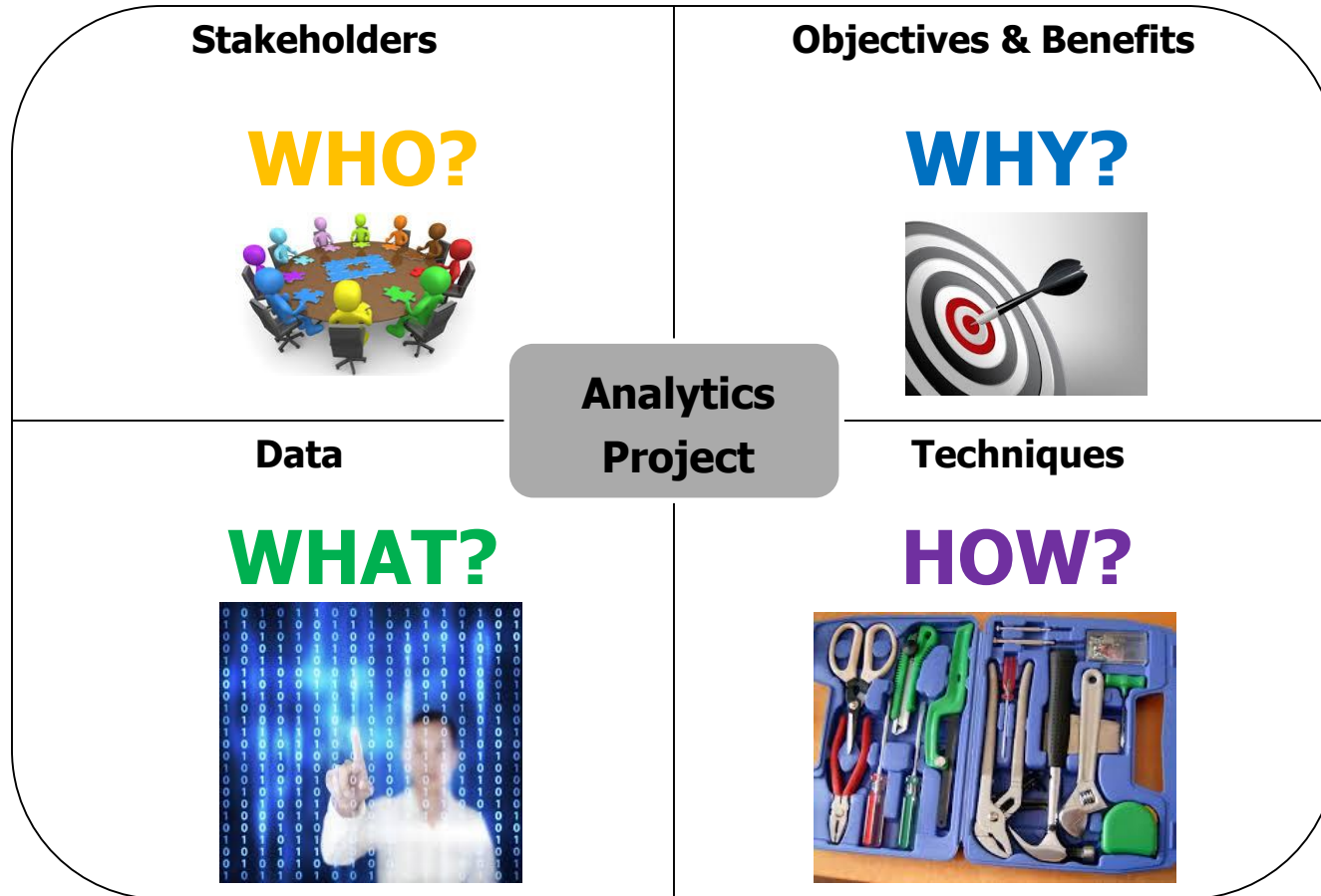


Text Analytics in Education

Text Analytics Framework for Education



Group Activity (5 min)

- List some examples of unstructured data that are related to education

Solution: Unstructured Data

Curriculum

Courses; Learning Outcomes

Course

Course Synopsis; Course Design Documents; Competencies; Course Content; Labs; Exercises; Questions; Course Feedback

Student

Application Essays; Project Reports; Assignment Essays; Student Blogs; Exam Answers; Discussion Forum Posts; CVs; Job/Internship Applications

Others

Surveys; Job advertisements; LinkedIn profiles; Social media posts; Social media profiles; Library reading behavior

Group Activity (5 min)

- List some examples of objectives or applications using these data

Solution: Text Analytics Applications in Education

**Curriculum
Effectiveness**

**Course
Competency**

**Course
Effectiveness**

**Student
Admissions**

**Student
Performance**

Assessment

Skills Gaps

**Student
Participation**

**Library
Resource
Analytics**

Solution: Objectives for Curriculum Effectiveness

Determine whether the outcomes have been achieved

Identify the gaps between the outcomes and assessments

Verify alignment with industry skills requirements

Study the impact on job opportunities

Study alignment of curriculum against learning taxonomies

Study alignment of content with learning activity

Group Activity (5 min)

- List some of the stakeholder who can benefits from text analytics applications in education

Solution: Stakeholders

Librarian

Lecturers

Government

**Deans/
Associate Deans**

Students

Employers

Alumni

**Professional
Bodies**

Counsellors

Text Analytics Challenges

- Data Privacy
 - Personal Data Protection Act
- Availability of Data
 - Not available in electronic format
 - In different systems and difficult to extract
 - Different formats
 - Lack of metadata
 - Missing and incorrect data
- Moral Obligation
 - Is it fair to predict student performance in current course based on past performance?
- Technical Limitation
 - Ineffectiveness of current algorithms

Technical Background

Tool 1 - Natural language processing

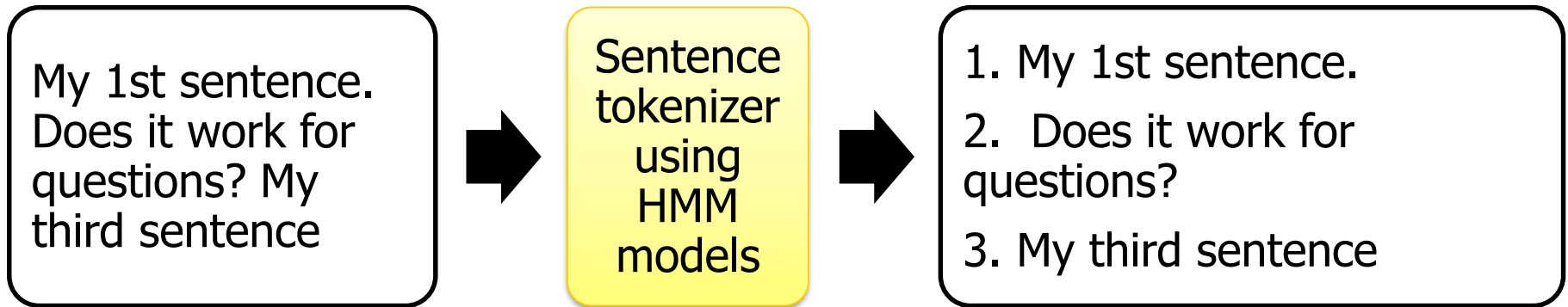
- Natural language processing (NLP) is a field of **computer science and linguistics** concerned with the **interactions** between computers and human (natural) languages
- Why Natural Language Processing ?
 - **Huge** amounts of data
 - **Internet** = at least 20 billions pages
 - Intranet Applications for processing large amounts of **texts** require NLP expertise
- Natural language understanding systems **convert samples of human language into more formal representations such as parse trees**

Tool 1 (Cont..) - Parts of Speech Tagger



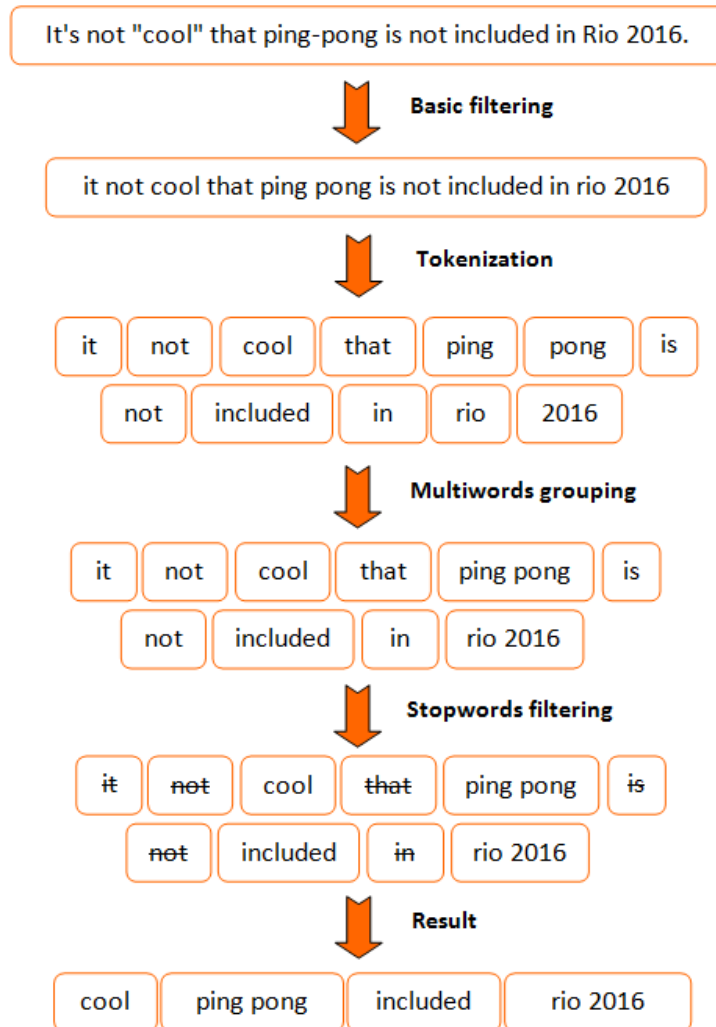
NP	Noun Phrase
NN	Noun, singular or mass
IN	Preposition or subordinating conjunction
VBD	Verb, past tense
DT	Determiner
JJ	Adjective

Tool 2 - Sentence Tokenizer



Tool 3 - Word Tokenizer

- Word tokenizer divides a text into a list of words.



Tool 4 - Key Phrase Extraction

- Useful for document categorization, clustering, indexing, search, and summarization;
- Clustering or Graph based algorithms

Keyword Extraction Result

Original URL/Text

Curriculum analysis unpacks the three education components namely the intended outcomes, content, and the learning activities with a purpose to evaluate how the individual components fit together in terms of consistency and alignment. It allows educators to analyse the strengths and weaknesses of the curricula in terms of all the three components. Furthermore, the analysis aims to justify the curriculum choices and assumptions, and it can be performed at the program level or at a specific course level.

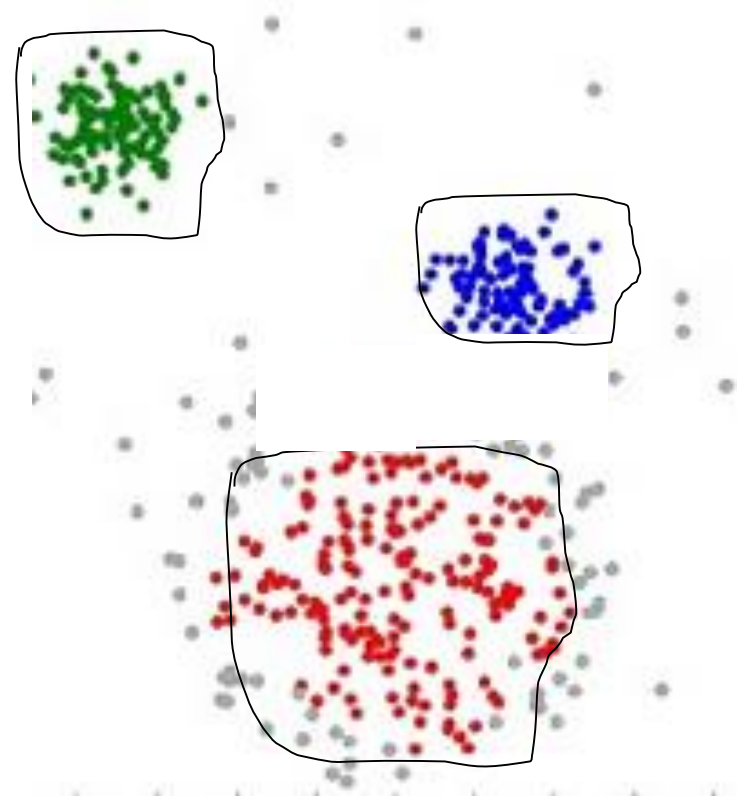
Extracted Keywords

specific course level
program level
analysis aims
education components
individual components
learning activities
analysis unpacks
curriculum choices

Tool 5 – Clustering Algorithm

- Technique of grouping data points into more meaningful sets with some identifying traits.

Store	Sales Turnover	Store Size	Staff Size	Profit Margin
New York	5.1	3.5	1.4	0.2
Los Angeles	4.9	3	1.4	0.2
Chicago	4.7	3.2	1.3	0.2
Houston	4.6	3.1	1.5	0.2
Philadelphia	5	3.6	1.4	0.2
Phoenix	5.4	3.9	1.7	0.4
San Antonio	4.6	3.4	1.4	0.3
San Diego	5	3.4	1.5	0.2
Dallas	4.4	2.9	1.4	0.2
San Jose	4.9	3.1	1.5	0.1
Jacksonville	5.4	3.7	1.5	0.2
Indianapolis	4.8	3.4	1.6	0.2
San Francisco	4.8	3	1.4	0.1
Austin	4.3	3	1.1	0.1
Columbus	5.8	4	1.2	0.2
Fort Worth	5.7	4.4	1.5	0.4
Charlotte	5.4	3.9	1.3	0.4
Detroit	5.1	3.5	1.4	0.3



Tool 5 – Use of Clustering in Topic Extraction

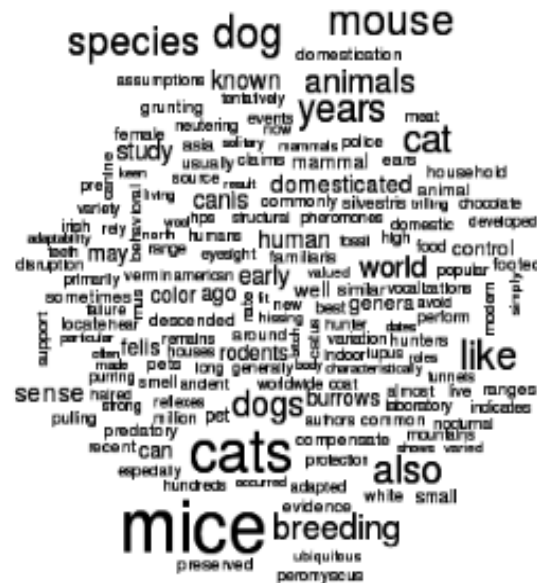
Forum discussions
/posts about the
pets



Topic extraction
models
(clustering
algorithms like K-
means clustering or
LDA models)

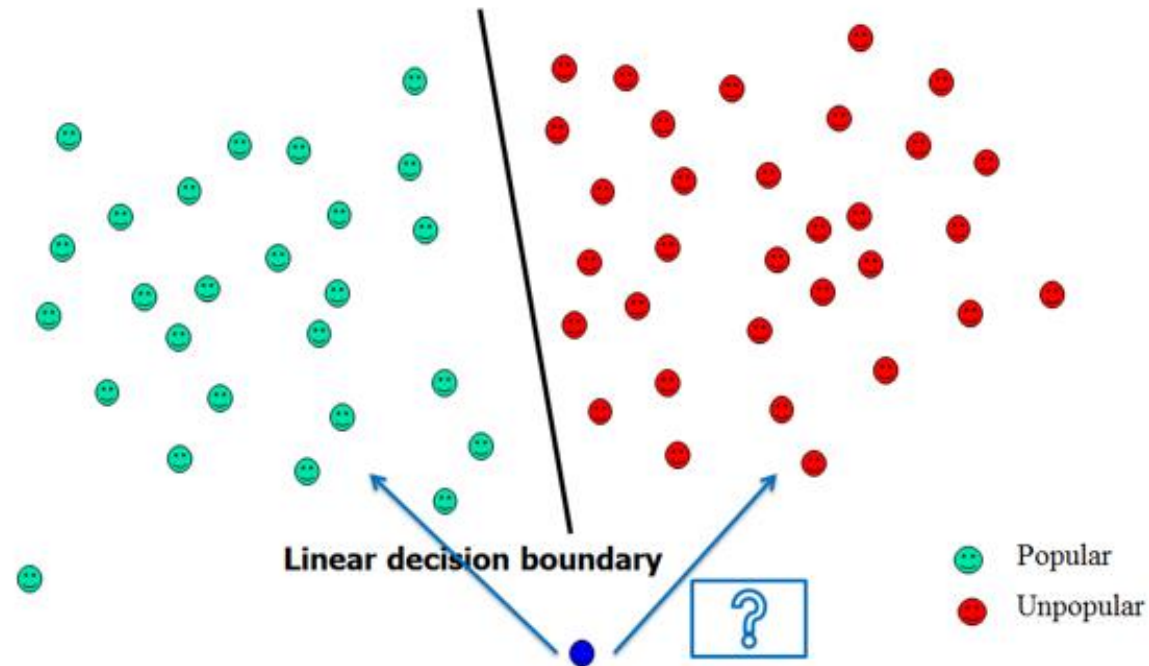


Clusters of words



Tool 6 - Classification Algorithm

- Classification is the problem of identifying to which category (cluster) a new observation belongs, on the basis of a training set of data containing observations whose category membership is known.



Tool 7 - Sentiment Classification

