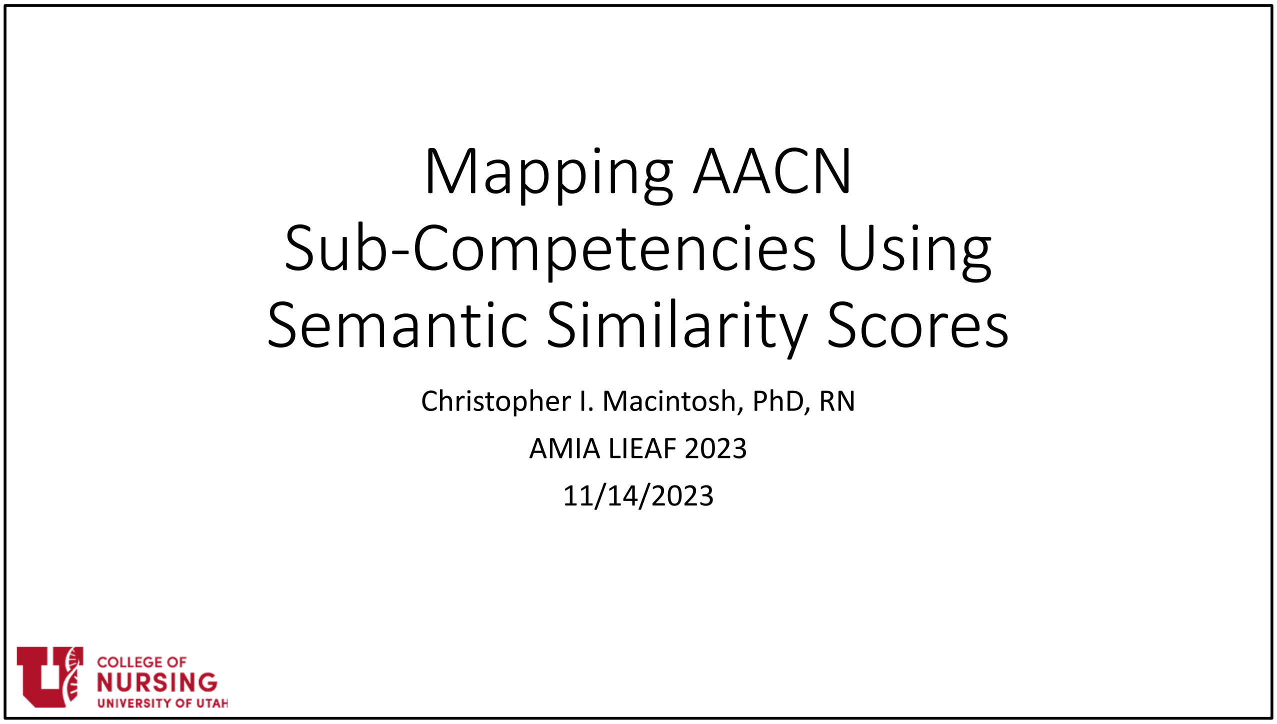


Mapping AACN Sub-Competencies Using Semantic Similarity Scores

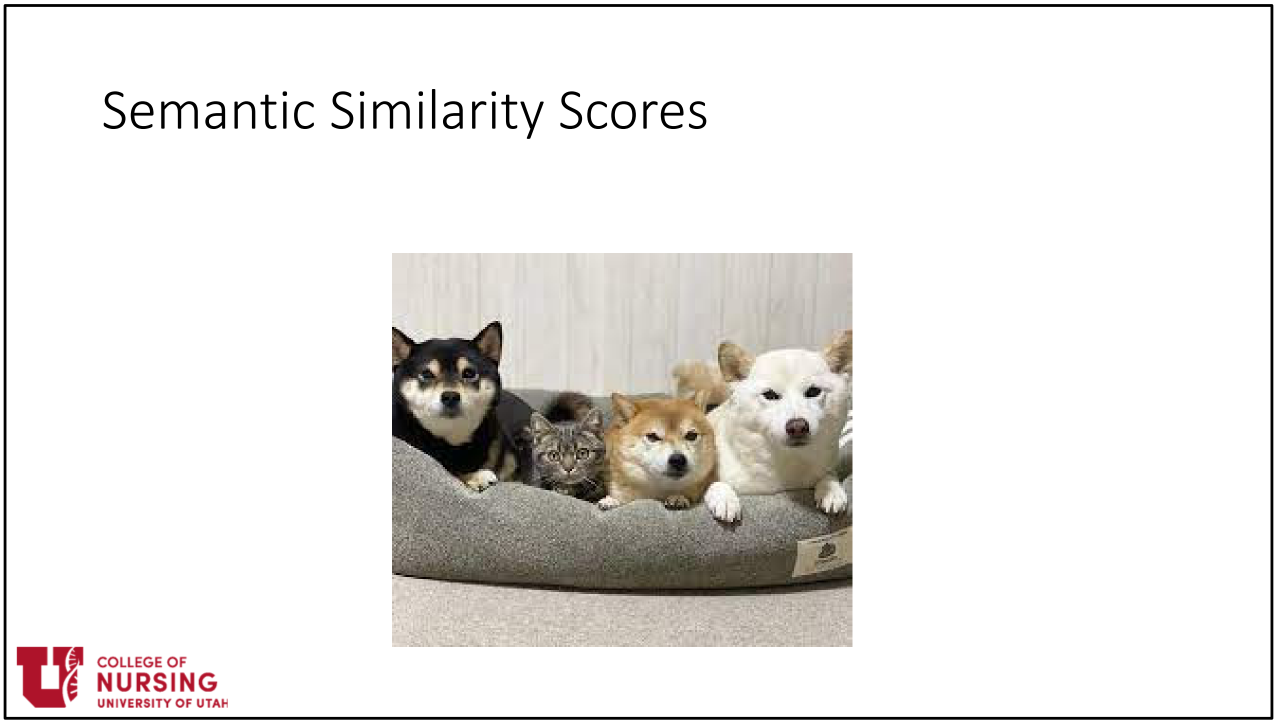
Christopher I. Macintosh, PhD, RN

AMIA LIEAF 2023

11/14/2023

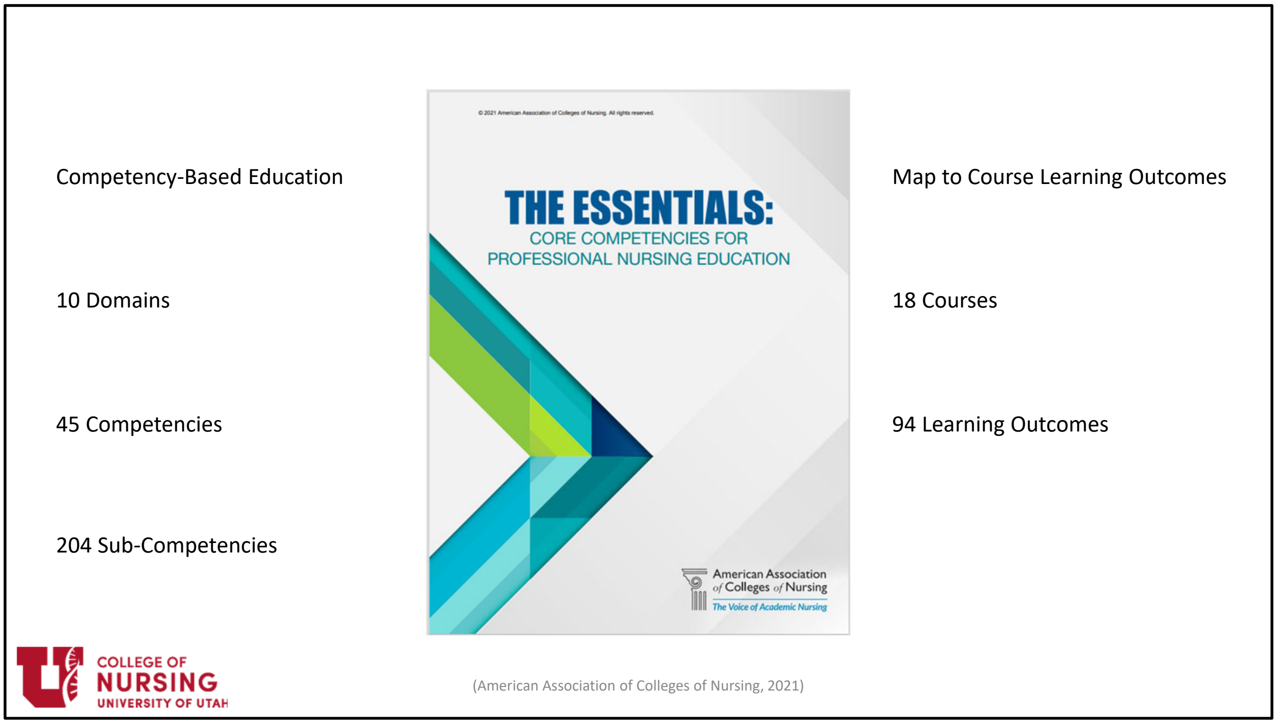
COLLEGE OF NURSING
UNIVERSITY OF UTAH

- Hello.
- I’m Chris Macintosh.
- I’m the director of the Nursing Informatics program at the University of Utah College of Nursing.
- I’m also adjunct faculty in the Biomedical Informatics program in the U of U School of Medicine.
- Today I’m going to demonstrate how to generate semantic similarity scores using a large language model (LLM).
- [click]



Semantic Similarity Scores

- The lines between quantitative and qualitative analysis are getting blurred.
- LLMs can be used to perform quantitative analysis on similarity in meaning of textual data.
- This is a demo of calculating semantic similarity scores for different groups of sentences.
- [click]



Competency-Based Education

10 Domains

45 Competencies

204 Sub-Competencies

© 2021 American Association of Colleges of Nursing. All rights reserved.

THE ESSENTIALS:
CORE COMPETENCIES FOR
PROFESSIONAL NURSING EDUCATION

American Association
of Colleges of Nursing
The Voice of Academic Nursing

Map to Course Learning Outcomes

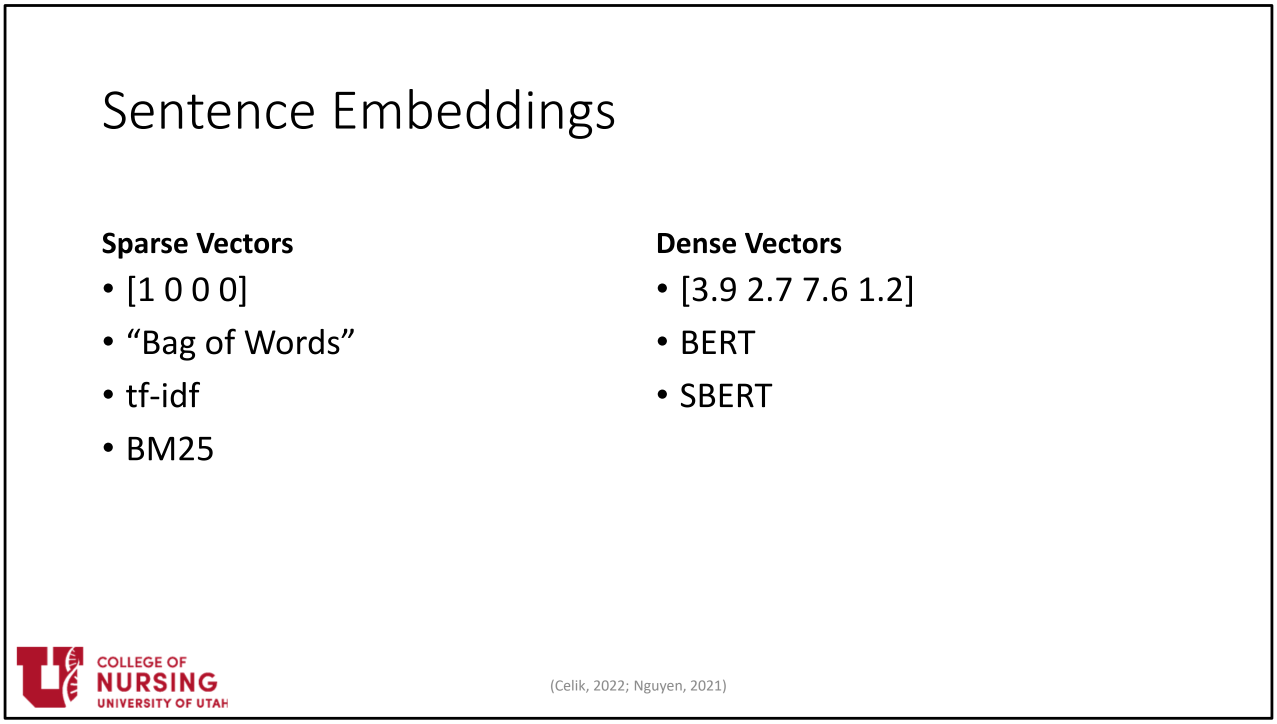
18 Courses

94 Learning Outcomes

COLLEGE OF
NURSING
UNIVERSITY OF UTAH

(American Association of Colleges of Nursing, 2021)

- With the publication of The Essentials: Core Competencies for Professional Nursing Education, the American Association of Colleges of Nursing (AACN) has outlined a strategy to implement competency-based education in undergraduate and graduate nursing programs.
- AACN has defined ten domains for undergraduate and graduate-level competencies.
- AACN has further defined 45 competencies that fall under the ten domains.
- The domains and competencies are identical for undergraduate and graduate education, but separate sub-competencies have been defined for entry-level and advanced-level nursing education.
- AACN has defined 204 sub-competencies for advanced-level education with the intent that measurable sub-competencies will help demonstrate competency attainment.
- Mapping AACN sub-competencies to course learning outcomes is not a trivial task, requiring comparing multiple learning outcomes for multiple courses in a graduate nursing program with the 204 AACN sub-competencies.
- I attempted this proof-of-concept using learning outcomes from our Nursing Informatics curricula. The POC demonstrated here mapped 94 learning outcomes from 18 of our courses with the 204 AACN sub-competencies.
- [click]



Sentence Embeddings

Sparse Vectors

- [1 0 0 0]
- “Bag of Words”
- tf-idf
- BM25

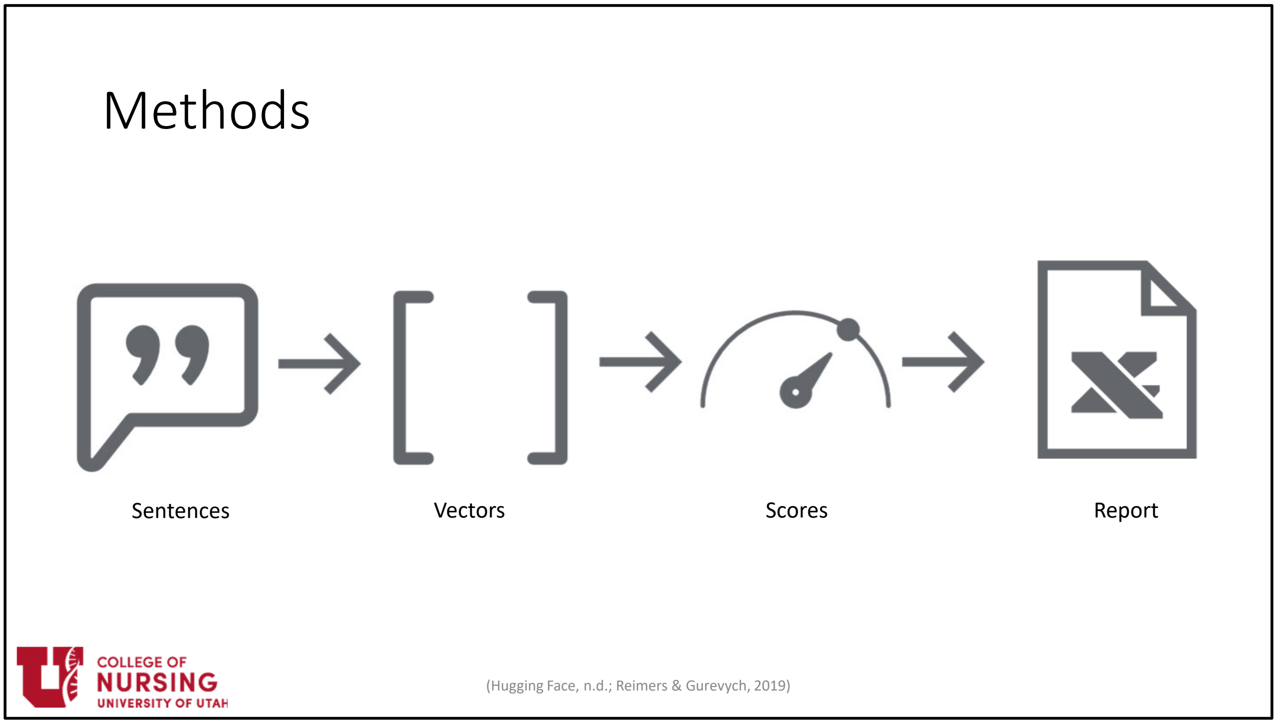
Dense Vectors

- [3.9 2.7 7.6 1.2]
- BERT
- SBERT

COLLEGE OF NURSING UNIVERSITY OF UTAH

(Celik, 2022; Nguyen, 2021)

- I won’t get into much detail about how sentence embedding is done.
- It is sufficient to know that sentence embedding replaces words with arrays of numbers called vectors, and mathematical techniques can be used to compare the vectors.
- Older embedding methods used sparse vectors, where the vectors consisted mostly of zeros.
- Older methods were called “bag of words”, and focused on the words present, but ignored word order.
- More sophisticated methods like term frequency – inverse document frequency (tf-idf) and best match 25 (BM25) expanded on “bag of words” by indicating the significance of words.
- More sophisticated neural network methods result in dense vectors that identify the location of a sentence in a large vector space.
- Most of the values in dense vectors are non-zero.
- Dense vector methods can encode semantics.
- (BERT) or Bidirectional encoder representations from transformers models also encode how words are related to each other.
- Sentence-BERT is an improvement on BERT that speeds up comparisons.
- The vector examples shown here are just for demonstration. The dense vectors created for this POC were several hundred dimensions.
- [click]



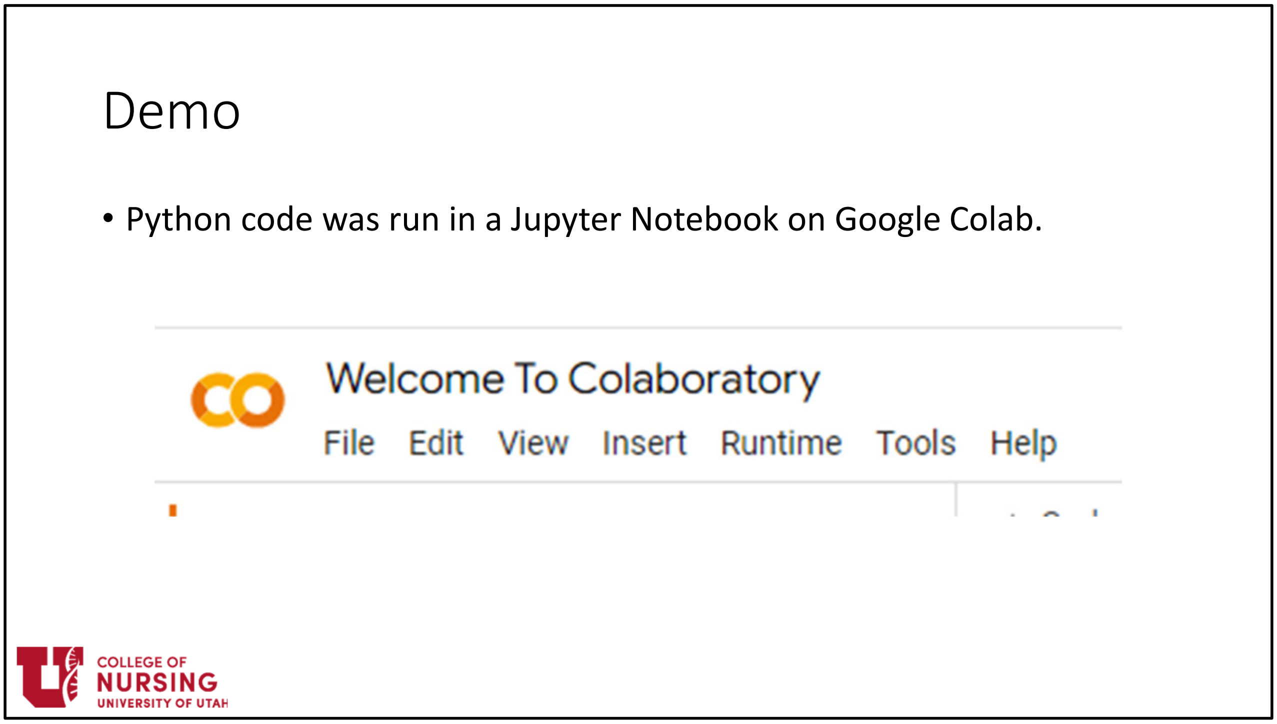
Methods

Sentences → Vectors → Scores → Report

COLLEGE OF NURSING UNIVERSITY OF UTAH

(Hugging Face, n.d.; Reimers & Gurevych, 2019)

- This is a basic overview of the steps I completed.
- Sentence vector embeddings were created for the AACN Essentials sub-competencies and course learning outcomes using the Python sentence-transformers package with the all-mpnet-base-v2 model.
- Cosine similarity scores were calculated for all sub-competency and learning outcome pairings.
- Scores were exported to an Excel workbook and conditional formatting and Excel list sorting were used to identify sub-competency and learning outcome pairs with high scores as potential mappings.
- [click]

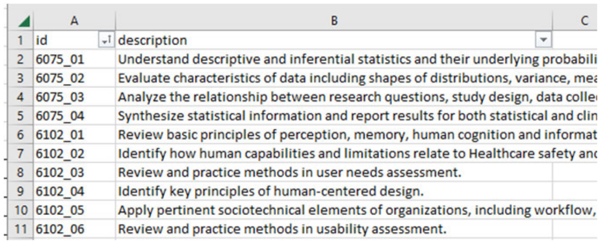


Demo

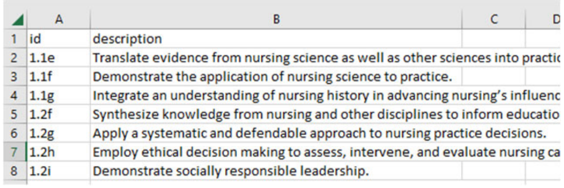
- Python code was run in a Jupyter Notebook on Google Colab.

- Python code was run in a Jupyter Notebook on Google Colab.
- [click]

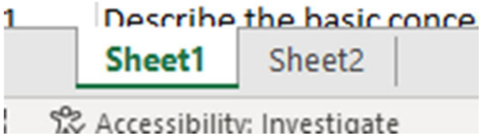
File Preparation



	A	B	C
1	id	description	
2	6075_01	Understand descriptive and inferential statistics and their underlying probabili	
3	6075_02	Evaluate characteristics of data including shapes of distributions, variance, mea	
4	6075_03	Analyze the relationship between research questions, study design, data colle	
5	6075_04	Synthesize statistical information and report results for both statistical and clin	
6	6102_01	Review basic principles of perception, memory, human cognition and informat	
7	6102_02	Identify how human capabilities and limitations relate to Healthcare safety anc	
8	6102_03	Review and practice methods in user needs assessment.	
9	6102_04	Identify key principles of human-centered design.	
10	6102_05	Apply pertinent sociotechnical elements of organizations, including workflow,	
11	6102_06	Review and practice methods in usability assessment.	



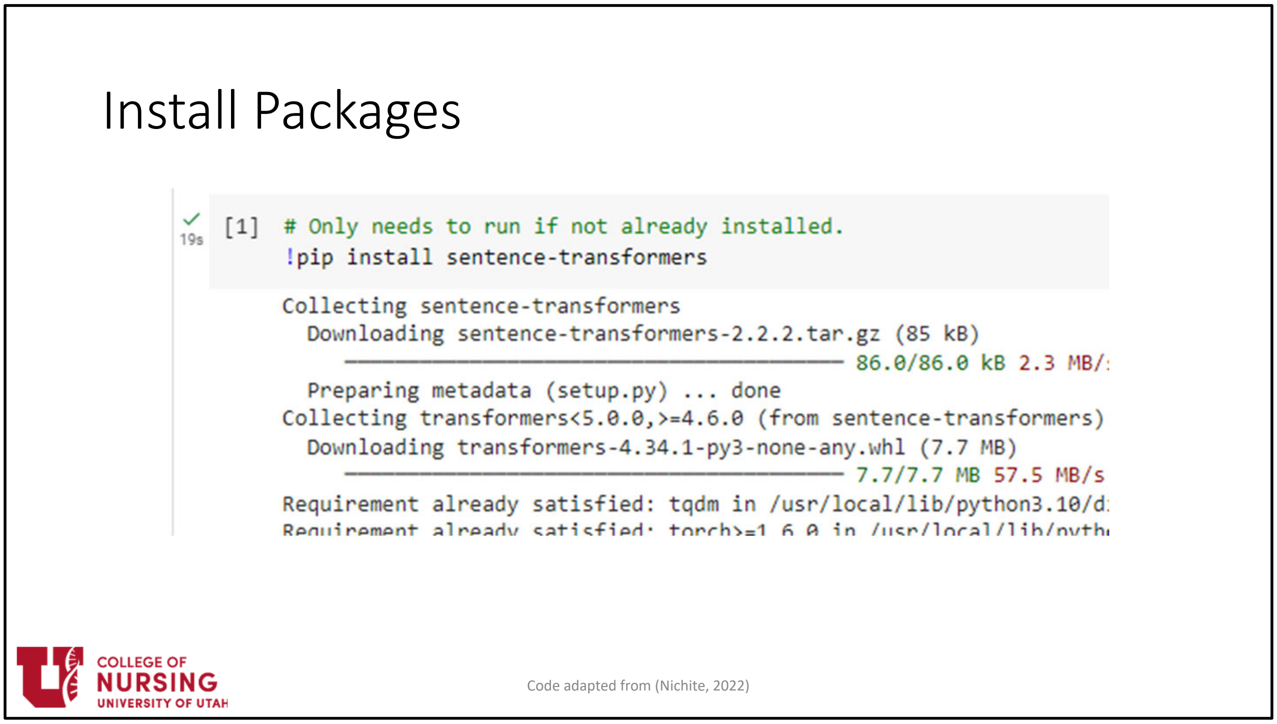
	A	B	C	D
1	id	description		
2	1.1e	Translate evidence from nursing science as well as other sciences into practic		
3	1.1f	Demonstrate the application of nursing science to practice.		
4	1.1g	Integrate an understanding of nursing history in advancing nursing's influenc		
5	1.2f	Synthesize knowledge from nursing and other disciplines to inform educatio		
6	1.2g	Apply a systematic and defendable approach to nursing practice decisions.		
7	1.2h	Employ ethical decision making to assess, intervene, and evaluate nursing ca		
8	1.2i	Demonstrate socially responsible leadership.		



1 Describe the basic conce

COLLEGE OF NURSING
UNIVERSITY OF UTAH

- First, I needed to prepare a data file with the text needed.
- An Excel file with two sheets was created.
- Both sheets had an ID column and a Description column.
- These could be any two sets of sentences you want to map with each other.
- [click]

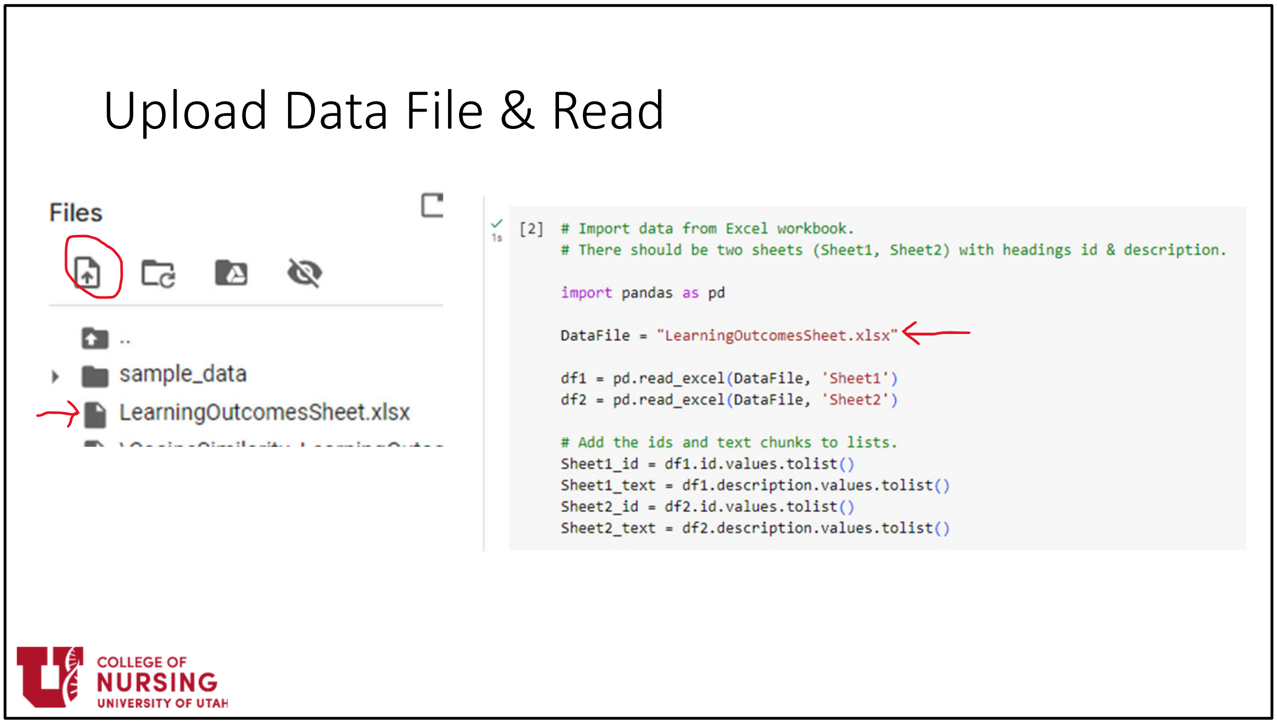


Install Packages

```
[1] # Only needs to run if not already installed.
    !pip install sentence-transformers

    Collecting sentence-transformers
      Downloading sentence-transformers-2.2.2.tar.gz (85 kB)
         ---------------------------------------- 86.0/86.0 kB 2.3 MB/:
      Preparing metadata (setup.py) ... done
    Collecting transformers<5.0.0,>=4.6.0 (from sentence-transformers)
      Downloading transformers-4.34.1-py3-none-any.whl (7.7 MB)
         ---------------------------------------- 7.7/7.7 MB 57.5 MB/s
    Requirement already satisfied: tqdm in /usr/local/lib/python3.10/d:
    Requirement already satisfied: torch>=1.6.0 in /usr/local/lib/pyth
```

Code adapted from (Nichite, 2022)

- This is Python code to install the sentence-transformers package.
- [click]



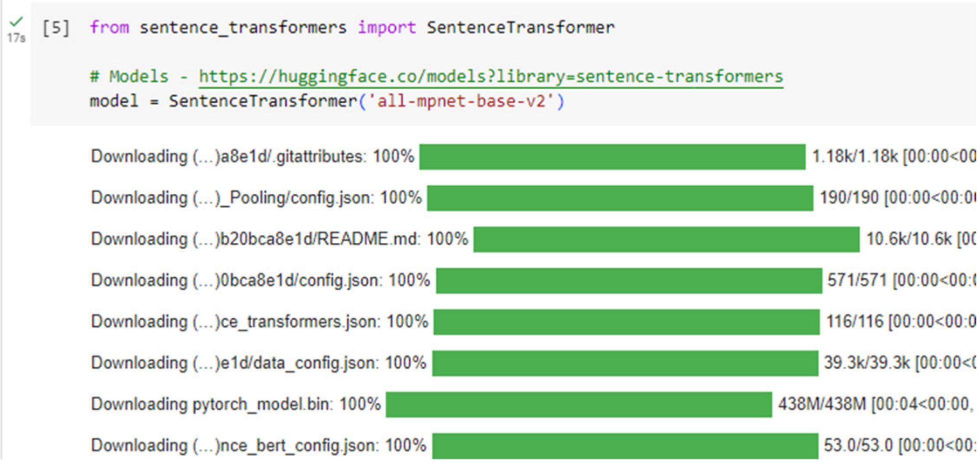
Upload Data File & Read

```
[2] # Import data from Excel workbook.
    # There should be two sheets (Sheet1, Sheet2) with headings id & description.

    import pandas as pd

    DataFile = "LearningOutcomesSheet.xlsx"

    df1 = pd.read_excel(DataFile, 'Sheet1')
    df2 = pd.read_excel(DataFile, 'Sheet2')

    # Add the ids and text chunks to lists.
    Sheet1_id = df1.id.values.tolist()
    Sheet1_text = df1.description.values.tolist()
    Sheet2_id = df2.id.values.tolist()
    Sheet2_text = df2.description.values.tolist()
```

COLLEGE OF NURSING
UNIVERSITY OF UTAH

- You need to upload the datafile on Google Colab and make sure the filename matches in the Python code.
- [click]

Import the Model



```
[5]  from sentence_transformers import SentenceTransformer

     # Models - https://huggingface.co/models?library=sentence-transformers
     model = SentenceTransformer('all-mpnet-base-v2')
```

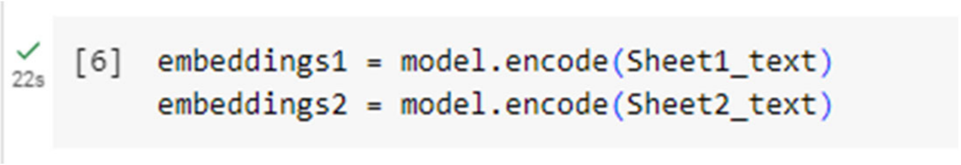
```
Downloading (…)a8e1d/.gitattributes: 100%   1.18k/1.18k [00:00<00
Downloading (…)_Pooling/config.json: 100%   190/190 [00:00<00:0
Downloading (…)b20bca8e1d/README.md: 100%   10.6k/10.6k [0(
Downloading (…)0bca8e1d/config.json: 100%   571/571 [00:00<00:(
Downloading (…)ce_transformers.json: 100%   116/116 [00:00<00:0
Downloading (…)e1d/data_config.json: 100%   39.3k/39.3k [00:00<(
Downloading pytorch_model.bin: 100%   438M/438M [00:04<00:00,
Downloading (…)nce_bert_config.json: 100%   53.0/53.0 [00:00<00:
```

COLLEGE OF NURSING
UNIVERSITY OF UTAH

Code adapted from (Nichite, 2022)

- This is Python code to import the desired model.
- There are many models on the HuggingFace site.
- I used a model that showed good performance in other demos I watched.
- [click]

Create the Sentence Embeddings



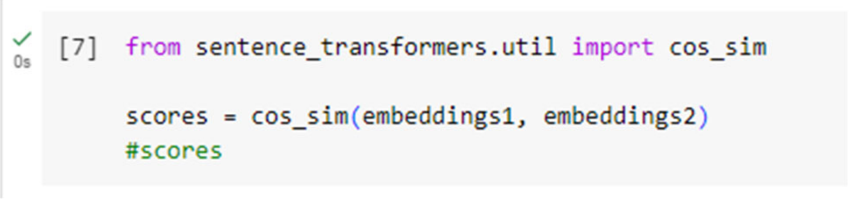
```
embeddings1 = model.encode(Sheet1_text)
embeddings2 = model.encode(Sheet2_text)
```

COLLEGE OF NURSING UNIVERSITY OF UTAH

Code adapted from (Nichite, 2022)

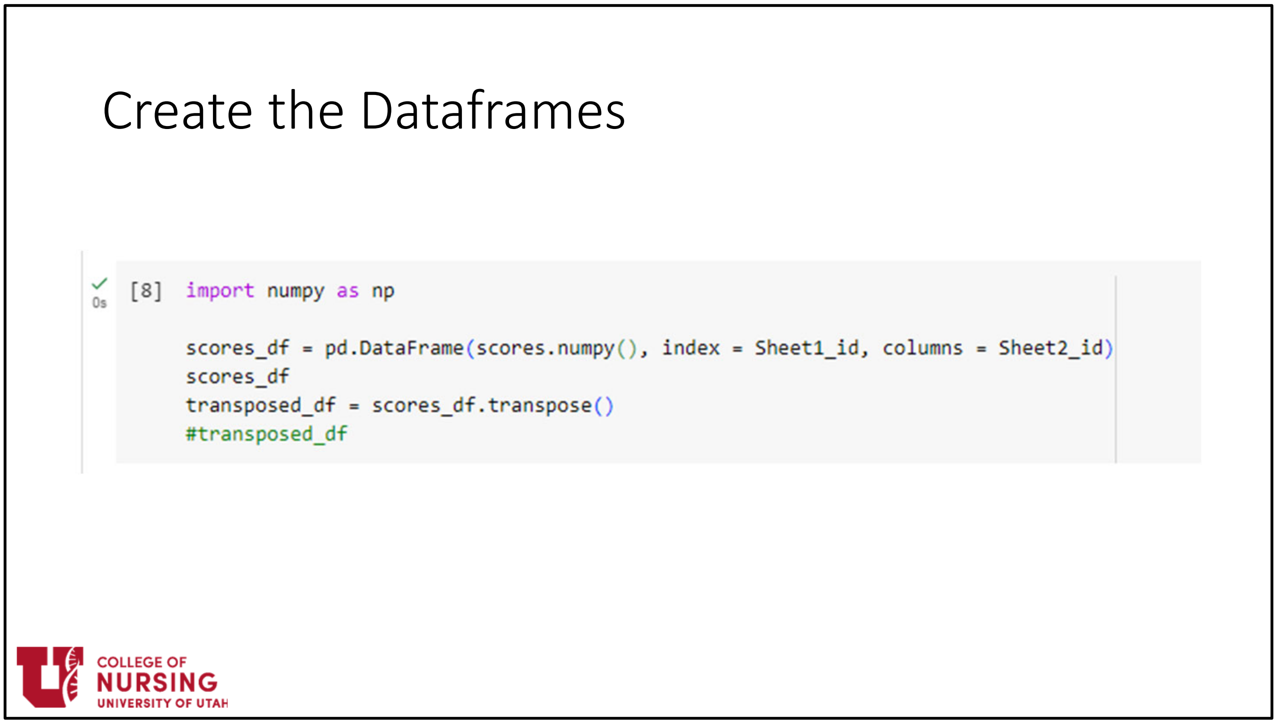
- This is Python code to create the sentence embeddings.
- [click]

Calculate Cosine Similarity Scores



```
from sentence_transformers.util import cos_sim

scores = cos_sim(embeddings1, embeddings2)
#scores
```

- This is Python code to calculate the cosine similarity scores.
- [click]

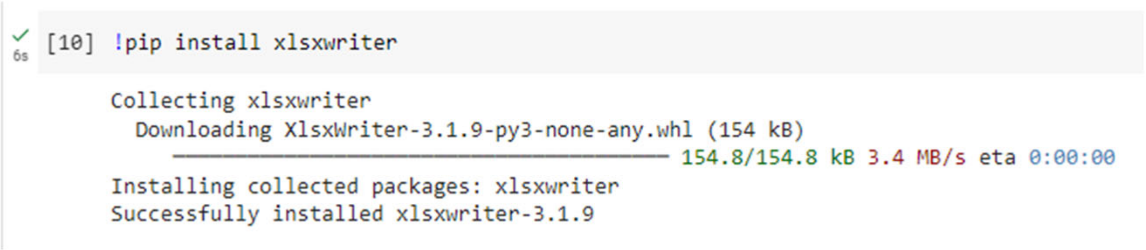


Create the Dataframes

```
import numpy as np

scores_df = pd.DataFrame(scores.numpy(), index = Sheet1_id, columns = Sheet2_id)
scores_df
transposed_df = scores_df.transpose()
#transposed_df
```

- This is Python code to create the dataframes.
- [click]

Package Needed for Excel Formatting

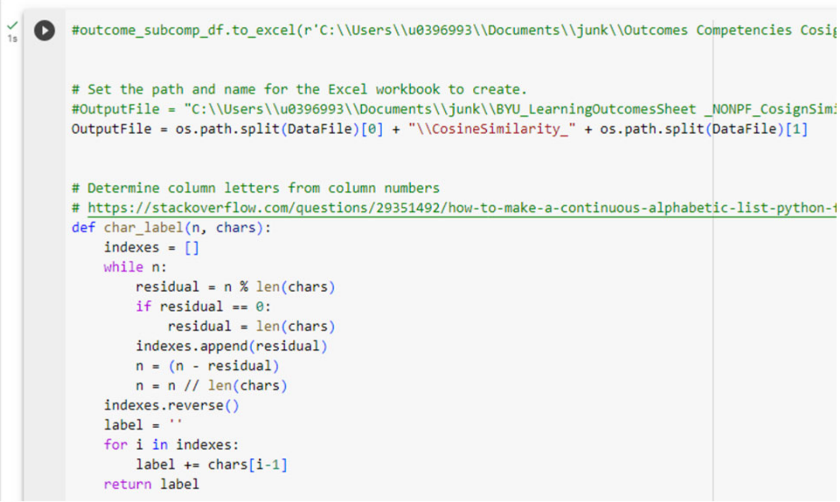


```
!pip install xlsxwriter
```

```
Collecting xlsxwriter
  Downloading XlsxWriter-3.1.9-py3-none-any.whl (154 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 154.8/154.8 kB 3.4 MB/s eta 0:00:00
Installing collected packages: xlsxwriter
Successfully installed xlsxwriter-3.1.9
```

- This is Python code to install the xlsxwriter package needed to format the worksheets.
- [click]

Export & Format Excel File

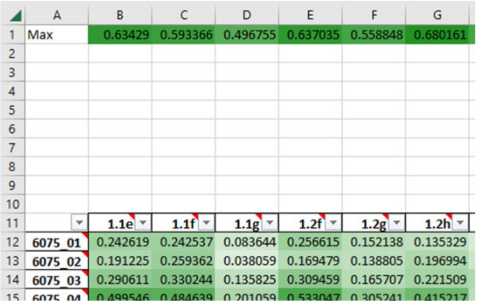


```
#outcome_subcomp_df.to_excel(r'C:\\Users\\u0396993\\Documents\\junk\\Outcomes Competencies Cosi


# Set the path and name for the Excel workbook to create.
#OutputFile = "C:\\Users\\u0396993\\Documents\\junk\\BYU_LearningOutcomesSheet _NONPF_CosignSimi
OutputFile = os.path.split(DataFile)[0] + "\\CosineSimilarity_" + os.path.split(DataFile)[1]


# Determine column letters from column numbers
# https://stackoverflow.com/questions/29351492/how-to-make-a-continuous-alphabetic-list-python-
def char_label(n, chars):
    indexes = []
    while n:
        residual = n % len(chars)
        if residual == 0:
            residual = len(chars)
        indexes.append(residual)
        n = (n - residual)
        n = n // len(chars)
    indexes.reverse()
    label = ''
    for i in indexes:
        label += chars[i-1]
    return label
```

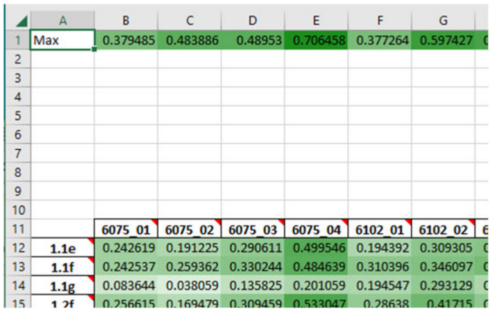
COLLEGE OF NURSING UNIVERSITY OF UTAH

- The Python code to format the Excel sheets was a little longer.
- I won’t show all of that here.
- [click]

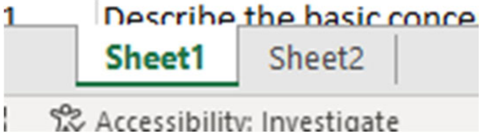
Report Export



	A	B	C	D	E	F	G
1	Max	0.63429	0.593366	0.496755	0.637035	0.558848	0.680161
2							
3							
4							
5							
6							
7							
8							
9							
10							
11		1.1e	1.1f	1.1g	1.2f	1.2g	1.2h
12	6075_01	0.242619	0.242537	0.083644	0.256615	0.152138	0.135329
13	6075_02	0.191225	0.259362	0.038059	0.169479	0.138805	0.196994
14	6075_03	0.290611	0.330244	0.135825	0.309459	0.165707	0.221509
15	6075_04	0.499546	0.484639	0.201059	0.533047	0.305241	0.415217



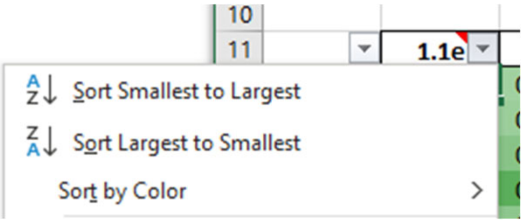
	A	B	C	D	E	F	G
1	Max	0.379485	0.483886	0.48953	0.706458	0.377264	0.597427
2							
3							
4							
5							
6							
7							
8							
9							
10							
11		6075_01	6075_02	6075_03	6075_04	6102_01	6102_02
12	1.1e	0.242619	0.191225	0.290611	0.499546	0.194392	0.309305
13	1.1f	0.242537	0.259362	0.330244	0.484639	0.310396	0.346097
14	1.1g	0.083644	0.038059	0.135825	0.201059	0.194547	0.293129
15	1.2f	0.256615	0.169479	0.309459	0.533047	0.28638	0.41715

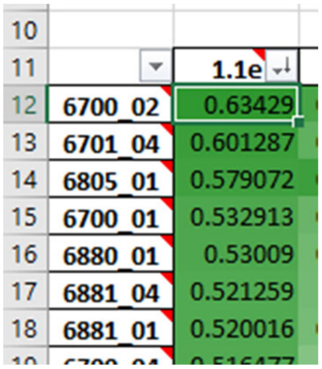


COLLEGE OF NURSING UNIVERSITY OF UTAH

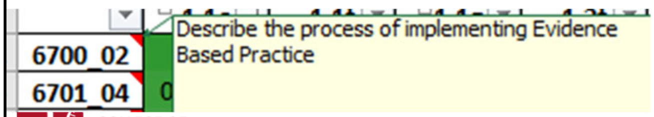
- A report was exported to Excel for easy use and sharing with others.
- An Excel file with two sheets was created.
- One sheet arranged AACN sub-competencies along the columns.
- One sheet arranged course learning outcomes along the columns.
- This allows users to sort the lists either way using the Data Filter.
- Conditional formatting was used to color cells according to the strength of the relationship.
- The maximum score for each column was displayed at the top of each sheet.
- [click]

Sort Scores to Find Suggested Mappings



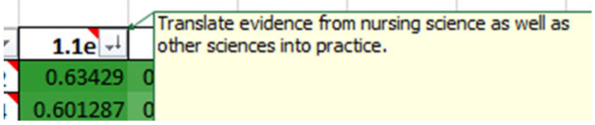


10		
11		1.1e
12	6700_02	0.63429
13	6701_04	0.601287
14	6805_01	0.579072
15	6700_01	0.532913
16	6880_01	0.53009
17	6881_04	0.521259
18	6881_01	0.520016



6700_02
6701_04

Describe the process of implementing Evidence Based Practice



1.1e
0.63429
0.601287

Translate evidence from nursing science as well as other sciences into practice.

COLLEGE OF NURSING
UNIVERSITY OF UTAH

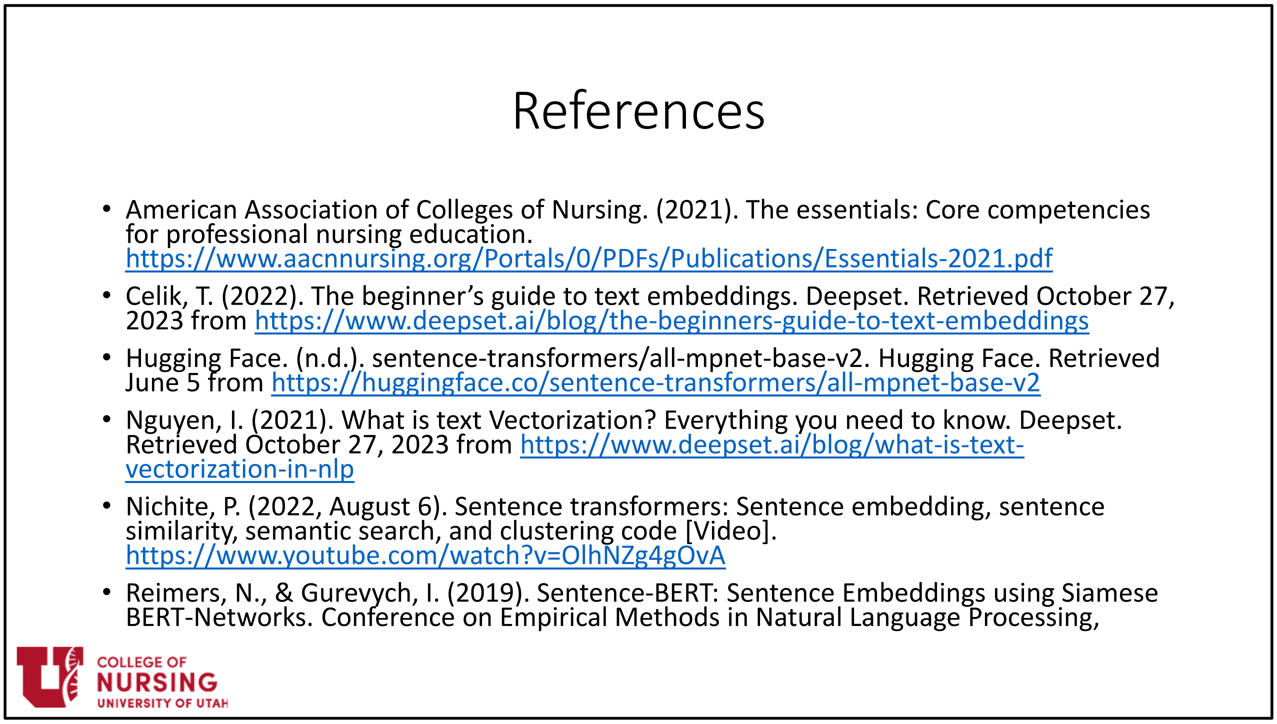
- Map learning outcomes to sub-competencies by sorting scores from largest to smallest.
- Higher scores indicate a better match.
- Python was able to save the text for sub-competencies and learning outcomes as notes that can be seen when you hover over the cell with the ID number.
- Human evaluation is still needed to verify good matches, but this process may speed up the mapping process and improve consistency.
- [click]

Demo Files

- Files for the demo can be found at https://github.com/cmcntsh/SemanticSimilarityReport_AMIA_LIEAF_2023

COLLEGE OF NURSING UNIVERSITY OF UTAH

- The files for the demo are on a GitHub repository.
- Anyone should be able to run the code in Google Colab with their own data file.
- [click]



References

- American Association of Colleges of Nursing. (2021). The essentials: Core competencies for professional nursing education. https://www.aacnnursing.org/Portals/0/PDFs/Publications/Essentials-2021.pdf
- Celik, T. (2022). The beginner’s guide to text embeddings. Deepset. Retrieved October 27, 2023 from https://www.deepset.ai/blog/the-beginners-guide-to-text-embeddings
- Hugging Face. (n.d.). sentence-transformers/all-mpnet-base-v2. Hugging Face. Retrieved June 5 from https://huggingface.co/sentence-transformers/all-mpnet-base-v2
- Nguyen, I. (2021). What is text Vectorization? Everything you need to know. Deepset. Retrieved October 27, 2023 from https://www.deepset.ai/blog/what-is-text-vectorization-in-nlp
- Nichite, P. (2022, August 6). Sentence transformers: Sentence embedding, sentence similarity, semantic search, and clustering code [Video]. https://www.youtube.com/watch?v=OlhNZg4gOvA
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Conference on Empirical Methods in Natural Language Processing,

COLLEGE OF
NURSING
UNIVERSITY OF UTAH

- Here are the references for this presentation.
- [End of Presentation]