it		0.17	27
ja		0.04	6
ko		0.67	107
ne		0.01	1
nl		0.03	4
no		0.01	1
pl		0.04	6
pt		0.05	8
ro		0.08	12
ru		0.04	7
sk		0.01	2
sv		0.01	2
tl		0.06	9
tr		0.08	13
und		1.13	180
zh		0.42	66

- Separate the original messages vs. re-tweets and assign them different weights / counts



	me	location	retweet_count	text	RC
1	n Cammett	Buffalo, NY -- Rochester, NY	55.000	RT @amy_harmon: Yeah, more sexual harassment in science - my piece on an episode @UChicago. h/t @azeen for leading the way https://t.co/v...	1
2	PN Chicago	Chicago	$null$	The University of Chicago is looking for an Assistant Director of Strategic Events https://t.co/wesfsmLGE1 #JobsMonday @UChicago	0
3	ghtwayInsurance	4375 Lynx Paw Trail Valrico FL	9.000	RT @CNNMoney: .@UChicago gets a $50 million donation to benefit low-income students with high potential https://t.co/d89RMcGDwn https://t.câ€¦	1
4	ne Thompson		2.000	RT @Namaste_Ish: .@jxchong Everyone at @UChicago is under enormous pressure to STFU. To save time, reputation, career. Thinking anything ...	1
5	stoy enthusiast		$null$	if this were a famous tragic opera i would be fantine and uchicago would be felix tholomyÃ¨s	0
6	ec Daves Truck	Chicago	1.000	RT @chiftf_uchicago: Trucks at University of Chicago today: @dapizzadude1 @aztec_daves @theroostkitchen	1
7	non Stevens	Coventry, UK	$null$	@UChicago I would be honoured if you would follow me back	0
8	iversities	San Mateo, CA	$null$	#UChicago Sequoia Venture Capitalist Donates $50 Million for University of Chicago's 'Lower-Inco... Read more: https://t.co/SrubxpUcKi	0
9	cular Students		$null$	Matt Enloe is a law student at the UChicago &amp; another one of our $2,500 winners. Read more: https://t.co/qEE8EoOBsU https://t.co/uQIauT5A...	0
10	rky walnut		1.000	RT @ViewChicago: Just your average walk to the library. #UChicago #Hogwarts https://t.co/EOa1cMmCb4 https://t.co/X4z8wnEQQm	1

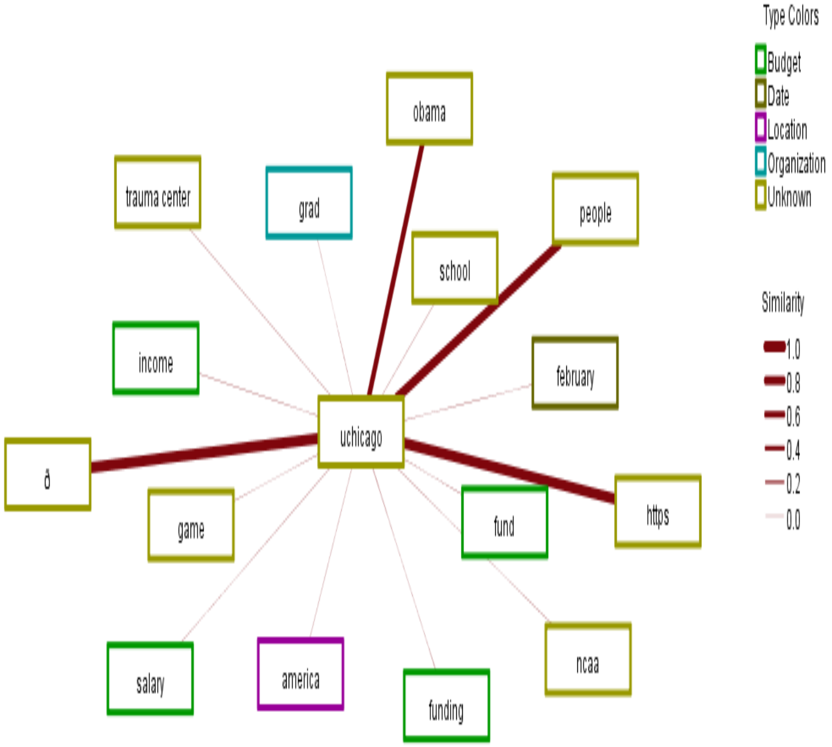
- The frequency of retweet count Vs non retweet counts



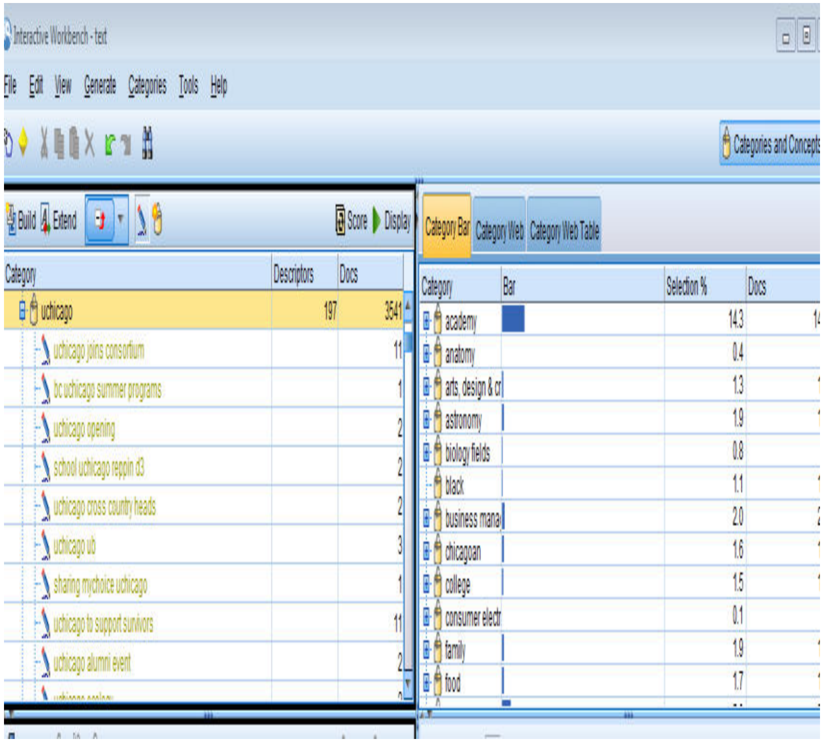
Value	Proportion	%	Count
0		48.78	7288
1		51.22	7654

Chicago Detail^2

Concept extracted



Descriptors and Documents in each Categories

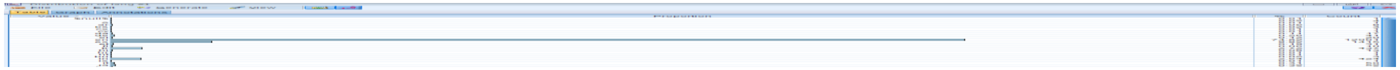

Interactive Workbench - text
File Edit View Generate Categories Tools Help
Categories and Concepts
Build Extend
Score Display
Category Bar Category Web Category Web Table

Category	Descriptors	Docs
uchicago	197	3541
uchicago joins consortium		11
bc uchicago summer programs		1
uchicago opening		2
school uchicago reppin d3		2
uchicago cross country heads		2
uchicago ub		3
sharing mychoice uchicago		1
uchicago to support survivors		11
uchicago alumni event		2

Category	Bar	Selection %	Docs
academy		14.3	14
anatomy		0.4	
arts, design & cr		1.3	1
astronomy		1.9	1
biology fields		0.8	
black		1.1	1
business mana		2.0	2
chicagoan		1.6	1
college		1.5	1
consumer electr		0.1	
family		1.9	1
food		1.7	1

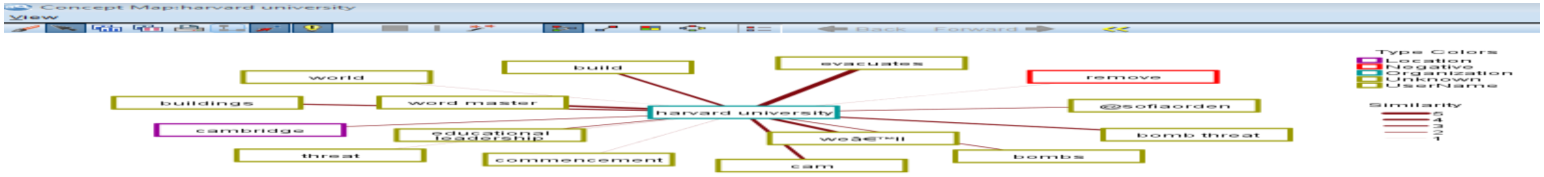
Harvard Detail^2

- language distribution

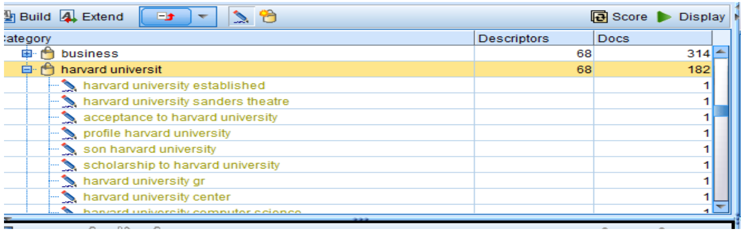


- Rewet count is over 8, 000 and right skewed

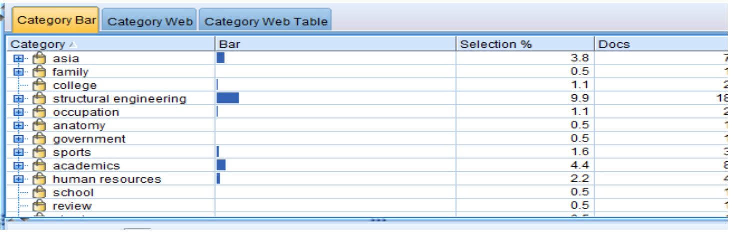
- Concept extracted



- Descriptors and Categories



:ategory	Descriptors	Docs
business	68	314
harvard universit	68	182
harvard university established		1
harvard university sanders theatre		1
acceptance to harvard university		1
profile harvard university		1
son harvard university		1
scholarship to harvard university		1
harvard university gr		1
harvard university center		1
harvard university computer science		1

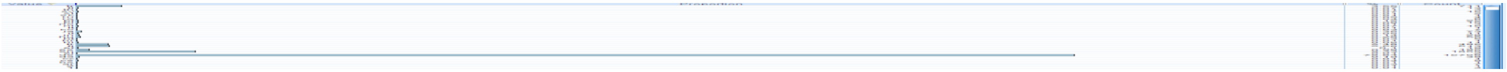


Category	Bar	Selection %	Docs
asia		3.8	7
family		0.5	1
college		1.1	2
structural engineering		9.9	18
occupation		1.1	2
anatomy		0.5	1
government		0.5	1
sports		1.6	3
academics		4.4	8
human resources		2.2	4
school		0.5	1
review		0.5	1

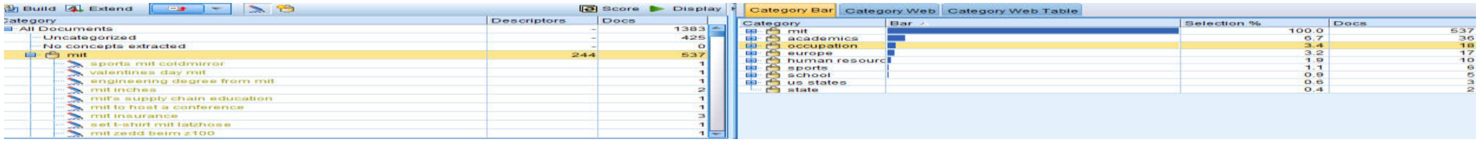
THE UNIVERSITY OF CHICAGO

MIT Detail 2

- language distribution



- Concept extracted

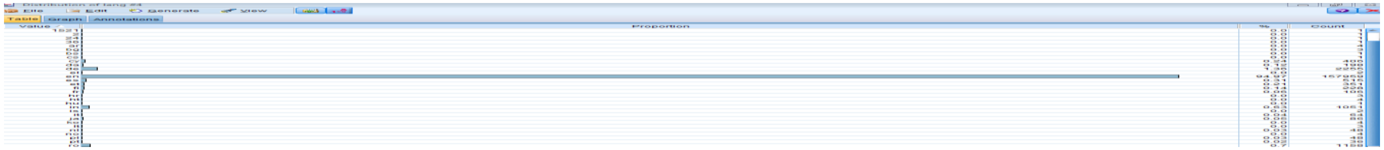


Category	Descriptors	Docs
All Documents	-	1383
Uncategorized	-	425
No concepts extracted	-	0
mit	244	537
sports mit coldmirror		1
valentines day mit		1
engineering degree from mit		1
mit inches		2
mit's supply chain education		1
mit to host a conference		1
mit insurance		3
set t-shirt mit latzhose		1
mit zedd beim z100		1

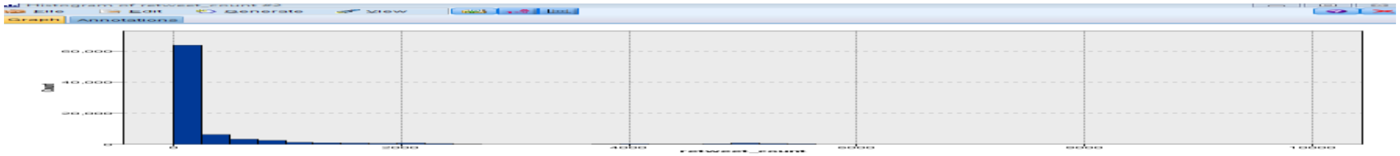
Category	Bar	Selection %	Docs
mit		100.0	537
academics		6.7	36
occupation		3.4	18
europe		3.2	17
human resourc		1.9	10
sports		1.1	6
school		0.9	5
us states		0.6	3
state		0.4	2

NorthWestern^2

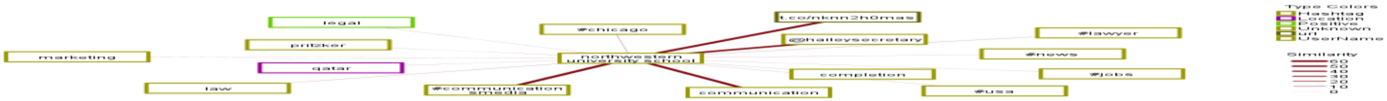
- language distribution



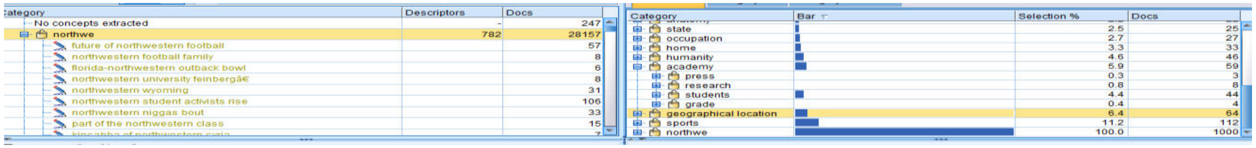
- Frequency of Rewet count



- Concept extracted



- Descriptors and Categories



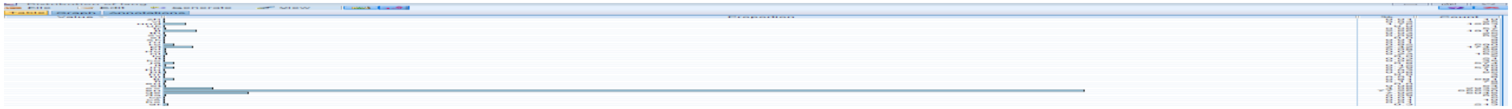
Category	Descriptors	Docs
No concepts extracted	-	247
northwe	782	28157
future of northwestern football		57
northwestern football family		8
florida-northwestern outback bowl		6
northwestern university feinbergâ€		8
northwestern wyoming		31
northwestern student activists rise		106
northwestern niggas bout		33
part of the northwestern class		15
[illegible]		7

Category	Bar	Selection %	Docs
state		2.5	25
occupation		2.7	27
home		3.3	33
humanity		4.6	46
academy		5.9	59
press		0.3	3
research		0.8	8
students		4.4	44
grade		0.4	4
geographical location		6.4	64
sports		11.2	112
northwe		100.0	1000

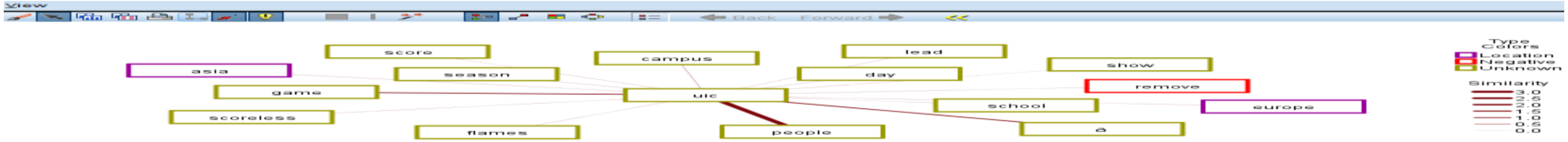
THE UNIVERSITY OF CHICAGO

UIC Detail 2

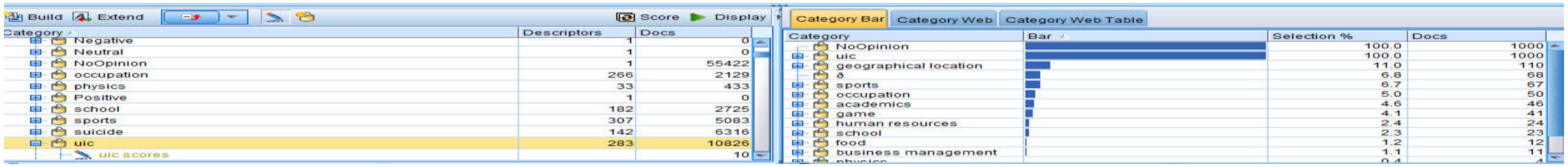
- language distribution



- Concept extracted



- Descriptors and Categories



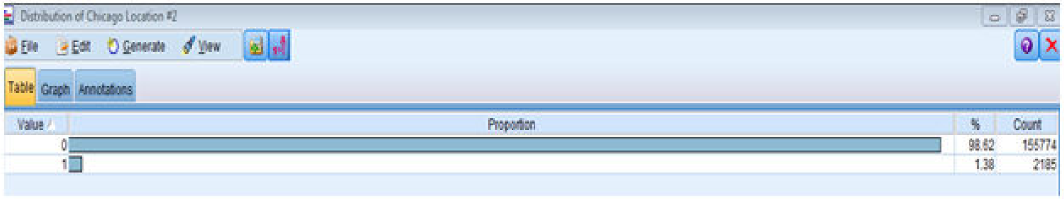
Category	Descriptors	Docs
Negative	1	0
Neutral	1	0
NoOpinion	1	55422
occupation	266	2129
physics	33	433
Positive	1	0
school	182	2725
sports	307	5083
suicide	142	6316
uic	283	10826
uic scores		10

Category	Bar	Selection %	Docs
NoOpinion		100.0	1000
uic		100.0	1000
geographical location		11.0	110
ò		6.8	68
sports		6.7	67
occupation		5.0	50
academics		4.6	46
game		4.1	41
human resources		2.4	24
school		2.3	23
food		1.2	12
business management		1.1	11
physics		0.4	4

Frequency of tweets from Metropolitan universities 2

Chicago

Frequency of tweets from 1.38% total 2,185 counts

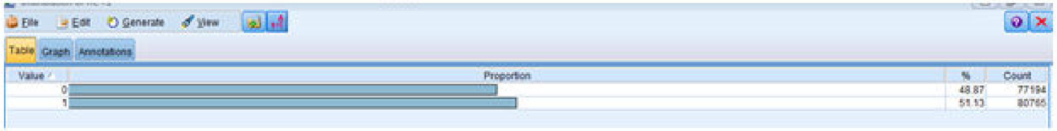


Distribution of Chicago Location #2

Value	Proportion	%	Count
0		98.62	155774
1		1.38	2185

Northwestern

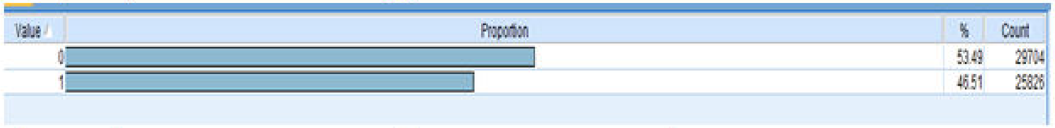
Frequency of retweet counts is : 51.13% total 80765 counts.



Value	Proportion	%	Count
0		48.87	77194
1		51.13	80765

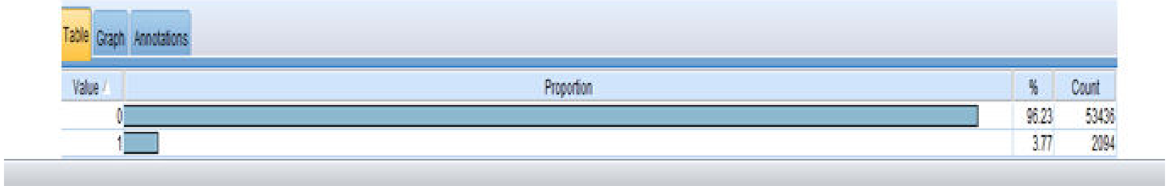
UIC

Frequency of retweet counts is : 46.51% total 25826 counts.



Value	Proportion	%	Count
0		53.49	29704
1		46.51	25826

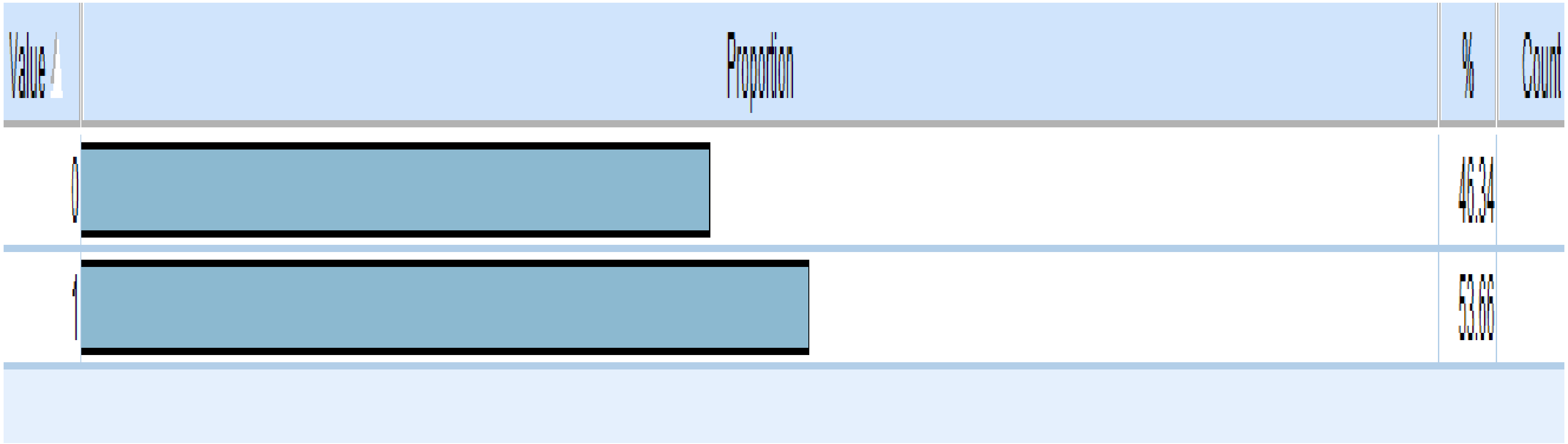
Chicago location : Frequency of tweets: 3.77% total 2094 counts



Value	Proportion	%	Count
0		96.23	53436
1		3.77	2094

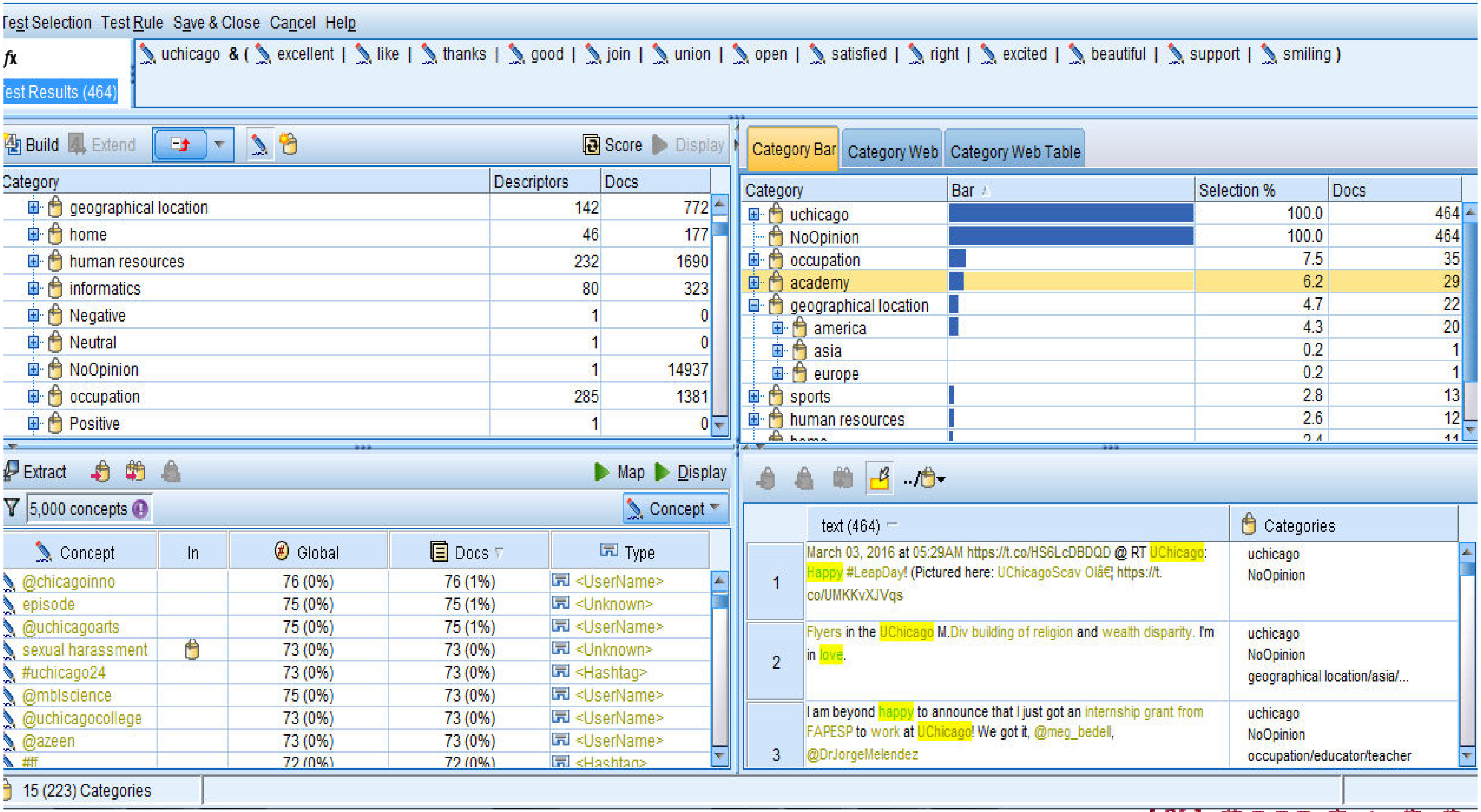
Clustering and segmentation of users with similar tweets / attitudes toward UChicago

- Highest tweet is 6,622
- Frequency of retweets count is 51.24%



Value ▲	Proportion	%	Count
0		46.34	
1		53.66	

Positive feeling on UChicago 3



uchicago & (excellent | like | thanks | good | join | union | open | satisfied | right | excited | beautiful | support | smiling)

Category	Descriptors	Docs
geographical location	142	772
home	46	177
human resources	232	1690
informatics	80	323
Negative	1	0
Neutral	1	0
NoOpinion	1	14937
occupation	285	1381
Positive	1	0

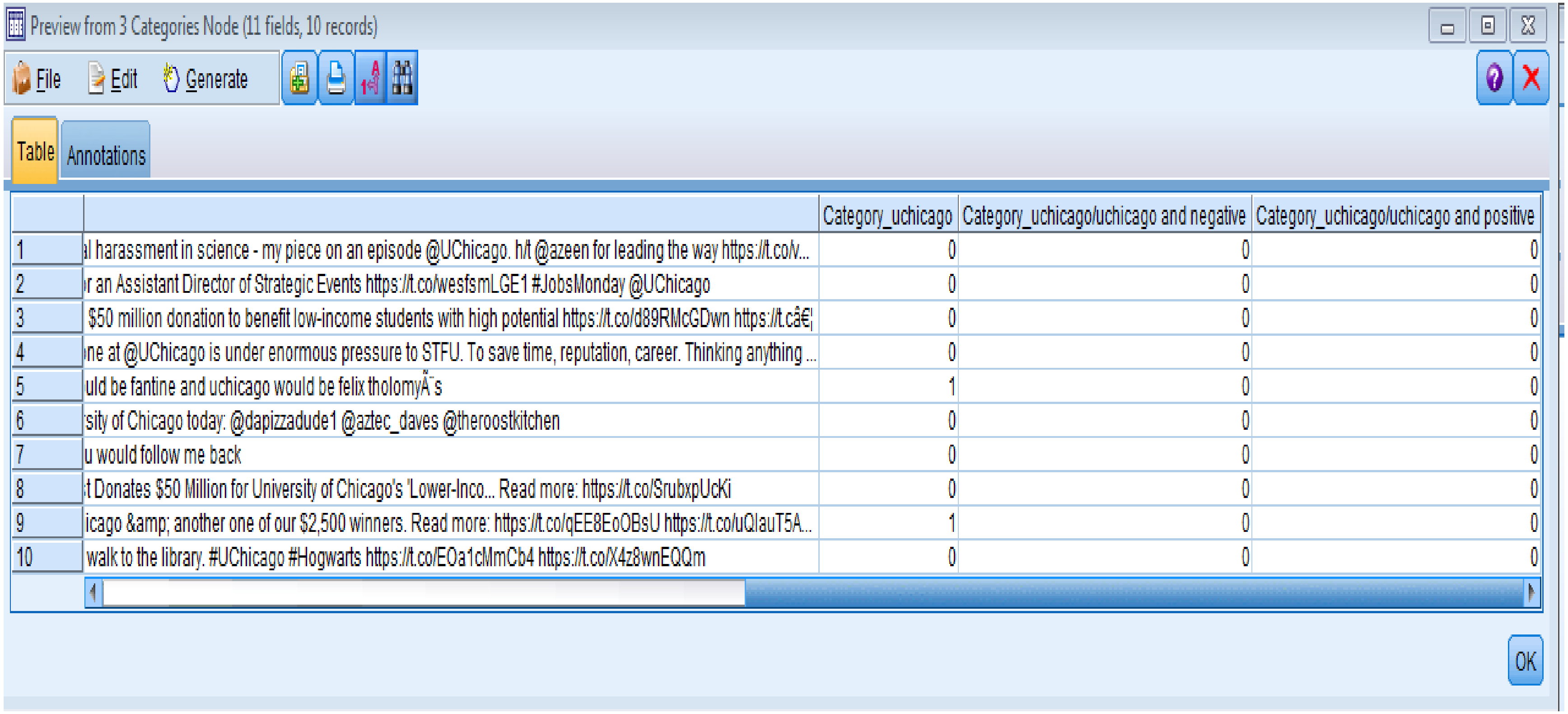
Category	Bar	Selection %	Docs
uchicago		100.0	464
NoOpinion		100.0	464
occupation		7.5	35
academy		6.2	29
geographical location		4.7	22
america		4.3	20
asia		0.2	1
europe		0.2	1
sports		2.8	13
human resources		2.6	12
home		2.4	11

Concept	In	Global	Docs	Type
@chicagoinno		76 (0%)	76 (1%)	<UserName>
episode		75 (0%)	75 (1%)	<Unknown>
@uchicagoarts		75 (0%)	75 (1%)	<UserName>
sexual harassment		73 (0%)	73 (0%)	<Unknown>
#uchicago24		73 (0%)	73 (0%)	<Hashtag>
@mblscience		75 (0%)	73 (0%)	<UserName>
@uchicagocollege		73 (0%)	73 (0%)	<UserName>
@azeen		73 (0%)	73 (0%)	<UserName>
#ff		72 (0%)	72 (0%)	<Hashtag>

15 (223) Categories

	text (464)	Categories
1	March 03, 2016 at 05:29AM https://t.co/HS6LcDBDQD @ RT UChicago: Happy #LeapDay! (Pictured here: UChicagoScav Olâ€¦ https://t.co/UMKKvXJVqs	uchicago NoOpinion
2	Flyers in the UChicago M.Div building of religion and wealth disparity. I'm in love.	uchicago NoOpinion geographical location/asia/...
3	I am beyond happy to announce that I just got an internship grant from FAPESP to work at UChicago! We got it, @meg_bedell, @DrJorgeMelendez	uchicago NoOpinion occupation/educator/teacher

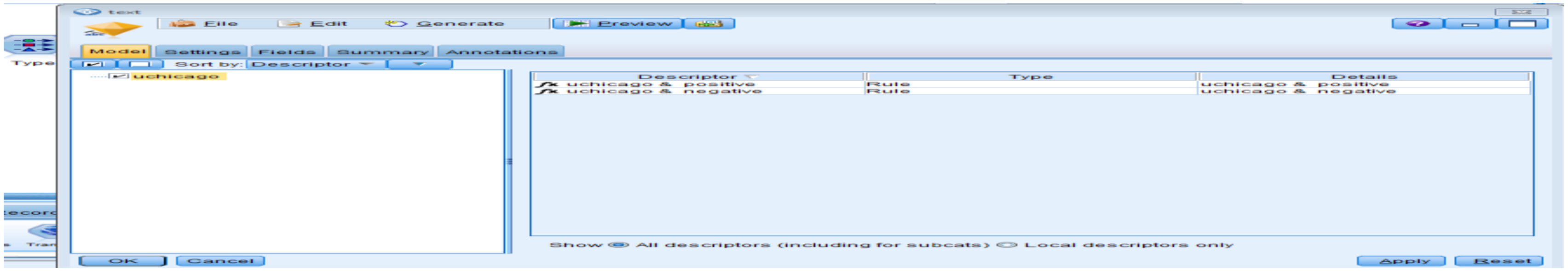
Added Categories as New Fields



		Category_uchicago	Category_uchicago/uchicago and negative	Category_uchicago/uchicago and positive
1	al harassment in science - my piece on an episode @UChicago. h/t @azeen for leading the way https://t.co/v...	0	0	0
2	r an Assistant Director of Strategic Events https://t.co/wesfsmLGE1 #JobsMonday @UChicago	0	0	0
3	$50 million donation to benefit low-income students with high potential https://t.co/d89RMcGDwn https://t.câ€¦	0	0	0
4	ne at @UChicago is under enormous pressure to STFU. To save time, reputation, career. Thinking anything ...	0	0	0
5	uld be fantine and uchicago would be felix tholomyÃ¨s	1	0	0
6	sity of Chicago today: @dapizzadude1 @aztec_daves @theroostkitchen	0	0	0
7	u would follow me back	0	0	0
8	t Donates $50 Million for University of Chicago's 'Lower-Inco... Read more: https://t.co/SrubxpUcKi	0	0	0
9	icago &amp; another one of our $2,500 winners. Read more: https://t.co/qEE8EoOBsU https://t.co/uQlauT5A...	1	0	0
10	walk to the library. #UChicago #Hogwarts https://t.co/EOa1cMmCb4 https://t.co/X4z8wnEQQm	0	0	0

Rules on Chicago sentiment

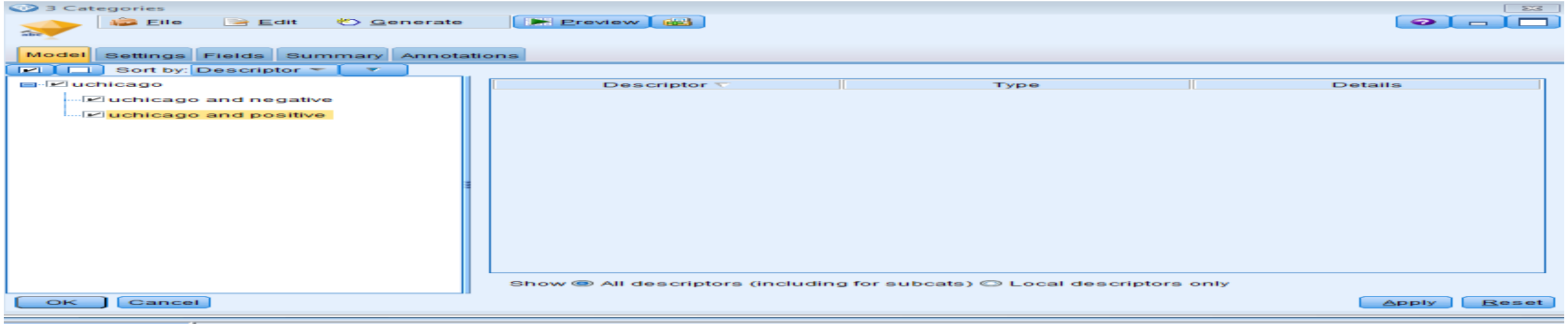
- Add rules



☑ uchicago

Descriptor	Type	Details
fx uchicago & positive	Rule	uchicago & positive
fx uchicago & negative	Rule	uchicago & negative

Show ● All descriptors (including for subcats) ○ Local descriptors only



☑ uchicago
- ☑ uchicago and negative
- ☑ uchicago and positive

Descriptor	Type	Details

Show ● All descriptors (including for subcats) ○ Local descriptors only

Pig script code

```
REGISTER hdfs:///jar/elephant-bird/json-simple-1.1.1.jar;
REGISTER hdfs:///jar/elephant-bird/elephant-bird-core-4.6-SNAPSHOT.jar;
REGISTER hdfs:///jar/elephant-bird/elephant-bird-pig-4.6-SNAPSHOT.jar;
REGISTER hdfs:///jar/elephant-bird/elephant-bird-hadoop-compat-4.6-SNAPSHOT.jar;
REGISTER hdfs:///jar/tutorial.jar;

A = load '/user/kadochnikov/twitter_full/' USING com.twitter.elephantbird.pig.load.JsonLoader('-nestedLoad') as tweets;
C = FOREACH A GENERATE
(CHARARRAY)tweets#'id' as id,
(CHARARRAY)tweets#'lang' AS lang,
(CHARARRAY)tweets#'created_at' AS created_at,
(CHARARRAY)tweets#'text' AS text,
tweets#'user' as user_info,
tweets#'retweeted_status' as retweet_info
;
D = FOREACH C GENERATE
id,
lang,
created_at,
(CHARARRAY)user_info#'screen_name' as screen_name,
(CHARARRAY)user_info#'name' as name,
(CHARARRAY)user_info#'location' as location,
(INT)retweet_info#'retweet_count' as retweet_count,
REPLACE(text, '\n', ' ') AS text
; E = FILTER D BY (text matches '(?i).*uchicago.*'); --(your other filter words go here)

rmf /user/joanzhang/Test_Tweet

--forcing a single file output using PARALLEL 1
reduced = FOREACH (GROUP E BY RANDOM()) GENERATE FLATTEN(E) PARALLEL 1;
STORE reduced INTO '/user/joanzhang/Test_Tweet' using PigStorage('\t','-schema');

fs -getmerge -nl /user/joanzhang/Test_Tweet /home/joanzhang/Tweet/reduce.txt

--on local Linux, run the command to reserver header and remove other staff of pig
awk 'NR==1 || NR>4 {print}' /home/joanzhang/Tweet/reduce.txt > /home/joanzhang/Tweet/tweet.txt
```