Data Driven Approach to Analyzing CSPS' Curriculum

Using natural language processing to reduce gaps and identify overlaps

CANDEV

Motivation

- Identify main themes in the CSPS' courses
- Identify redundancies
- **Improve Upkeep:** Compare potential future courses to existing ones

Solution

Sentence Embedding using Transformers

- A neural network that uses self-attention to associate sentence fragments
- SOTA accuracy at semantic understanding
  - Outperform bi-dir LSTMs
- Produces 512-dim vector to represent high-level representation
- Used transfer learning on CSPS data from pretrained model

Solution

t-Distributed Stochastic Neighbour Embedding **(t-SNE)**

- Dimensionality reduction technique

  (512 dim -> 3 dim)
- Data can then be visualized in 3-dimensional space
- Requires iteration to learn low-dimensional representation

t-SNE Demonstration

Determining Course Similarity

Search:

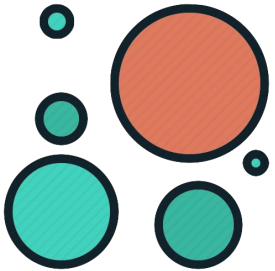
Example (new) Description: This course seeks to inform students of the complexities associated with language acquisition, and strategies to aid ESL students with this skill.

Most similar courses:

1. Master Key Competencies in Oral and Reading Comprehension - English as a Second Language (E441)
2. Master Key Competencies in Oral and Reading Comprehension - French as a Second Language (E433)
3. Managing Change: Building Positive Support for Change (X031)
4. Planning an Effective Presentation (C060)
5. Preparation for the English as a Second Language Evaluation: Written Expression and Reading Comprehension (C256)

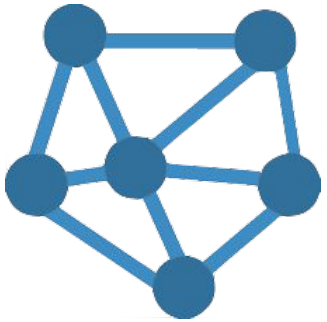
Further Applications

What can be done with this code?



Topic Allocation

Assess Course Ideas



Semantic Understanding