

Lab sheet 11 – Text Summarization

Kavishka Gamage – 17000475

Key Phrase Extraction

Collocations N-grams (Bigram and Trigram)

```
Bigram :
[('said alice', 123), ('mock turtle', 56), ('march hare', 31), ('said king', 29), ('thought alice', 26), ('white rabbit', 22), ('said hatter', 22), ('said mock', 20), ('said caterpillar', 18), ('said gryphon', 18)]
Trigram :
[('said mock turtle', 20), ('said march hare', 10), ('poor little thing', 6), ('little golden key', 5), ('certainly said alice', 5), ('white kid gloves', 5), ('march hare said', 5), ('mock turtle said', 5), ('know said alice', 4), ('might well say', 4)]
```

NLTK built in collocation with raw frequency and PMI (Pointwise Mutual Information)

```
NLTK collocation using raw frequencies - Bigram
[('said', 'alice'), ('mock', 'turtle'), ('march', 'hare'), ('said', 'king'), ('thought', 'alice'), ('said', 'hatter'), ('white', 'rabbit'), ('said', 'mock'), ('said', 'caterpillar'), ('said', 'gryphon')]
NLTK collocation using PMI - Bigram
[('abide', 'figures'), ('acceptance', 'elegant'), ('accounting', 'tastes'), ('accustomed', 'usurpation'), ('act', 'crawling'), ('adjourn', 'immediate'), ('adoption', 'energetic'), ('affair', 'trusts'), ('agony', 'terror'), ('alarmed', 'proposal')]
NLTK collocation using PMI - Bigram
[('said', 'mock', 'turtle'), ('said', 'march', 'hare'), ('poor', 'little', 'thing'), ('little', 'golden', 'key'), ('march', 'hare', 'said'), ('mock', 'turtle', 'said'), ('white', 'kid', 'gloves'), ('beau', 'ootiful', 'soo'), ('certainly', 'said', 'alice'), ('might', 'well', 'say')]
NLTK collocation using PMI - trigram
[('accustomed', 'usurpation', 'conquest'), ('adjourn', 'immediate', 'adoption'), ('adoption', 'energetic', 'remedies'), ('ancient', 'modern', 'seaography'), ('apple', 'roast', 'turkey'), ('arithmetic', 'ambition', 'distraction'), ('brother', 'latin', 'grammar'), ('canvas', 'bag', 'tied'), ('cherry', 'tart', 'custard'), ('circle', 'exact', 'shape')]
```

Weighted tag based phrase extraction

Toy corpus

```
Chunks
[['elephants', 'large mammals', 'family elephantidae', 'order proboscidea'], ['species', 'african elephant', 'asian elephant'], ['elephants', 'sub-saharan africa', 'south asia', 'southeast asia'], ['male african elephants', 'extant terrestrial animals'], ['elephants', 'long trunk', 'many purposes', 'breathing', 'water', 'grasping objects'], ['incisors', 'tusks', 'weapons', 'tools', 'objects', 'digging'], ['elephants', 'large ear flaps', 'body temperature'], ['pillar-like legs', 'great weight'], ['african elephants', 'ears', 'backs', 'asian elephants', 'ears', 'convex', 'level backs']]
Key Phrases for Toy Dataset
[('extant terrestrial animals', 0.707), ('male african elephants', 0.707), ('great weight', 0.707), ('pillar-like legs', 0.707), ('body temperature', 0.684), ('large ear flaps', 0.684), ('ears', 0.667), ('african elephant', 0.577), ('asian elephant', 0.577), ('species', 0.577)]
```

Alice corpus

```
key phrases for Alice corpora
[('chapter', 1.0), ('rabbit', 1.0), ('thought alice', 1.0), ('sides', 1.0), ('ignorant little girl', 1.0), ('dinah', 1.0), ('thump', 1.0), ('whiskers', 1.0), ('***', 1.0), ('***', 1.0), ('*****', 1.0)]
```

Topic Modeling

For Toy corpus

LSI Gensim model - TFIDF

```
LSI Genseim model
Topic #1 with weights
[('programming', 0.46), ('language', 0.46), ('python', 0.34), ('java', 0.34), ('popular', 0.34)]

Topic #2 with weights
[('dog', -0.46), ('fox', -0.46), ('jump', -0.44), ('smarter', -0.32), ('cat', -0.32)]
```

LSI custom built model

```
LSI custome Built in model
Topic #1 without weights
['dog', 'fox', 'jump', 'smarter', 'cat', 'slow', 'lazy', 'quick', 'clever']

Topic #2 without weights
['programming', 'language', 'python', 'java', 'popular', 'ruby', 'excellent', 'program']
```

LDA genism model

```
LDA Gensim model
Topic #1 with weights
[('popular', 0.07), ('python', 0.07), ('language', 0.06), ('programming', 0.06), ('excellent', 0.06)]

Topic #2 with weights
[('fox', 0.07), ('programming', 0.07), ('language', 0.07), ('dog', 0.07), ('jump', 0.06)]

Topic #1 with weights
[('fox', 1.06), ('dog', 1.06), ('jump', 1.10), ('clever', 1.12), ('quick', 1.12), ('lazy', 1.12), ('slow', 1.12), ('cat', 1.06)]

Topic #2 with weights
[('programming', 1.8), ('language', 1.8), ('java', 1.04), ('python', 1.04), ('program', 1.3), ('ruby', 1.11), ('excellent', 1.11), ('popular', 1.06)]

Topic #1 with weights
[('programming', 0.55), ('language', 0.55), ('python', 0.4), ('java', 0.4), ('popular', 0.24), ('ruby', 0.23), ('excellent', 0.23), ('program', 0.09), ('small', 0.0)]

Topic #2 with weights
[('dog', 0.57), ('fox', 0.57), ('jump', 0.35), ('smarter', 0.26), ('cat', 0.26), ('quick', 0.13), ('slow', 0.13), ('clever', 0.13), ('lazy', 0.13)]

I base the value of a game on the amount of enjoyable gameplay I can get out of it and this one was definitely worth the price!
Topic #1 without weights
['skyrim', 'quest', 'one', 'like', 'play', 'oblivion', 'go', 'get', 'time', 'good']

Topic #2 without weights
['recommend', 'love', 'even', 'best', 'great', 'highly', 'level', 'play', 'elder', 'scroll']
```

For Amazon skyrim reviews

LDA genism model

```
Amazon skyrim reviews - Gensim LDA model
Topic #1 without weights
['play', 'dragon', 'fun', 'one', 'good', 'like', 'explore', 'love', 'go', 'really']

Topic #2 without weights
['skyrim', 'one', 'much', 'character', 'graphic', 'good', 'hour', 'play', 'hundred', 'level']

Topic #3 without weights
['love', 'buy', 'one', 'time', 'great', 'much', 'really', 'best', 'far', 'skyrim']

Topic #4 without weights
['love', 'great', 'buy', 'wonderful', 'good', 'elder', 'scroll', 'play', 'quest', 'like']

Topic #5 without weights
['fun', 'get', 'quest', 'play', 'hour', 'like', 'skyrim', 'go', 'oblivion', '5']
```

Sklearn LDA

```
Amazon skyrim reviews - Sklearn LDA
Topic #1 without weights
['estatic', 'booklet', 'wonder4ful', 'electricity', 'stays', 'heat', 'trhats', 'amazingly', 'interfere', '12yr']

Topic #2 without weights
['estatic', 'booklet', 'wonder4ful', 'electricity', 'stays', 'heat', 'trhats', 'amazingly', 'interfere', '12yr']

Topic #3 without weights
['nuff', 'pointy', 'castles', 'beards', 'mead', 'wolfhounds', 'axes', 'irish', 'helmets', 'waist']

Topic #4 without weights
['game', 'play', 'get', 'one', 'skyrim', 'great', 'like', 'time', 'quest', 'much']

Topic #5 without weights
['de', 'compras', 'pagar', 'futuras', 'skyrimseguridad', 'recomiendo', 'responsabilidad', 'crédito', 'momento', 'tarjeta']
```

Sklearn Non Negative Matrix Factorization

```
Amazon skyrim reviews - sklearn Non Negative Matrix Factorization
Topic #1 without weights
['game', 'get', 'skyrim', 'play', 'time', 'quest', 'like', 'one', 'go', 'much']

Topic #2 without weights
['ever', 'best', 'game', 'play', 'rpg', 'one', 'hour', 'great', 'definitely', 'decade']

Topic #3 without weights
['scroll', 'elder', 'series', 'always', 'love', 'pass', 'buy', 'far', 'franchise', 'dont']

Topic #4 without weights
['game', 'recommend', 'love', 'great', 'highly', 'play', 'wonderful', 'like', 'would', 'graphic']

Topic #5 without weights
['fun', 'game', 'much', 'graphic', 'improvement', 'expect', 'mission', 'see', 'hour', 'couple']
```

Document Summarization

Gensim summarization model has removed from genism. So its result not included here.

Custom summarization algorithm result

All elephants have a long trunk used for many purposes, particularly breathing, lifting water and grasping objects. Their incisors grow into tusks, which can serve as weapons and as tools for moving objects and digging. African elephants have larger ears and concave backs while Asian elephants have smaller ears and convex or level backs.

Pagerank - networkX

Elephants are large mammals of the family Elephantidae and the order Proboscidea. Male African elephants are the largest extant terrestrial animals. African elephants have larger ears and concave backs while Asian elephants have smaller ears and convex or level backs.

Generic LSA based text summarizer result

LSA
The Elder Scrolls V: Skyrim is an open world action role-playing video game developed by Bethesda Game Studios and published by Bethesda Softworks. Players may navigate the game world more quickly by riding horses, or by utilizing a fast-travel system which allows them to warp to previously discovered locations. Players have the option to develop their character. At the beginning of the game, players create their character by selecting one of several races, including humans, orcs, elves and anthropomorphic cat or lizard-like creatures, and then customizing their character's appearance.

Generic TextRank based text summarizer

The Elder Scrolls V: Skyrim is an open world action role-playing video game developed by Bethesda Game Studios and published by Bethesda Softworks. Players may navigate the game world more quickly by riding horses, or by utilizing a fast-travel system which allows them to warp to previously discovered locations. Players have the option to develop their character. Skyrim is the first entry in The Elder Scrolls to include Dragons in the game's wilderness.