SeekΣdgar

Rajendra P. Srivastava
PhD (Physics) & PhD (Accounting)
EY Professor of Accounting and Information Systems
School of Business, The University of Kansas
Lawrence, KS 66045

Email 1: rsrivastava@ku.edu
Email 2: rsrivastava@seekedgar.com
Cell Phone: 785-218-4409

SeekEdgar website: https://www.seekedgar.com

SeekEdgar

Introduces

SeekiNF

Anytime, Anywhere, …

SEC Filings and PCAOB Reports at your finger tips in seconds

For a quick introduction, watch this 6 minutes video at
https://www.youtube.com/watch?v=wR3TqoQpYQk&feature=youtu.be

SeekiNF

Example 12

Example 12a: Cosine Similarity Measure

&

Example 12b: Word Variations Year-to-Year

SeekiNF

Example 12a: Cosine Similarity Measure

For illustration, we consider two Indian IT companies, Satyam Computer Services Ltd, CIK = 1106056, and WIPRO Ltd., CK = 1123799, both filed 20-F with the SEC. In January 2009, Satyam's Chairman Ramalinga Raju sent a confession letter to SEBI (Securities and Exchange Board of India) admitting that he inflated the financial results and cooked books of Satyam. We would like to compare the Cosine Similarity measure of Satyam, a fraudulent company, with a competing non-fraudulent company, WIPRO Ltd. Here is what we need to do to get the Cosine Similarity measure for 20-F for the period 2002 to 2008.

Step 1: Type "a|the" without the quotes in the slot of "With the exact phrase". Or, you can use any search criteria including Boolean logic, if that is relevant for your search.
Step 2: Type either "Satyam Computer Services" and select option "Name" from the right or type the CIK code and select "CIK" option from the right. Select the time period 2002 – 2008.
Step 3: Select "Paragraph (All)".
Step 4: Select 20-F and 20-F/A
Step 5: Submit

You will get the result in seconds. The next slide shows how you enter the above information and the following slide shows the Cosine Similarity and Word Variation buttons at the top right of the display window. Click on this button. The system will open another widow. Select the CIK that is already present (CIK of the company just searched) or type in another CIK that you want to get the compare data. Select the filing, in the present case 20-F, and the period 2002-2008 and then "Search". The system will list all the filings for your compare data. Now you select the specific year filings for the compare data or select all the years. In the present example, I will do it two at a time keeping 2003 year common. This is because I want to get Cosine Similarity measure relative to 2003 Filing. See the subsequent slides for details.

HOME | ABOUT US | SeekiNF | FRAANK | FAQs | CONTACT US | Welcome Srivastava ▼

Recent Press and Other Releases | 10-K Exhibit 21(Subsidiaries) | Conference Call Transcripts | Search | Request Form | Special Request | SeekiNF Examples |

Example 12a: Cosine Similarity Measure

STEP 1: Please enter Phrase(s)/Keyword(s) you wish to search:

With the exact phrase: `a|the`

The following features are for Single Phrase/Word only: (In case of Multiple Phrases/Words, please use the Request Form to submit the request.)

☐ Phrase/Word Counter

☐ Display words before & after

Number of words before: _____   Number of words After: _____

Proximity Search: _____

Please input within how many words: _____

With all of the words: _____

With at least one of the words: _____

Without the words: _____

STEP 2: Please select the options below (Default - ALL):

COMPANY: SATYAM COMPUTER SERVICES LTD,   [Name ▼]

From: [2002 ▼]   To: [2008 ▼]

SIC (Optional): [All SICs ▼]

STEP 3: Please select the search Resolution here:

● Paragraph(All)   ○ Footnote   ○ SOX 404 Mgt Report   ○ MD&A
○ Table   ○ Audit Report   ○ SOX 404 Audit Report

STEP 4: Please select the Document(s) you want to search here:

○ All
● Select (Please CLICK HERE for Filing Description):   ○ Un-Check All

Annual Reports   ○ Check All Annual Reports   ○ Un-Check All Annual Reports
☐ 10-K   ☐ 10-KT   ☐ 10-K405   ☐ 10KT405   ☐ 10KSB   ☐ 10KSB40
☐ 10-K/A   ☐ 10-KT/A   ☐ 10-K405/A   ☐ 10KT405/A   ☐ 10KSB/A   ☐ 10KSB40/A

Quarterly, Current & Foreign Annual Reports
☐ 10-Q   ☐ 10-QT   ☐ 10QSB   ☐ 8-K   ☐ 6-K   ☑ 20-F   ☐ 40-F
☐ 10-Q/A   ☐ 10-QT/A   ☐ 10QSB/A   ☐ 8-K/A   ☐ 6-K/A   ☑ 20-F/A   ☐ 40-F/A

Proxies & Registrations
☐ DEF 14A   ☐ DEFM14A   ☐ S-1   ☐ S-4   ☐ 15-12B   ☐ 15-12G
☐ DEFA14A   ☐ DEFM14C   ☐ S-1/A   ☐ S-4/A   ☐ 15-12B/A   ☐ 15-12G/A
☐ 15-12D   ☐ 15F-12B   ☐ 15F-12G   ☐ 15F-12D   ☐ PRE 14A
☐ 15-12D/A   ☐ 15F-12B/A   ☐ 15F-12G/A   ☐ 15F-12D/A

Ownership & Prospectuses
☐ Form 3   ☐ Form 4   ☐ Form 5   ☐ 424B1   ☐ 424B3   ☐ 424B5
☐ Form 3/A   ☐ Form 4/A   ☐ Form 5/A   ☐ 424B2   ☐ 424B4   ☐ 424B7
☐ 13F-HR   ☐ 13F-NT   ☐ SC 13D   ☐ SC 13G   ☐ 424B8
☐ 13F-HR/A   ☐ 13F-NT/A   ☐ SC 13D/A   ☐ SC 13G/A

Other Filings
☐ N-CSR   ☐ N-CSRS   ☐ NSAR-A   ☐ NSAR-AT   ☐ NSAR-B
☐ N-CSR/A   ☐ N-CSRS/A   ☐ NSAR-A/A   ☐ NSAR-AT/A   ☐ NSAR-B/A
☐ NSAR-BT   ☐ NSAR-U   ☐ N-Q   ☐ SD   ☐ CORRESP (Comment Letters)
☐ NSAR-BT/A   ☐ NSAR-U/A   ☐ N-Q/A   ☐ SD/A   ☐ UPLOAD (Response Letters)
☐ AAER**

10-K, 10-Q & 8-K Exhibits & Shareholders Letters
☐ PRESS RELEASE   ☐ SHAREHOLDERS MEETINGS   ☐ CONFERENCE CALLS
☐ OTHER 8-K EX   ☐ SHAREHOLDERS LETTER*   ☐ EXHIBIT 21 (10-K & 10-Q)
☐ EXHIBIT 95 (10-K & 10-Q)   ☐ EXHIBIT 10 (10-K)   ☐ EXHIBIT 10 (10-Q)
☐ EXHIBIT 99 (10-K)   ☐ EXHIBIT 99 (10-Q)   ☐ ITEM 1A (10-K)

Public Company Accounting Oversight Board
☐ INSPECTION REPORTS**   ☐ SETTLED DISCIPLINARY ORDERS**
☐ ADJUDICATED DISCIPLINARY ACTIONS**

*Beta Version: We are working on improving the accuracy of the letters
**AAER & PCAOB do not have CIK's, they have File number and Firm number respectively

STEP 5: SUBMIT

SeekiNF

Example 12a: Compare: Cosine Similarity Measure

SeekiNF

Anytime, Anywhere...
SEC filings at your fingertips in seconds with SeekiNF

Example 12: Compare

| HOME | ABOUT US | **SeekiNF** | FRAANK | FAQs | CONTACT US | Welcome Srivastava ▾ |

| Recent Press and Other Releases | 10-K Exhibit 21(Subsidiaries) | Conference Call Transcripts | Search | Request Form | Special Request | Guidelines and Examples |

CIK : `1106056` [Change CIK]   **Year :** [2002 ▾] to [2008 ▾]

File Type:

○10-K	○8-K	●20-F	○Form-3	○424B1	○424B4	○15F-12B	○DEF 14A	○N-CSR	○UPLOAD	○MDNA 10-K
○10-Q	○6-K	○40-F	○Form-4	○424B2	○424B5	○15F-12G	○DEFM14A	○NSAR-A	○CORRESP	○MDNA 10-Q
○N-Q	○S-1	○11-K	○Form-5	○424B3	○424B7	○15F-12D	○DEFM14C	○NSAR-B	○ITEM 1A	○FOOTNOTE 10-K
○SD	○S-4	○15-12B	○15-12G	○15-12D	○424B8	○SC 13D	○SC 13G	○NSAR-U	○13F-HR	○FOOTNOTE 10-Q

Cosine Similarity measures how close two documents are. **Word Variation** compares and provides the frequency of words that appear in two documents.

Search

The list of 20-F 's filed between 2002 - 2008 years for CIK : 1106056

Cosine Similarity			**Word Variation**
Filing	**SEC File Link**	**Filing Date**	**Word Distribution**
☑20-F	https://www.sec.gov/Archives/edgar/data/1106056/0001145549-08-001441-index.htm	08-08-2008	Download Distribution
☐20-F	https://www.sec.gov/Archives/edgar/data/1106056/0001145549-07-000670-index.htm	04-30-2007	Download Distribution
☐20-F	https://www.sec.gov/Archives/edgar/data/1106056/0001145549-06-000579-index.htm	04-28-2006	Download Distribution
☐20-F	https://www.sec.gov/Archives/edgar/data/1106056/0001145549-05-000706-index.htm	04-28-2005	Download Distribution
☐20-F/A	https://www.sec.gov/Archives/edgar/data/1106056/0001145549-04-000961-index.htm	07-26-2004	Download Distribution
☐20-F	https://www.sec.gov/Archives/edgar/data/1106056/0001145549-04-000850-index.htm	06-29-2004	Download Distribution
☑20-F	https://www.sec.gov/Archives/edgar/data/1106056/0001145549-03-000795-index.htm	06-30-2003	Download Distribution
☐20-F	https://www.sec.gov/Archives/edgar/data/1106056/0001021408-02-009039-index.htm	07-01-2002	Download Distribution

SeekiNF

Example 12a: Cosine Similarity Measure Data for Satyam and WIPRO

Once your click on the Cosine Similarity button after selecting the two years filings, the system will yield the result in a CSV file which you can open and save in Excel. Repeating this process for each pair, (2003, 2008), (2003, 2007), (2003, 2006), (2003, 2005), (2003, 2005), and (2003, 2004), I obtained the following data.

Cosine Similarity Scores 20F		File 1 VS File 2					
Satyam	CIK	2008 VS 2003	2007 VS 2003	2006 VS 2003	2005 VS 2003	2004 20F/A VS 2003	2004 VS 2003
	1106056	0.986219031	0.986518731	0.986687789	0.983219142	0.884167909	0.995569951
WIPRO	CIK	2008 VS 2003	2007 VS 2003	2006 VS 2003	2005 VS 2003	2004/A VS 2003	2004 VS 2003
	1123799	0.992107956	0.993143381	0.993785122	0.992941528	0.802915992	0.996115982

The following data was obtained by rearranging the above data. See the next slide for the plots

Year	Satyam Cosine Measure	WIPRO Cosine Measure
2008	0.986219031	0.992107956
2007	0.986518731	0.993143381
2006	0.986687789	0.993785122
2005	0.983219142	0.992941528
2004	0.995569951	0.996115982
2003	1.0	1.0

Anytime, Anywhere, ..

Example 12a: Graph of Cosine Similarity Data for Satyam and WIPRO

Cosine Measure of Similarity for Satyam with WIPRO in relation to 2003 20-F
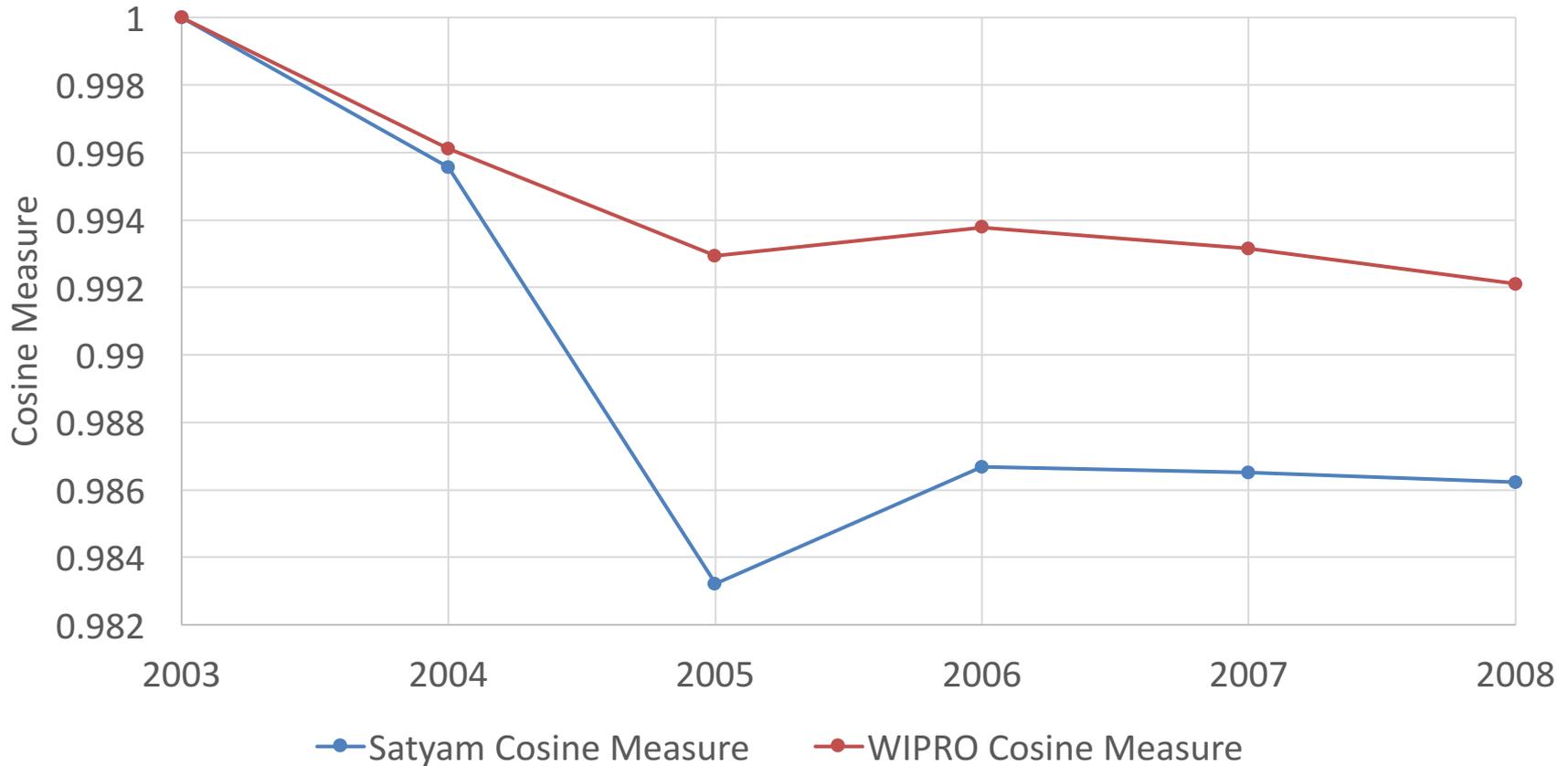
SeekiNF

Example 12: Cosine Similarity Measure and Word Variability

Example 12b:

Word Variation Year-to-Year

SeekiNF

Example 12b: Word Variations Year-to-Year

- Let us consider the example of Satyam, again. In Example 12a, we selected "Cosine Similarity". For this example, after selecting two years 20-F, we select "Word Variation". The system will yield the data in a csv file that you can open and save in Excel.

- Next slide shows the un-sorted data for the years 2007 and 2008.

- The second slide shows, after sorting, the words that are added in 20-F for 2008 that were not present in 20-F of 2007.

- The third slide shows the words (after sorting again) that are removed from 20-F of 2008, but were present in 20-F of 2007.

- One can sort the file and obtain changes in certain key words from year to year.

SeekiNF

WORD VARIATIONS FOR 1106056		
	https://\	https://www.sec.gov/Archives/edgar/data/1106056/0001145549-07-000670-index.html
WORDS	2008	2007
begin	1	1
privacy	3	3
enhanced	6	6
message	2	2
proc	1	1
type	14	15
2001,mic	1	1
clear	1	1
originator	2	2
name	16	20
webmaster	1	1
www.sec.gov	2	2
key	37	36
asymmetric	1	1
mfgwcgyevqgbaqicaf8d	1	1
twsm7vrzladbmyqaionv	1	1
b	55	59
twidaqab	1	1
mic	1	1
info	1	1
rsa	1	1

Example 12b: Word Variations Year-to-Year

un-sorted data for Satyam for the years 2007 and 2008

SeekiNF

Example 12b: New Words in 20-F of 2008 that were not present in 20-F of 2007 for Satyam

WORD VARIATIONS FOR 1106056					
	https://w	https://www.sec.gov/Archives/edgar/data/1106056/0001145549-07-000670-index.html			
WORDS	2008	2007			
nitor	22	0			
afm	13	0			
bridge	10	0			
euronext	9	0			
belts	6	0			
deconsolidation	6	0			
lieu	6	0			
pharma	6	0			
hetzel	5	0			
caterpillar	4	0			
chicago	4	0			
conditional	4	0			
contributed	4	0			
fbt	4	0			
fmsa	4	0			
infocity	4	0			

Words in 20-F of 2008 that are not present in 20-F of 2007 for Satyam

SeekiNF

Example 12b: Words absent in 20-F of 2008, but were present in 20-F of 2007 for Satyam

WORD VARIATIONS FOR 1106056						
	https://w	https://www.sec.gov/Archives/edgar/data/1106056/0001145549-07-000670-index.html				
WORDS	2008	2007				
acceleration	0	6				
contingency	0	6				
fulfillment	0	6				
resolved	0	6				
upto	0	6				
allotment	0	5				
allotted	0	5				
bommakanti	0	5				
concerned	0	5				
rank	0	5				
redeem	0	5				
acquirer	0	4				
allocation	0	4				
consequence	0	4				
decided	0	4				
get	0	4				
heirs	0	4				
murty	0	4				
procedure	0	4				
quantifying	0	4				
quantum	0	4				
referral	0	4				
tranche	0	4				
triggering	0	4				

Words in 20-F of 2007 that are not present in 20-F of 2008 for Satyam

SeekiNF

Thanks!

Email: rsrivastava@seekedgar.com