Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC-Bank Indonesia Satellite Seminar on *"Big Data"* at the ISI Regional Statistics Conference 2017

Bali, Indonesia, 21 March 2017

Finding similar words in Big Data - Text mining approach of semantic similar words in the Federal Reserve Board members' speeches[1]

Christian Dembiermont and Byeungchun Kwon,
Bank for International Settlements

---

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Finding similar words in Big Data

Text mining approach of semantic similar words in the Federal Reserve Board members' speeches

Christian Dembiermont and Byeungchun Kwon

Data Bank Services, Monetary and Economic Department,

Bank for International Settlements

Irving Fisher Committee - Bank Indonesia Satellite Seminar on "Big Data"

Bali, 21 March 2017

Overview

- Finding words in a corpus of thousands of documents is a difficult task
- Finding similar words in this corpus is a daunting task

- Business case: finding similar words to "forward"
- Solution: new text mining technology "Word2Vec"

Detection of words with similar meaning

Central hypothesis:

Linguistic items with similar distributions have similar meaning

sentence #1 – The Economic Outlook and Monetary Policy, Janet L. Yellen, Apr 2012

Because any economic forecast is inherently uncertain, the FOMC's **forward** policy guidance states explicitly that the Committee "currently anticipates

sentence #2 – International Financial Regulatory Reform, Jerome H. Powell, Jul 2013

I believe there is a trust between us that is the basis for collaboration. I look **forward** to working with you to make the financial system safer and stronger.

⋮

sentence #XXXXX – Monetary Policy & the Housing Bubble, Ben S. Bernanke, Jan 2010

The low policy rates during the 2002-06 period were accompanied at various times by "**forward** guidance" on policy from the Committee.

Big data

- 1,241 speeches
- over 100,000 sentences
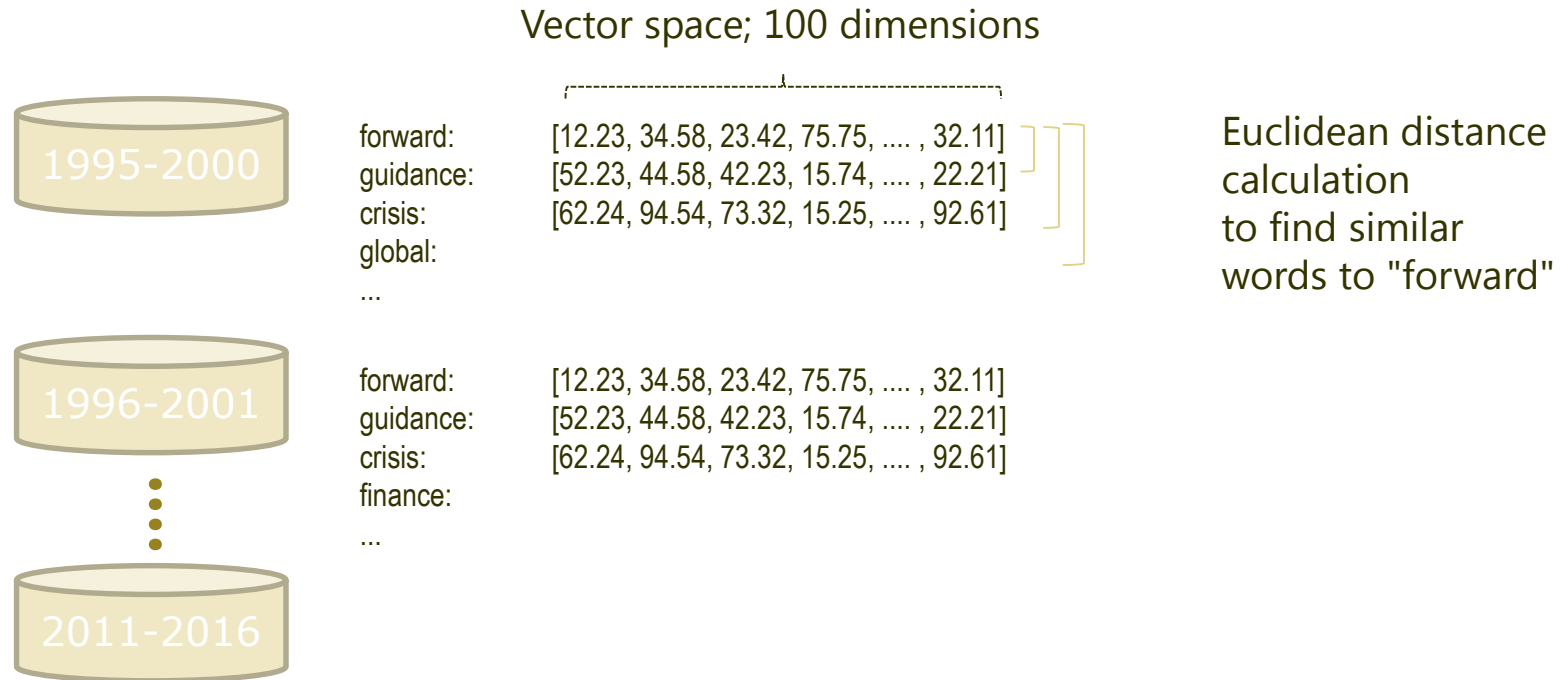
Text mining

- Two-layer neural networks
- Assign corpus to a vector space

Semantic Similarity Database

Semantic Similarity Database

Calculation of the Euclidean distance between two Word vectors

Vector space; 100 dimensions

1995-2000		
forward:	[12.23, 34.58, 23.42, 75.75, ... , 32.11]	
guidance:	[52.23, 44.58, 42.23, 15.74, ... , 22.21]	
crisis:	[62.24, 94.54, 73.32, 15.25, ... , 92.61]	
global:		
...		

Euclidean distance calculation to find similar words to "forward"

1996-2001		
forward:	[12.23, 34.58, 23.42, 75.75, ... , 32.11]	
guidance:	[52.23, 44.58, 42.23, 15.74, ... , 22.21]	
crisis:	[62.24, 94.54, 73.32, 15.25, ... , 92.61]	
finance:		
...		

2011-2016

Behind the Semantic Similarity Database: Word2Vec

- Word2vec
    - created by a team of researchers led by Tomas Mikolov (Google)
    - input: a large corpus of text
    - output:
        - a vector space, typically of several hundred dimensions
        - each unique word in the corpus being assigned a corresponding vector in the space
        - word vectors are positioned in the vector space such that words that share common contexts in the corpus are located in close proximity to one another in the space
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. ICLR.

Live demo *(available on http://centralbankersapp.com/)*

Results of the findings

Similar words to "forward" are:

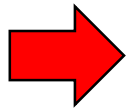Federal Reserve Board members' speeches: 1995-2007
 forecasts, incoming, ahead, carefully

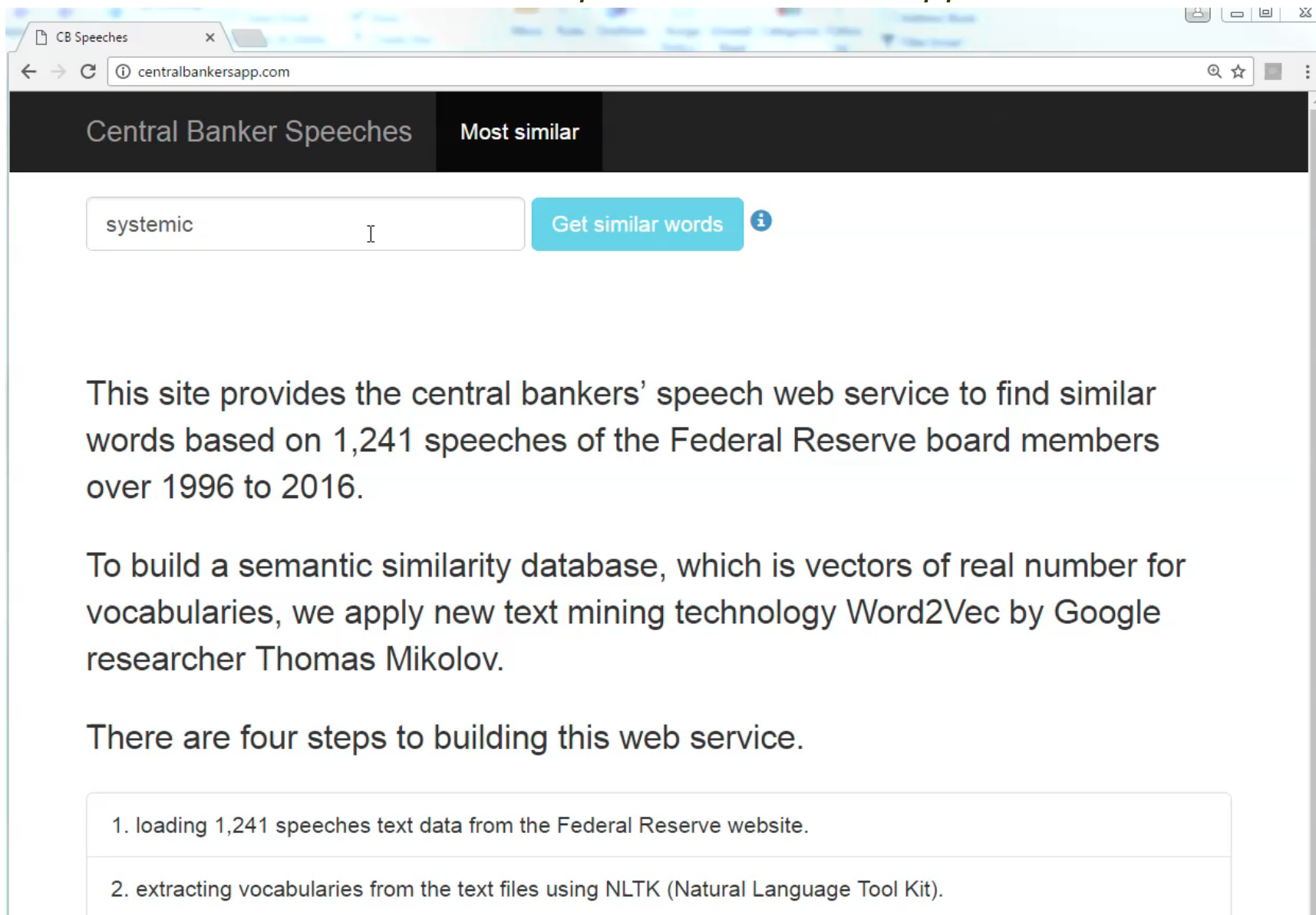Federal Reserve Board members' speeches: 2008-2016
 guidance, communicate, intention, path

p.m.: Standard dictionary:
 ahead, leading, onward, forth

Live demo *(available on http://centralbankersapp.com/)*
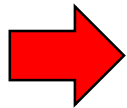
Results of the findings

Similar words to "systemic" are:

Federal Reserve Board members' speeches: 1995-2007
 hazard, moral, soundness, operations, sensitivity, taking

Federal Reserve Board members' speeches: 2008-2016
macroprudential, interconnectedness, failure, structure
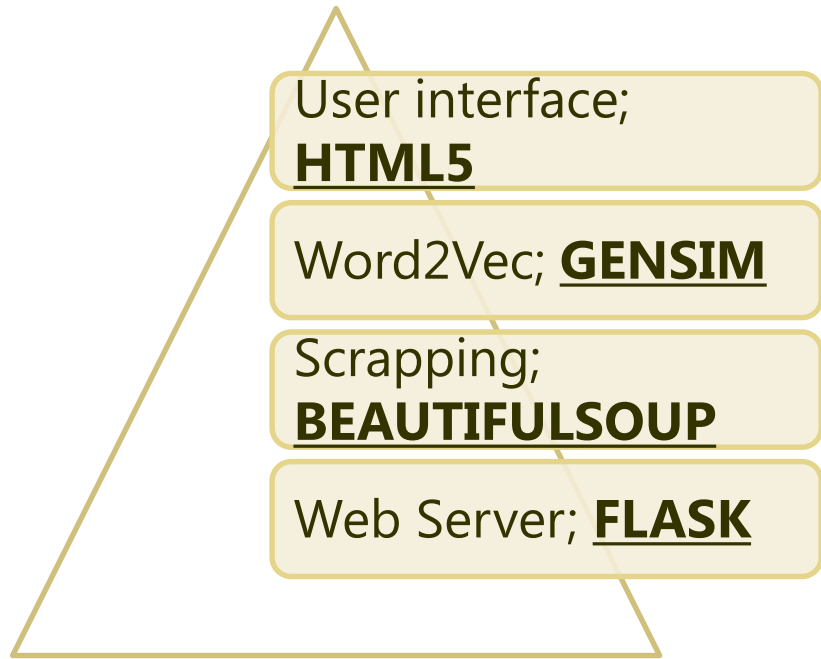
BANK FOR
INTERNATIONAL
SETTLEMENTS

Conclusion

Characteristics of the Word2Vec text mining technique

- Applied for the first time to a central bank domain

- Applied for the first time to analyze similarity between words

- Improved the text mining beyond the word cloud (used in CBs)

- Able to trace similarity evolution over time

- Does not provide any economic forecasts or causality analysis

Source code

- All codes are written in Python language and are available at
  http://github.com/Byeungchun/centralbankersword2vec

User interface;
HTML5

Word2Vec; **GENSIM**

Scrapping;
BEAUTIFULSOUP

Web Server; **FLASK**

Thank you!