

# Natural Language Processing

# Communication With Machines



~50-70s

```
Elite Edit Edit_Settings Menu Utilities Compilers Test Help
EDIT      BS90.DEV13.CLIOPRO(TIMMIES) - 01.31
Command > **** Top of Data ****
000001 /* REXX EXEC ****
000002 /*
000003 /* TIMMIES FACTOR - COMPOUND INTEREST CALCULATOR
000004 /*
000005 /* AUTHOR: PAUL GRIMBLE
000006 /* DATE: OCT 1/2007
000007 /*
000008 /*
000009 ****
000010 /*
000011 /*
000012 say '*****'
000013 say 'Welcome Coffee drinker.'
000014 say '*****'
000015 DO WHILE DATATYPE(CoffeeAmt) \= 'NUM'
000016   say ""
000017   say "What is the price of your coffee?","
000018   say "  "(e.g. 1.58 = $1.58)"
000019   parse pull CoffeeAmt
000020 END
000021 /*
000022 DO WHILE DATATYPE(CoffeeWk) \= 'NUM'
000023   say ""
000024   say "How many coffees a week do you have?"
000025   parse pull CoffeeWk
000026 END
000027 /*
000028 DO WHILE DATATYPE(Rate) \= 'NUM'
000029   say ""
000030   say "What annual interest rate would you like to see on that money?","
000031   say "  "(e.g. 8 = 8%)"
000032   parse pull Rate
000033 END
000034 Rate = Rate * .01 /* CHG TO DECIMAL NUMBER */
```

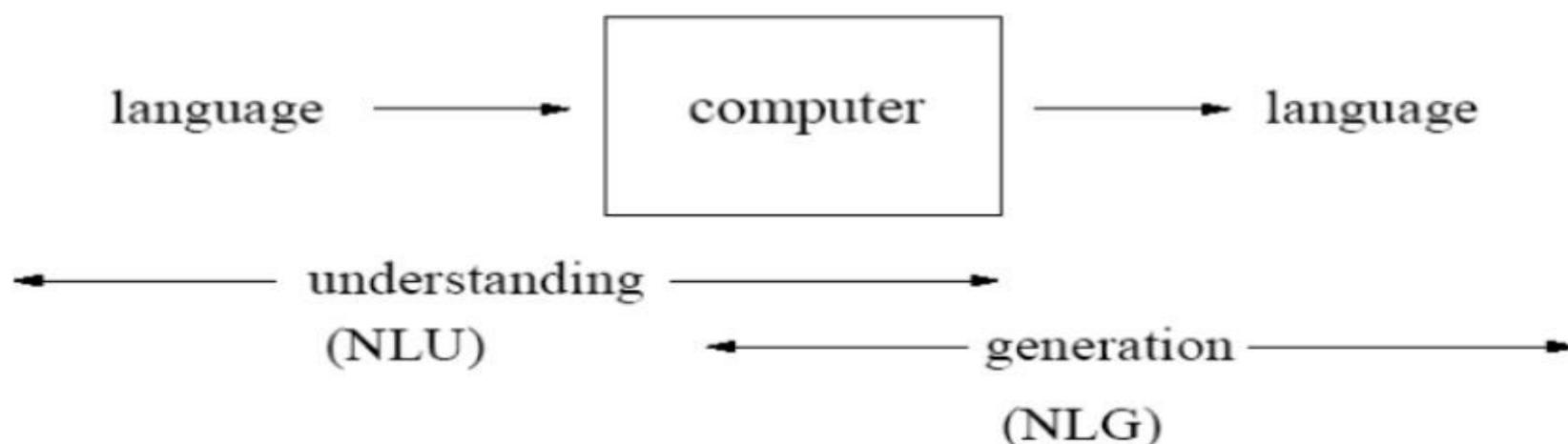
~80s

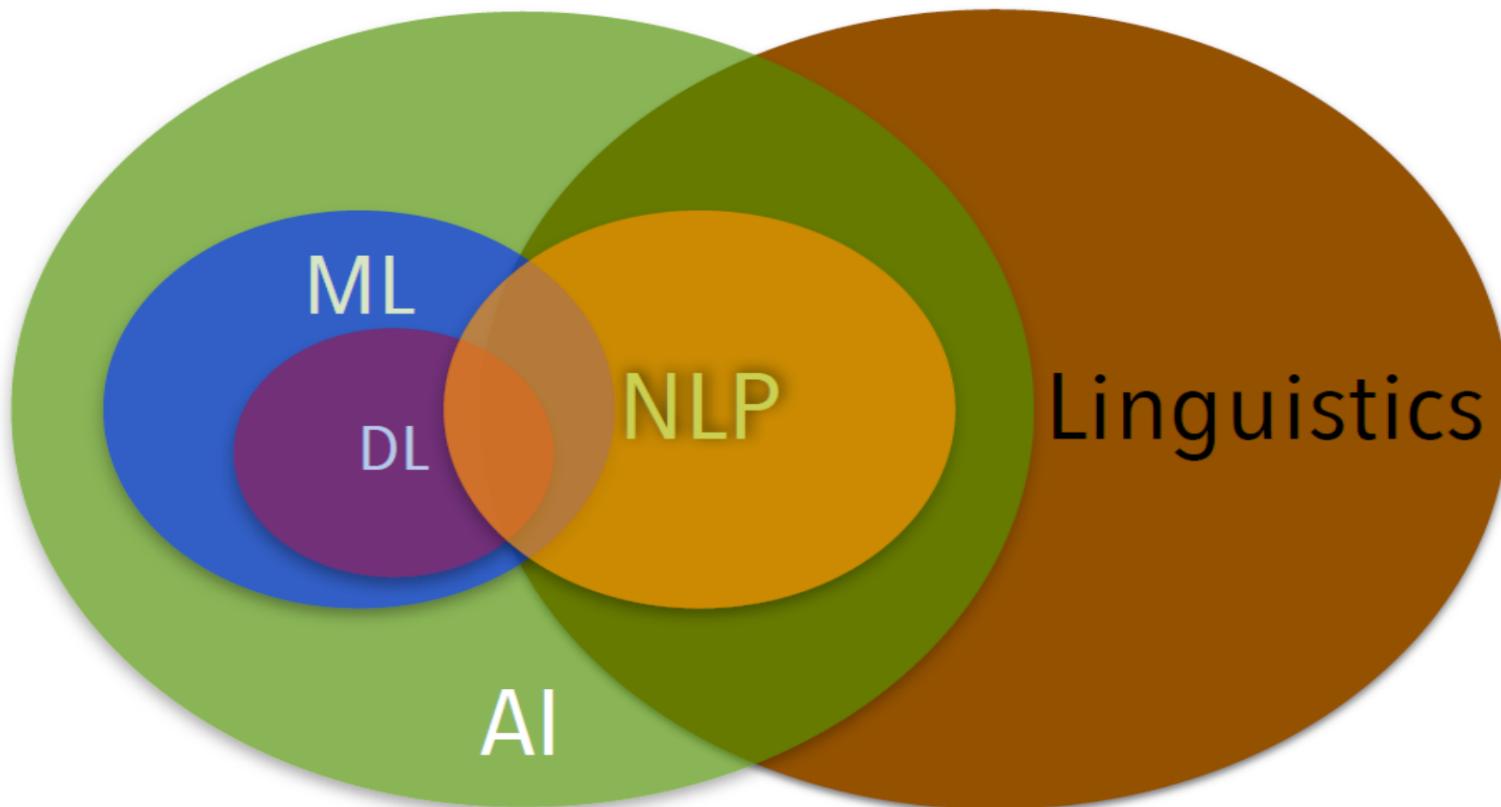


today

# What is NLP?

- Natural Language Processing (NLP) is a field in Artificial Intelligence (AI) devoted to creating computers that use natural language as input and/or output.

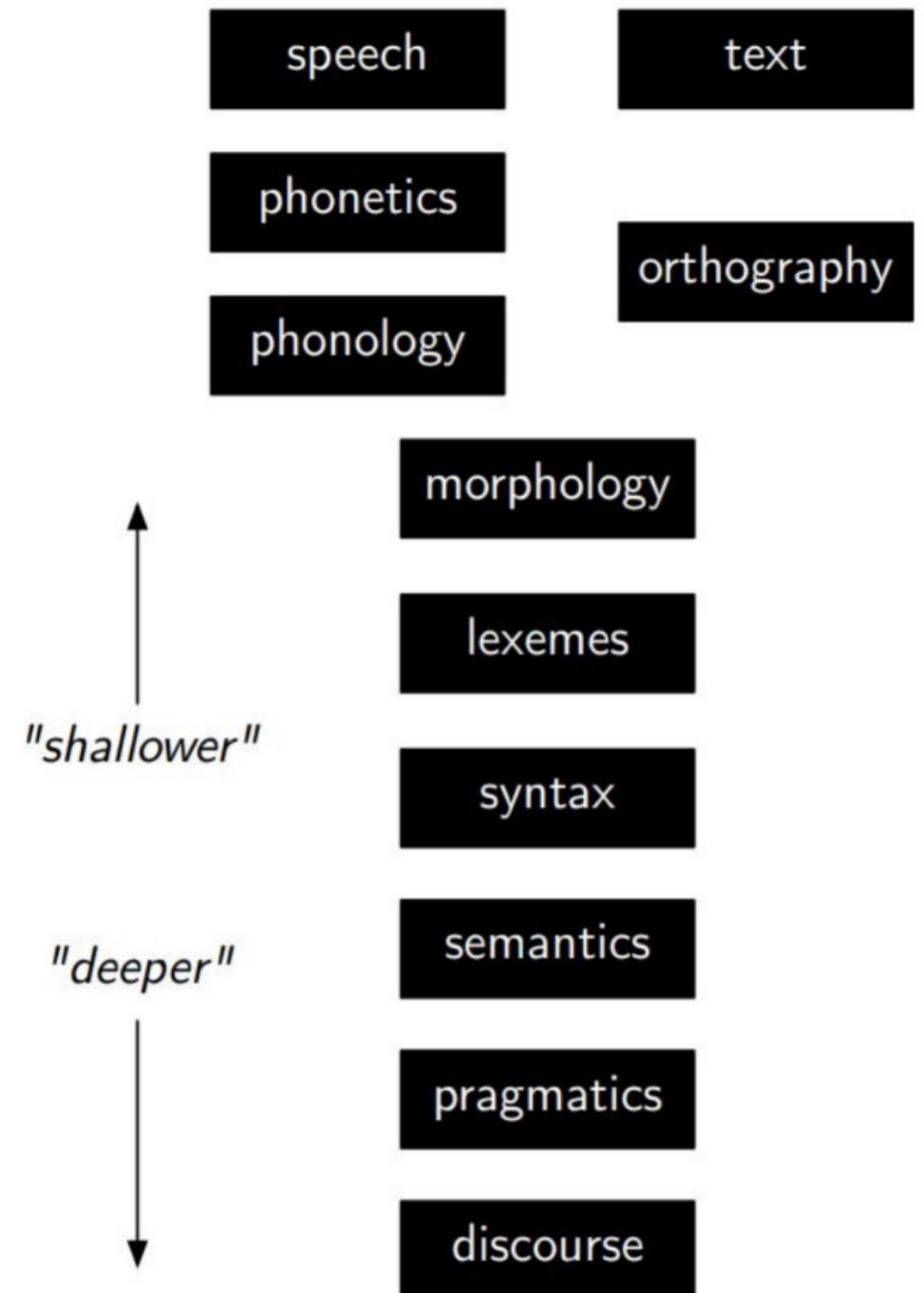




# Components of NLP

- **Natural Language Understanding**
  - Mapping the given input in the natural language into a useful representation.
  - Different level of analysis required:  
*morphological analysis,*  
*syntactic analysis,*  
*semantic analysis,*  
*discourse analysis*, ...
- **Natural Language Generation**
  - Producing output in the natural language from some internal representation.
  - Different level of synthesis required:  
*deep planning* (what to say),  
*syntactic generation*
- NL Understanding is much harder than NL Generation.  
But, still both of them are hard.

# Level Of Linguistic Knowledge



# Why NLP?

- To interact with computing devices using human (natural) languages. For example,
  - Building intelligent robots (AI).
  - Enabling voice-controlled operation.
- To access (large amount of) information and knowledge stored in the form of human languages quickly.

## Test for Intelligence – Turing Test

- Alan Turing (1950) proposed a test of a machine's capability to perform human-like conversation.
- A human judge engages in a natural language conversation with two other parties, one a human and the other a machine; if the judge cannot reliably tell which is which, then the machine is said to pass the test.



# Early Conversational Programs

- ELIZA (by Joseph Weizenbaum), 1966
  - A psychotherapist
  - No real understanding; simple pattern-matching to respond to user input -> canned responses

```
(my ?x depresses me) (why does your ?x depress you) 5  
(life ?x) (why do you say it ?x) 3  
(I could ?x) (you could ?x) 2  
(because ?x) (that is a good reason) 3  
(?x) (tell me more) 0
```

x - □

ELIZA

ELIZA: How do you do? Please tell me your problem.

YOU: I am unhappy.

ELIZA: How do you feel about being unhappy?

YOU: I need some help, that much seems certain.

ELIZA: Would it really help you to get some help that much seems certain?

YOU: Perhaps I could learn to get along with my mother.

ELIZA: Good family relations are important.

YOU: My mother takes care of me.

ELIZA: Why do you say that your mother takes care of you?

# **Modern NLP: NLP in the Days of Big Data**

Three trends:

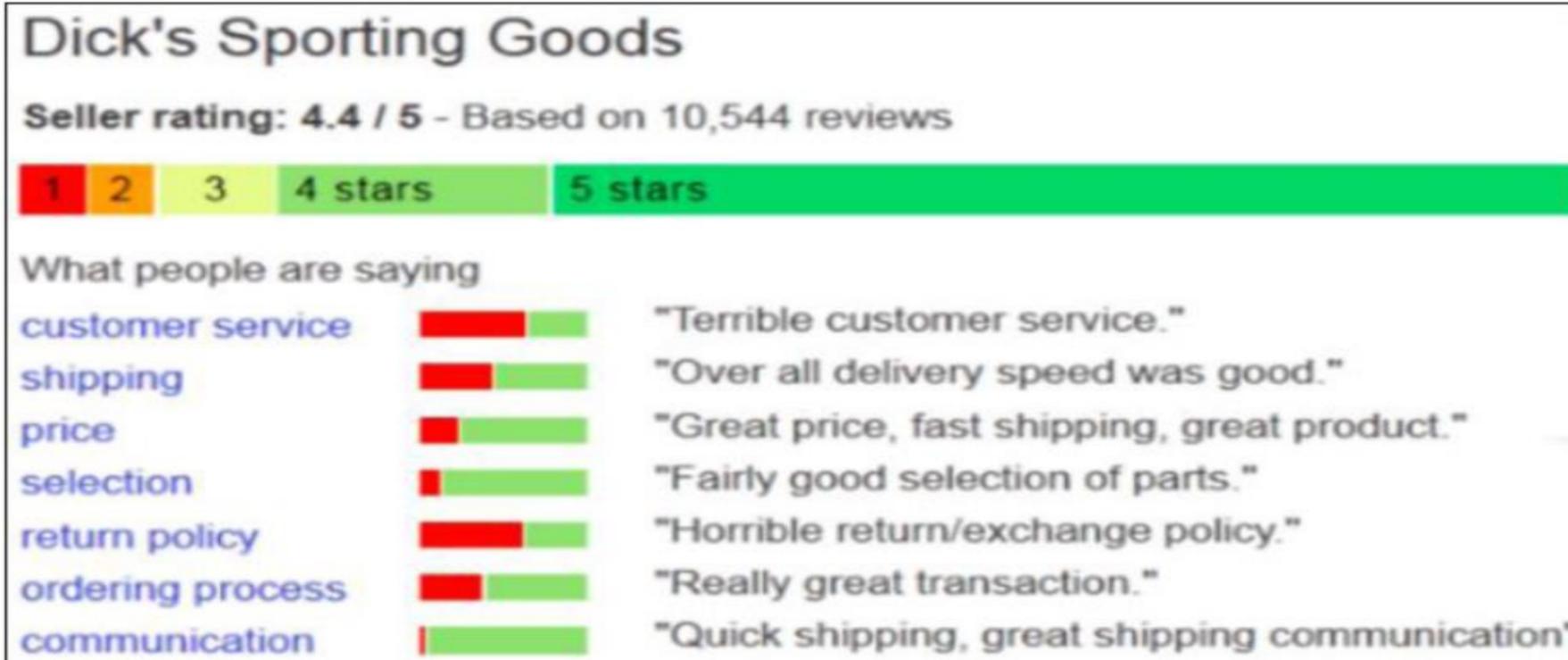
1. An **enormous amount of information** is now available in machine readable form as natural language text (newspapers, web pages, medical records, financial filings, product reviews, discussion forums, etc.)
2. Conversational agents are becoming an important form of human-computer **communication**
3. Much of human-human interaction is now mediated by computers via **social media**

# NLP Applications

- Three prominent application areas:
  - Text analytics/mining (from “***unstructured data***”)
    - Sentiment analysis
    - Topic identification
    - Digital Humanities (“*new ways of doing scholarship that involve collaborative, transdisciplinary, and computationally engaged research, teaching, and publishing.*”)
  - Conversational agents
    - Siri, Cortana, Amazon Alexa, Google Assistant
    - Chatbots
  - Machine translation

# Text Analytics

- Data-mining of weblogs, microblogs, discussion forums, user reviews, and other forms of user-generated media.



## Text Analytics (cont.)

- Typically this involves the extraction of **limited** kinds of semantic and pragmatic information from texts
  - Entity mentions
  - Concept identification
  - Sentiment

API TEST TOOL

English Entities Graphical

I really enjoyed using the Canon EOS 80D in Madrid on March 4. The Panasonic Lumix DMC-GH4 is a bit disappointing, but the Canon EOS M camera is not bad at all. All I want when taking photos is point it and then just press the button. For only 200 dollars \$, a really fair price, this camera is perfect for me. Besides, I have had a good customer service experience. John Faraday was very nice!

LEGEND color key

Type of entities:

Icon	Entity Type
Person name	Person name
Car license plate	Date
Place	Hour
Phone number	Money
Email address	Address
Twitter hashtag	Twitter hashtag

ENTITIES

# Demo

- Sentiment Analysis with Python NLTK Text Classification

<http://text-processing.com/demo/sentiment/>

- Tweet Sentiment Visualization Tool

[https://www.csc2.ncsu.edu/faculty/healey/tweet\\_viz/tweet\\_app/](https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/)

- Concept Extraction

<https://www.ibm.com/demos/live/natural-language-understanding/self-service/home>

# Conversational Agents

- **Combine**
  - Speech recognition/synthesis
  - Question answering
    - From the web and from structured information sources (freebase, dbpedia, yago, etc.)
  - Simple agent-like abilities
    - Create/edit calendar entries
    - Reminders
    - Directions
    - Invoking/interacting with other apps

# Question Answering

- Traditional *information retrieval* provides documents/resources that provide users with what they need to satisfy their information needs.
- *Question answering* on the other hand directly provides an answer to information needs posed as questions.

# IBM Watson



# **Machine Translation (MT)**

- The automatic translation of texts between languages is one of the oldest non-numerical applications in Computer Science.
- In the past 15 years or so, MT has gone from a niche academic curiosity to a robust commercial industry.

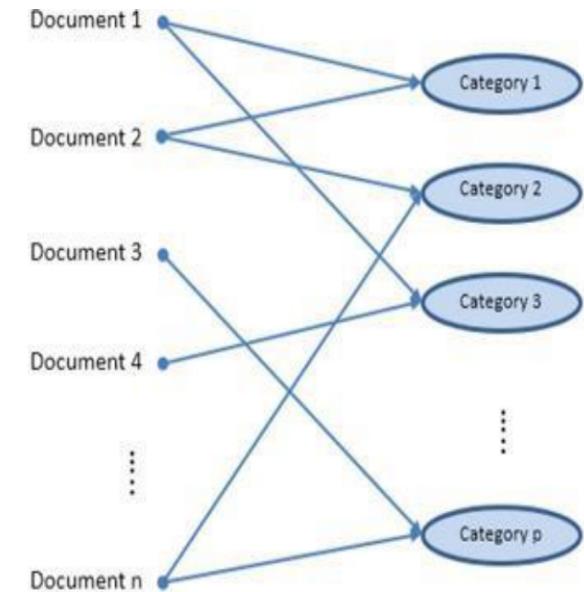
# NLP Application

- Supervised:
  - Spam Detection
  - Sentiment Analysis
  - Intent Classification
  - Multi-Label, Multi-Class Text Classification
- Unsupervised:
  - Topic Modeling
  - Keyword Extraction
  - Trend/Outlier detection

# Text Mining Applications –Supervised

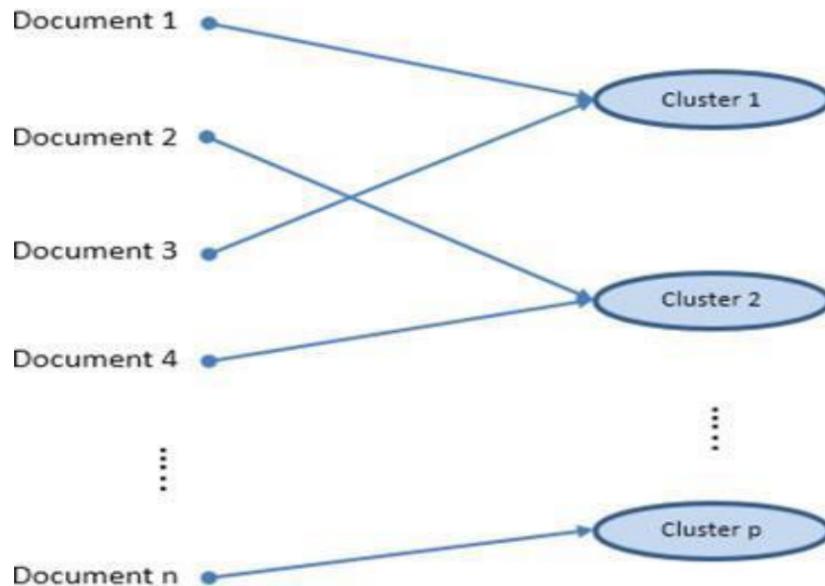
—Many typical predictive modeling or classification applications can be enhanced by incorporating textual data in addition to traditional input variables.

- churning propensity models that include customer center notes, website forms, e-mails, and Twitter messages
- hospital admission prediction models incorporating medical records notes as a new source of information
- insurance fraud modeling using adjustor notes
- sentiment categorization
- stylometry or forensic applications that identify the author of a particular writing sample



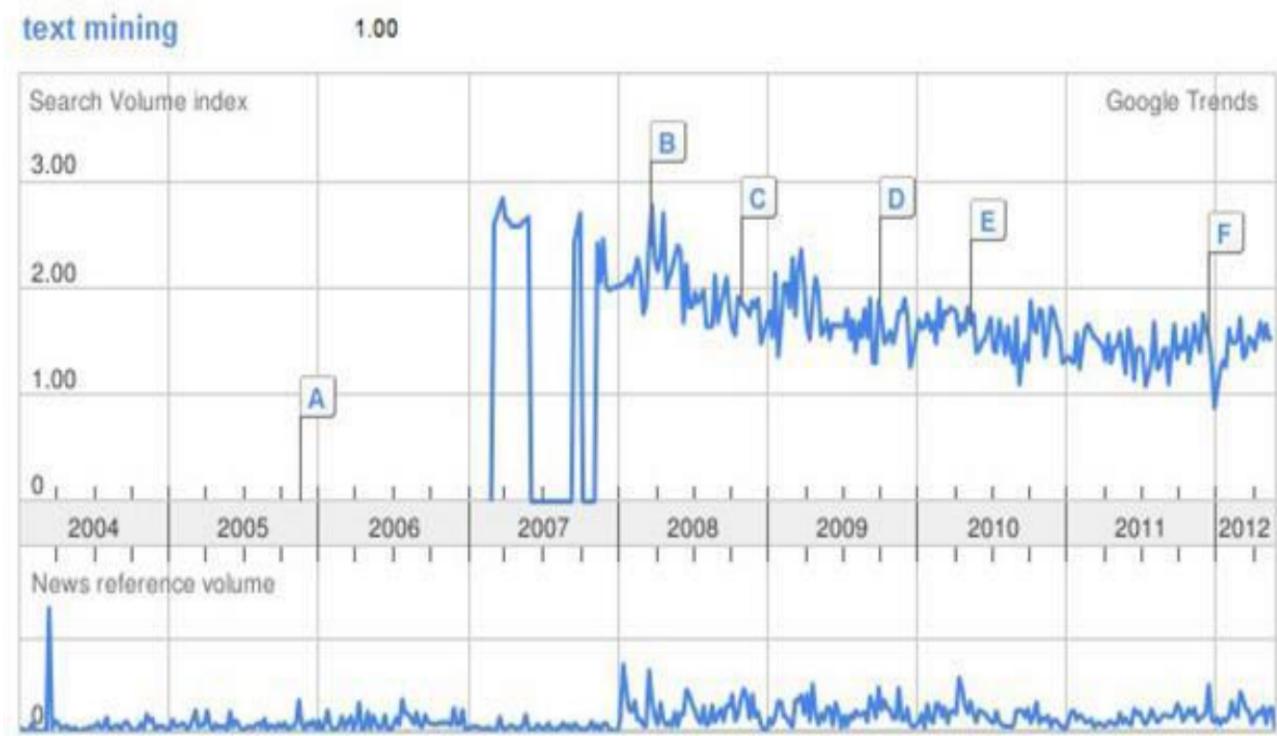
# Text Mining Applications –Unsupervised

- Text clustering



Cluster No.	Comment	Key Words
1	1, 3, 4	doctor, staff, friendly, helpful
2	5, 6, 8	treatment, results, time, schedule
3	2, 7	service, clinic, fast

- Trend analysis



Trend for the Term “text mining” from Google Trends

