

CSDL7013 Natural Language Processing Lab 9

- NOTE: Prepend your Roll Number to the name of this file
- YET ANOTHER NOTE: Take print out of the relevant code cells only.

Word Sense Disambiguation Lab

1. **Title:** Word Sense Disambiguation
2. **Objective/Aim:** To illustrate Word Sense Disambiguation in a given Document
3. **Due Date:** Friday October 07, 2022

Name : Harshvardhan Singh

Roll Number : 1019161

Setup Wordnet interaction

```
import nltk
nltk.download('wordnet')
nltk.download('omw-1.4')
from nltk.corpus import wordnet as wn
from nltk.wsd import lesk
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd
from sklearn.metrics.pairwise import cosine_similarity

[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Package wordnet is already up-to-date!
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...
[nltk_data] Package omw-1.4 is already up-to-date!
```

Play around with glosses

```
for ss in wn.synsets('liturgy'):
    print(ss, ss.definition())
```

```
Synset('holy_eucharist.n.01') a Christian sacrament commemorating the
Last Supper by consecrating bread and wine
Synset('liturgy.n.02') a rite or body of rites prescribed for public
worship
```

```
[(s, s.pos(),s.definition()) for s in wn.synsets('parliament')]

[(Synset('parliament.n.01'),
 'n',
 'a legislative assembly in certain countries'),
 (Synset('fantan.n.02'),
 'n',
 'a card game in which you play your sevens and other cards in
```

sequence in the same suit as the sevens; you win if you are the first to use all your cards']]

Check Lesk Algorithm

Set up sentence

```
message = "Your mission, should you choose to accept, is to complete  
this assignment and burn it."  
split_message = message.split()
```

Check the gloss of the word under consideration

```
[(s, s.pos(),s.definition()) for s in wn.synsets('burn')]  
  
[(Synset('burn.n.01'), 'n', 'pain that feels hot as if it were on  
fire'),  
 (Synset('tan.n.01'),  
  'n',  
  'a browning of the skin resulting from exposure to the rays of the  
sun'),  
 (Synset('burn.n.03'),  
  'n',  
  'an injury caused by exposure to heat or chemicals or radiation'),  
 (Synset('burn.n.04'),  
  'n',  
  "a place or area that has been burned (especially on a person's  
body)"),  
 (Synset('burn.n.05'), 'n', 'damage inflicted by fire'),  
 (Synset('burn.v.01'), 'v', 'destroy by fire'),  
 (Synset('burn.v.02'), 'v', 'shine intensely, as if with heat'),  
 (Synset('burn.v.03'), 'v', 'undergo combustion'),  
 (Synset('bite.v.02'), 'v', 'cause a sharp or stinging pain or  
discomfort'),  
 (Synset('burn.v.05'), 'v', 'cause to burn or combust'),  
 (Synset('burn.v.06'),  
  'v',  
  'feel strong emotion, especially anger or passion'),  
 (Synset('burn.v.07'), 'v', 'cause to undergo combustion'),  
 (Synset('burn.v.08'), 'v', 'burn at the stake'),  
 (Synset('burn.v.09'), 'v', 'spend (significant amounts of money)'),  
 (Synset('burn.v.10'), 'v', 'feel hot or painful'),  
 (Synset('cauterize.v.01'),  
  'v',  
  'burn, sear, or freeze (tissue) using a hot iron or electric current  
or a caustic agent'),  
 (Synset('sunburn.v.01'), 'v', 'get a sunburn by overexposure to the  
sun'),  
 (Synset('cut.v.21'), 'v', 'create by duplicating data'),
```

```
(Synset('burn_off.v.01'), 'v', 'use up (energy)'),
(Synset('burn.v.15'), 'v', 'burn with heat, fire, or radiation')]
```

Perform WSD using inbuilt Lesk Algorithm

```
print(lesk(split_message, 'burn'))
```

```
Synset('burn.v.05')
```

Illustrate your own Lesk Algorithm

```
sent = "The moment he touched the cricket pitch, raised his bat to
face the ball, he went for a duck"
```

Create the Document Term Matrix

```
vectorizer = CountVectorizer(stop_words='english')
```

#vectorizer = TfidfVectorizer(stop_words='english') # Use any one.

```
doc1 = sent
```

```
for ss in wn.synsets('duck'):
```

```
    doc2 = ss.definition()
```

```
    print(ss, doc2)
```

```
    documents = [doc1, doc2]
```

```
    sparse_matrix = vectorizer.fit_transform(documents)
```

```
    doc_term_matrix = sparse_matrix.todense()
```

```
    df = pd.DataFrame(doc_term_matrix,
                      columns=vectorizer.get_feature_names_out(),
                      index=['Test Sentence', 'Definition'])
```

```
    print(df)
```

```
    print(cosine_similarity(df, df))
```

```
Synset('duck.n.01') small wild or domesticated web-footed broad-billed
swimming bird usually having a depressed body and short legs
```

```
ball bat billed bird body broad cricket
```

```
depressed \
```

```
Test Sentence      1      1      0      0      0      0      1
```

```
0
```

```
Definition         0      0      1      1      1      1      0
```

```
1
```

```
domesticated duck ... pitch raised short small
```

```
swimming \
```

```
Test Sentence      0      1 ...      1      1      0      0
```

```
0
```

```
Definition         1      0 ...      0      0      1      1
```

```
1
```

```
touched usually web went wild
```

```
Test Sentence      1      0      0      1      0
```

```
Definition         0      1      1      0      1
```

```
[2 rows x 25 columns]
```

[[1. 0.]
[0. 1.]]

Synset('duck.n.02') (cricket) a score of nothing by a batsman
ball bat batsman cricket duck face moment pitch
raised \
Test Sentence 1 1 0 1 1 1 1 1
1
Definition 0 0 1 1 0 0 0 0
0

score touched went
Test Sentence 0 1 1
Definition 1 0 0
[[1. 0.18257419]
[0.18257419 1.]]

Synset('duck.n.03') flesh of a duck (domestic or wild)
ball bat cricket domestic duck face flesh moment
pitch \
Test Sentence 1 1 1 0 1 1 0 1
1
Definition 0 0 0 1 1 0 1 0
0

raised touched went wild
Test Sentence 1 1 1 0
Definition 0 0 0 1
[[1. 0.15811388]
[0.15811388 1.]]

Synset('duck.n.04') a heavy cotton fabric of plain weave; used for
clothing and tents
ball bat clothing cotton cricket duck fabric
face \
Test Sentence 1 1 0 0 1 1 0
1
Definition 0 0 1 1 0 0 1
0

heavy moment pitch plain raised tents touched
used \
Test Sentence 0 1 1 0 1 0 1
0
Definition 1 0 0 1 0 1 0
1

weave went
Test Sentence 0 1
Definition 1 0
[[1. 0.]
[0. 1.]]

Synset('duck.v.01') to move (the head or body) quickly downwards or

away								
head \	away	ball	bat	body	cricket	downwards	duck	face
Test Sentence	0	1	1	0	1	0	1	1
0								
Definition	1	0	0	1	0	1	0	0
1								

	moment	pitch	quickly	raised	touched	went
Test Sentence	1	1	0	1	1	1
Definition	0	0	1	0	0	0

[[1. 0.]

[0. 1.]]

Synset('duck.v.02') submerge or plunge suddenly

	ball	bat	cricket	duck	face	moment	pitch	plunge
raised \								
Test Sentence	1	1	1	1	1	1	1	0
1								
Definition	0	0	0	0	0	0	0	1
0								

	submerge	suddenly	touched	went
Test Sentence	0	0	1	1
Definition	1	1	0	0

[[1. 0.]

[0. 1.]]

Synset('dip.v.10') dip into a liquid

	ball	bat	cricket	dip	duck	face	liquid	moment
pitch \								
Test Sentence	1	1	1	0	1	1	0	1
1								
Definition	0	0	0	1	0	0	1	0
0								

	raised	touched	went
Test Sentence	1	1	1
Definition	0	0	0

[[1. 0.]

[0. 1.]]

Synset('hedge.v.01') avoid or try to avoid fulfilling, answering, or performing (duties, questions, or issues)

	answering	avoid	ball	bat	cricket	duck	duties
face \							
Test Sentence	0	0	1	1	1	1	0
1							
Definition	1	2	0	0	0	0	1
0							

	fulfilling	issues	moment	performing	pitch
questions \					

Test Sentence	0	0	1	0	1
0					
Definition	1	1	0	1	0
1					

	raised	touched	try	went
Test Sentence	1	1	0	1
Definition	0	0	1	0

```
[[1. 0.]
 [0. 1.]]
```

Solve your exercises below this.

Q.1. [03 Marks] Check the different senses of the following words.

drab, lead, hook, bold, clean

#Solution to Question 1

```
for ss in wn.synsets('drab'):
    print(ss,ss.definition())
for ss in wn.synsets('lead'):
    print(ss,ss.definition())
for ss in wn.synsets('hook'):
    print(ss,ss.definition())
for ss in wn.synsets('bold'):
    print(ss,ss.definition())
for ss in wn.synsets('clean'):
    print(ss,ss.definition())
```

Synset('olive_drab.n.01') a dull greyish to yellowish or light olive brown

Synset('drab.s.01') lacking in liveliness or charm or surprise

Synset('drab.s.02') lacking brightness or color; dull

Synset('olive-drab.s.01') of a light brownish green color

Synset('blue.s.08') causing dejection

Synset('lead.n.01') an advantage held by a competitor in a race

Synset('lead.n.02') a soft heavy toxic malleable metallic element; bluish white when freshly cut but tarnishes readily to dull grey

Synset('lead.n.03') evidence pointing to a possible solution

Synset('lead.n.04') a position of leadership (especially in the phrase 'take the lead')

Synset('lead.n.05') the angle between the direction a gun is aimed and the position of a moving target (correcting for the flight time of the missile)

Synset('lead.n.06') the introductory section of a story

Synset('lead.n.07') (sports) the score by which a team or individual is winning

Synset('star.n.04') an actor who plays a principal role

Synset('lead.n.09') (baseball) the position taken by a base runner

preparing to advance to the next base
Synset('tip.n.03') an indication of potential opportunity
Synset('lead.n.11') a news story of major importance
Synset('spark_advance.n.01') the timing of ignition relative to the position of the piston in an internal-combustion engine
Synset('leash.n.01') restraint consisting of a rope (or light chain) used to restrain an animal
Synset('lead.n.14') thin strip of metal used to separate lines of type in printing
Synset('lead.n.15') mixture of graphite with clay in different degrees of hardness; the marking substance in a pencil
Synset('jumper_cable.n.01') a jumper that consists of a short piece of wire
Synset('lead.n.17') the playing of a card to start a trick in bridge
Synset('lead.v.01') take somebody somewhere
Synset('leave.v.07') have as a result or residue
Synset('lead.v.03') tend to or result in
Synset('lead.v.04') travel in front of; go in advance of others
Synset('lead.v.05') cause to undertake a certain action
Synset('run.v.03') stretch out over a distance, space, time, or scope; run or extend between two points or beyond a certain point
Synset('head.v.02') be in charge of
Synset('lead.v.08') be ahead of others; be the first
Synset('contribute.v.03') be conducive to
Synset('conduct.v.02') lead, as in the performance of a composition
Synset('go.v.25') lead, extend, or afford access
Synset('precede.v.04') move ahead (of others) in time or space
Synset('run.v.23') cause something to pass or lead somewhere
Synset('moderate.v.01') preside over
Synset('hook.n.01') a catch for locking a door
Synset('hook.n.02') a sharp curve or crook; a shape resembling a hook
Synset('bait.n.01') anything that serves as an enticement
Synset('hook.n.04') a mechanical device that is curved or bent to suspend or hold or pull something
Synset('hook.n.05') a curved or bent implement for suspending or pulling something
Synset('hook.n.06') a golf shot that curves to the left for a right-handed golfer
Synset('hook.n.07') a short swinging punch delivered from the side with the elbow bent
Synset('hook_shot.n.01') a basketball shot made over the head with the hand that is farther from the basket
Synset('hook.v.01') fasten with a hook
Synset('overcharge.v.01') rip off; ask an unreasonable price
Synset('crochet.v.02') make a piece of needlework by interlocking and looping thread with a hooked needle
Synset('hook.v.04') hit a ball and put a spin on it so that it travels to the left
Synset('hook.v.05') take by theft
Synset('pilfer.v.01') make off with belongings of others

Synset('hook.v.07') hit with a hook
Synset('hook.v.08') catch with a hook
Synset('addict.v.01') to cause (someone or oneself) to become dependent (on something, especially a narcotic drug)
Synset('hook.v.10') secure with the foot
Synset('hook.v.11') entice and trap
Synset('hook.v.12') approach with an offer of sexual favors
Synset('boldface.n.01') a typeface with thick heavy lines
Synset('bold.a.01') fearless and daring
Synset('bold.s.02') clear and distinct
Synset('bluff.s.01') very steep; having a prominent and almost vertical front
Synset('clean_and_jerk.n.01') a weightlift in which the barbell is lifted to shoulder height and then jerked overhead
Synset('clean.v.01') make clean by removing dirt, filth, or unwanted substances from
Synset('clean.v.02') remove unwanted substances from, such as feathers or pits
Synset('houseclean.v.01') clean and tidy up the house
Synset('cleanse.v.01') clean one's body or parts thereof, as by washing
Synset('clean.v.05') be cleanable
Synset('clean.v.06') deprive wholly of money in a gambling game, robbery, etc.
Synset('clean.v.07') remove all contents or possession from, or empty completely
Synset('clean.v.08') remove while making clean
Synset('scavenge.v.04') remove unwanted substances from
Synset('clean.v.10') remove shells or husks from
Synset('clean.a.01') free from dirt or impurities; or having clean habits
Synset('clean.s.02') free of restrictions or qualifications
Synset('clean.s.03') (of sound or color) free from anything that dulls or dims
Synset('clean.s.04') free from impurities
Synset('clean.s.05') (of a record) having no marks of discredit or offense
Synset('clean.a.06') ritually clean or pure
Synset('clean.a.07') not spreading pollution or contamination; especially radioactive contamination
Synset('clean.a.08') (of behavior or especially language) free from objectionable elements; fit for all observers
Synset('uninfected.s.01') free from sepsis or infection
Synset('clean.s.10') morally pure
Synset('clean.s.11') (of a manuscript) having few alterations or corrections
Synset('blank.s.01') (of a surface) not written or printed on
Synset('clean.s.13') exhibiting or calling for sportsmanship or fair play
Synset('clean.s.14') without difficulties or problems

Synset('clean.s.15') thorough and without qualification
Synset('clean.s.16') not carrying concealed weapons
Synset('clean.s.17') free from clumsiness; precisely or deftly executed
Synset('clean.s.18') free of drugs
Synset('clean.r.01') completely; used as intensifiers
Synset('fairly.r.03') in conformity with the rules or laws and without fraud or cheating

[03 Marks] Q. 2. Check the sense of the word face in the given sentence, using the in-built lesk algorithm.

Face the sun and you can't see your shadow.

#Solution to Question 2

```
message = "Face the sun and you can't see your shadow."  
split_message = message.split()  
print(lesk(split_message, 'burn'))
```

```
Synset('tan.n.01')
```

Q. 3. [04 Marks] Use cosine distance technique to check the sense of the word face, in the given sentence.

#Solution to Question 3

```
vectorizer = TfidfVectorizer(stop_words='english')  
doc1 = sent  
for ss in wn.synsets('face'):  
    doc2 = ss.definition()  
    print(ss, doc2)  
    documents = [doc1, doc2]  
    sparse_matrix = vectorizer.fit_transform(documents)  
    doc_term_matrix = sparse_matrix.todense()  
    df = pd.DataFrame(doc_term_matrix,  
                      columns=vectorizer.get_feature_names_out(),  
                      index=['Test Sentence', 'Definition'])
```

```
print(df)  
print(cosine_similarity(df, df))
```

Synset('face.n.01') the front of the human head from the forehead to the chin and ear to ear
Synset('expression.n.01') the feelings expressed on a person's face
Synset('face.n.03') the general outward appearance of something
Synset('face.n.04') the striking or working surface of an implement
Synset('face.n.05') a part of a person that is used to refer to a person
Synset('side.n.04') a surface forming part of the outside of an object
Synset('face.n.07') the part of an animal corresponding to the human face
Synset('face.n.08') the side upon which the use of a thing depends (usually the most prominent surface of an object)

Synset('grimace.n.01') a contorted facial expression
 Synset('font.n.01') a specific size and style of type within a type family
 Synset('face.n.11') status in the eyes of others
 Synset('boldness.n.02') impudent aggressiveness
 Synset('face.n.13') a vertical surface of a building or cliff
 Synset('confront.v.02') deal with (something unpleasant) head on
 Synset('confront.v.01') oppose, as in hostility or a competition
 Synset('front.v.01') be oriented in a certain direction, often with respect to another reference point; be opposite to
 Synset('face.v.04') be opposite
 Synset('face.v.05') turn so as to face; turn the face in a certain direction
 Synset('confront.v.03') present somebody with something, usually to accuse or criticize
 Synset('face.v.07') turn so as to expose the face
 Synset('face.v.08') line the edge (of a garment) with a different material
 Synset('face.v.09') cover the front or surface of

	ball	bat	cover	cricket	duck
face \					
Test Sentence	0.316228	0.316228	0.000000	0.316228	0.316228
0.316228					
Definition	0.000000	0.000000	0.707107	0.000000	0.000000
0.000000					

	moment	pitch	raised	surface	touched
went					
Test Sentence	0.316228	0.316228	0.316228	0.000000	0.316228
0.316228					
Definition	0.000000	0.000000	0.000000	0.707107	0.000000
0.000000					

[[1. 0.]
 [0. 1.]]