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

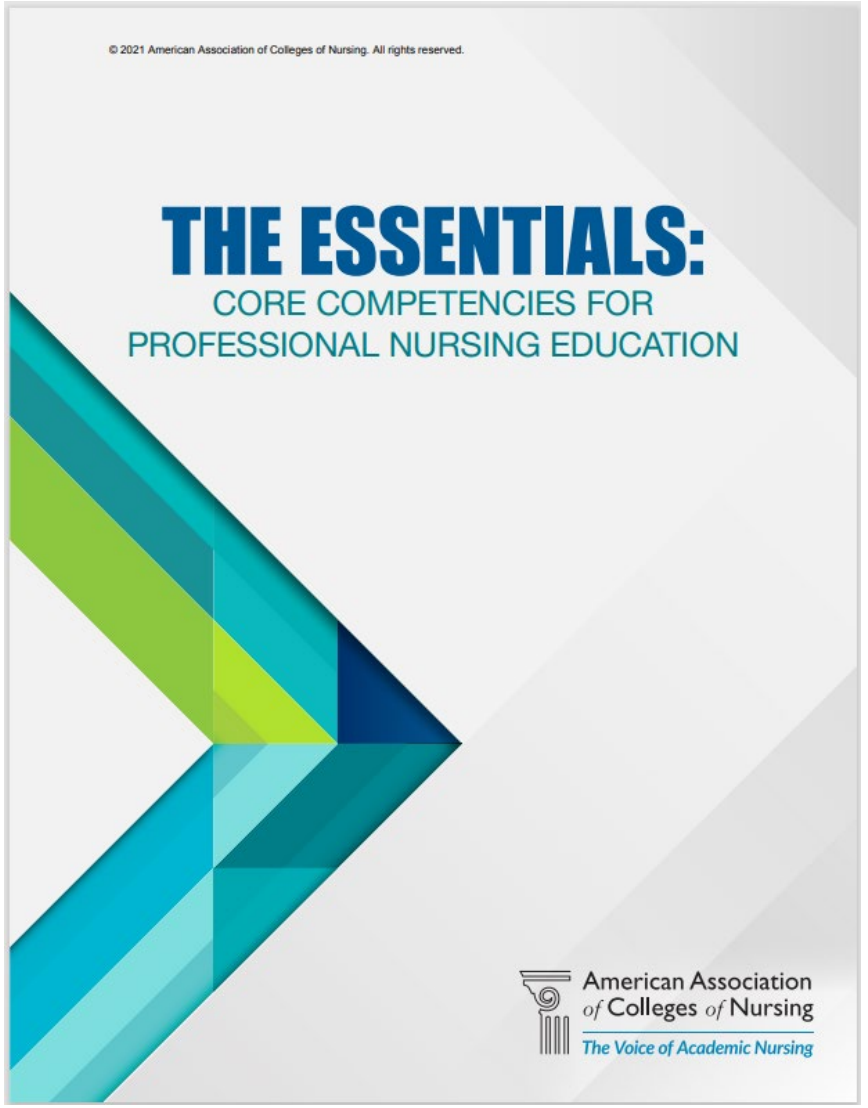
Semantic Similarity Scores

Competency-Based Education

10 Domains

45 Competencies

204 Sub-Competencies



Map to Course Learning Outcomes

18 Courses

94 Learning Outcomes

COLLEGE OF NURSING
UNIVERSITY OF UTAH

(American Association of Colleges of Nursing, 2021)

Sentence Embeddings

Sparse Vectors

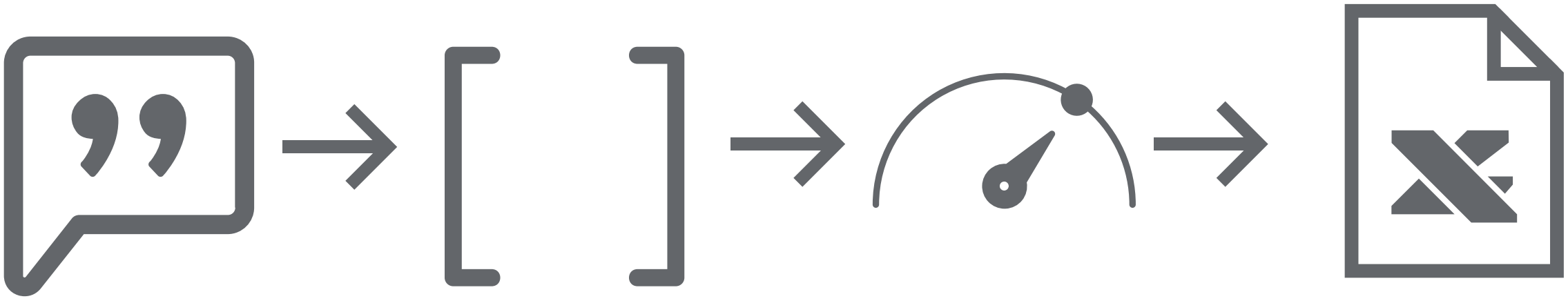
- [1 0 0 0]
- “Bag of Words”
- tf-idf
- BM25

Dense Vectors

- [3.9 2.7 7.6 1.2]
- BERT
- SBERT

(Celik, 2022; Nguyen, 2021)

Methods

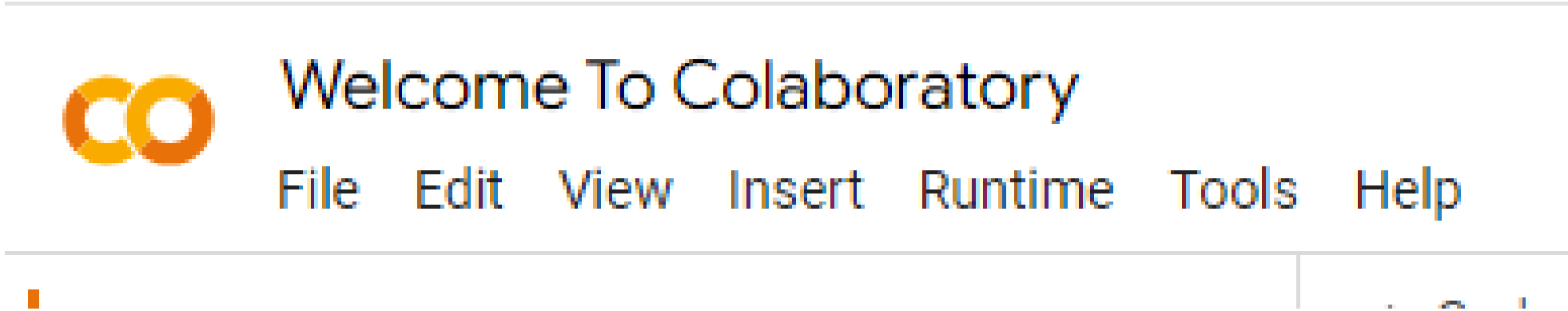


Sentences → Vectors → Scores → Report

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

Demo

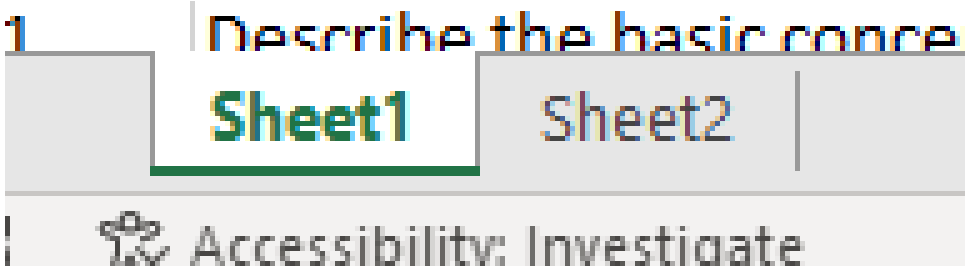
- Python code was run in a Jupyter Notebook on Google Colab.



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 and	
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.		



COLLEGE OF NURSING
UNIVERSITY OF UTAH

Install Packages

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

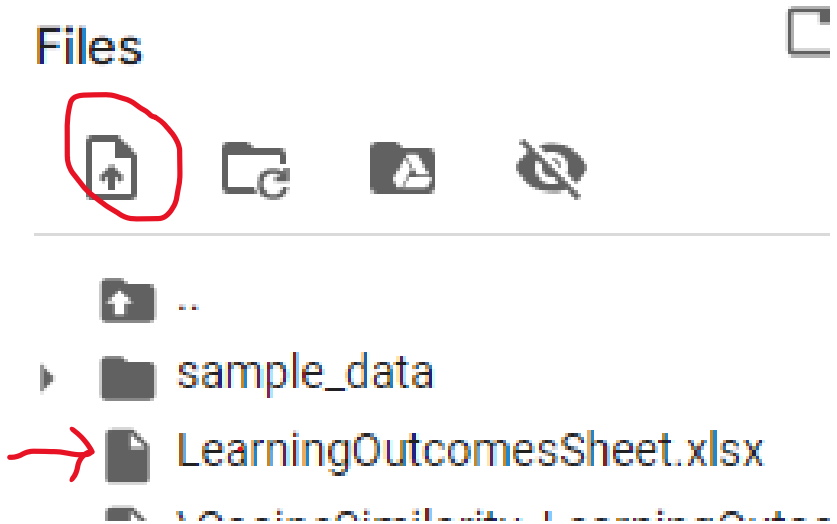
19s

```
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
```

COLLEGE OF NURSING
UNIVERSITY OF UTAH

Code adapted from (Nichite, 2022)

Upload Data File & Read

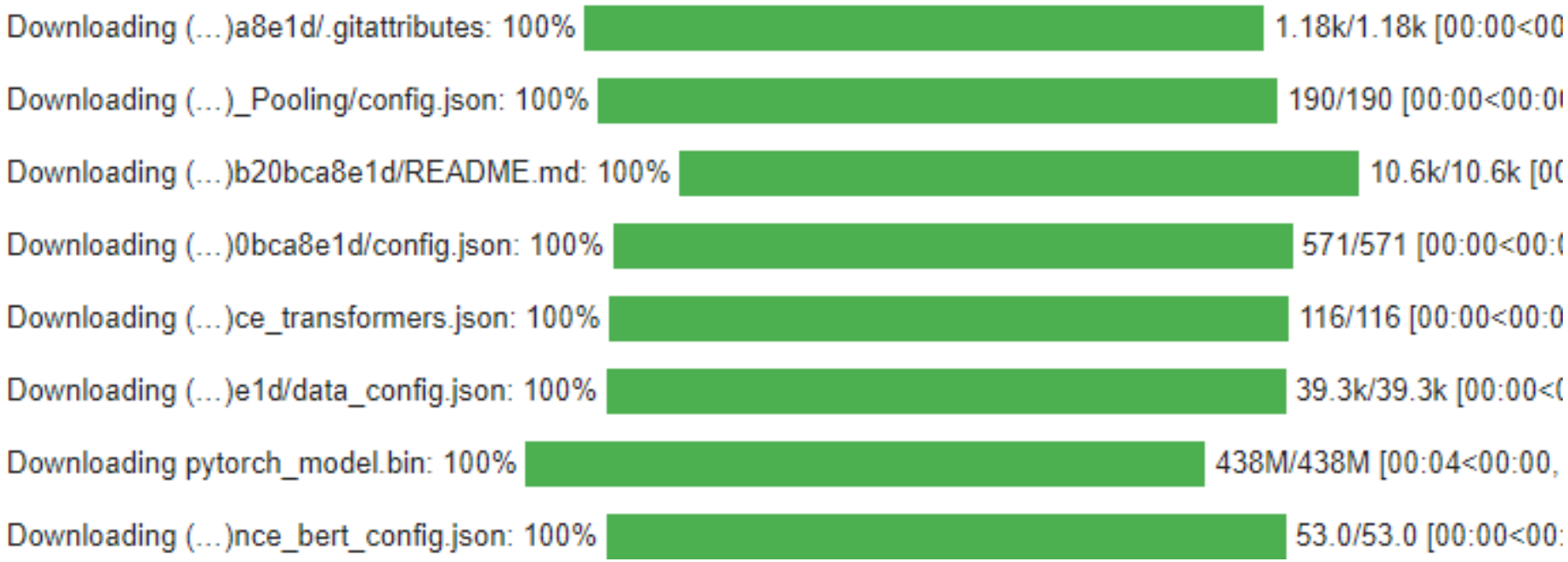



```
[2]  # Import data from Excel workbook.
     # There should be two sheets (Sheet1, Sheet2) with headings id & description.

     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()
```

Import the Model

```
[5]  from sentence_transformers import SentenceTransformer

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

17s



COLLEGE OF NURSING
UNIVERSITY OF UTAH

Code adapted from (Nichite, 2022)

Create the Sentence Embeddings

```
[6]  embeddings1 = model.encode(Sheet1_text)
     embeddings2 = model.encode(Sheet2_text)
```

22s

Code adapted from (Nichite, 2022)

Calculate Cosine Similarity Scores

```
[7]  from sentence_transformers.util import cos_sim

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

0s

Create the Dataframes

```
[8]  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
```

Package Needed for Excel Formatting

```
[10] !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
```

Export & Format Excel File

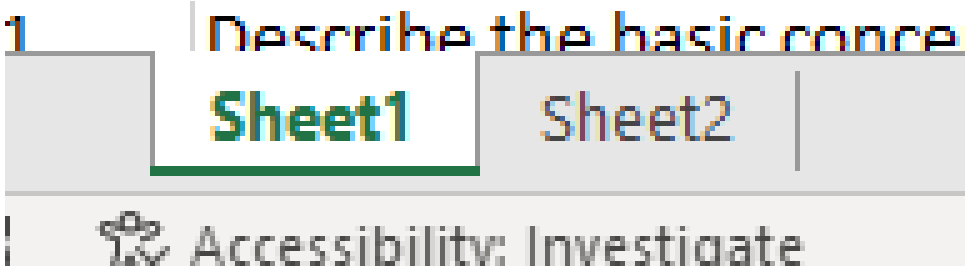
```
#outcome_subcomp_df.to_excel(r'C:\\Users\\u0396993\\Documents\\junk\\Outcomes Competencies Cosig


# 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-f
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

Report Export

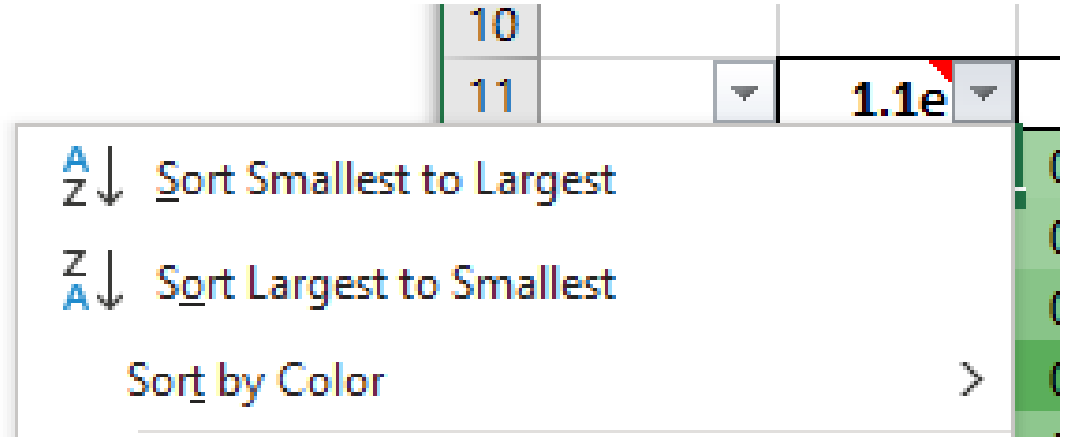
	A	B	C	D	E	F	G
1	Max	0.63429	0.593366	0.496755	0.637035	0.558848	0.680161
2							
3							
4							
5							
6							
7							
8							
9							
10							
11		1.1e	1.1f	1.1g	1.2f	1.2g	1.2h
12	6075_01	0.242619	0.242537	0.083644	0.256615	0.152138	0.135329
13	6075_02	0.191225	0.259362	0.038059	0.169479	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

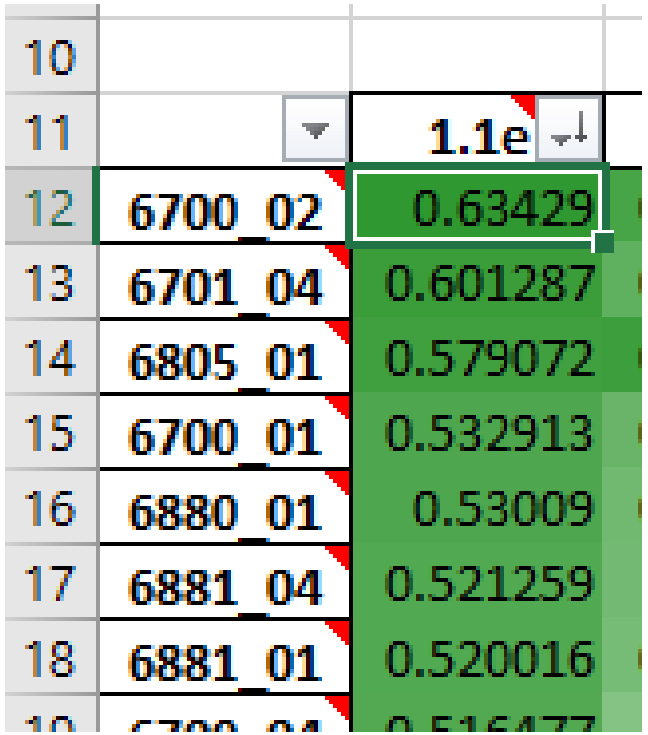
	A	B	C	D	E	F	G
1	Max	0.379485	0.483886	0.48953	0.706458	0.377264	0.597427
2							
3							
4							
5							
6							
7							
8							
9							
10							
11		6075_01	6075_02	6075_03	6075_04	6102_01	6102_02
12	1.1e	0.242619	0.191225	0.290611	0.499546	0.194392	0.309305
13	1.1f	0.242537	0.259362	0.330244	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

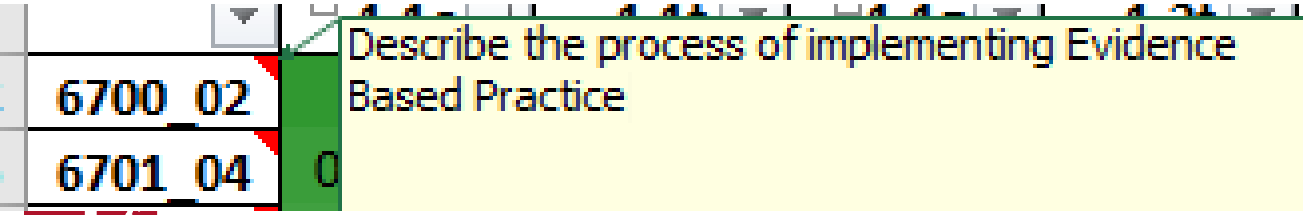


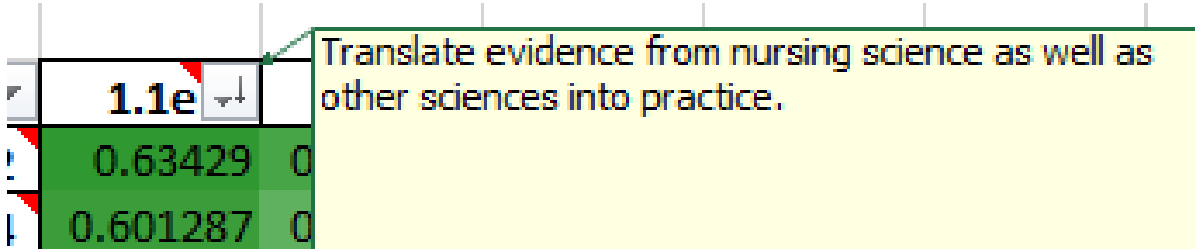
COLLEGE OF NURSING
UNIVERSITY OF UTAH

Sort Scores to Find Suggested Mappings









COLLEGE OF NURSING
UNIVERSITY OF UTAH

Demo Files

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

COLLEGE OF NURSING
UNIVERSITY OF UTAH

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,