**Clustering Student Responses**

Clustering provides teachers with a capability to lead discussions based on the students’ responses in groups of belief.

Extract features from Question and Reference Answer to create a feature vector:

\[ \langle F_1, F_2, ..., F_n \rangle \]

Weigh each student response against feature vector:

- student 1: \[ f_1 f_2 ... f_n \]
- student 2: \[ f_1 f_2 ... f_n \]
- student n: \[ f_1 f_2 ... f_n \]

**k-means clustering**

Iterate 10 times unless convergence, then iterate through each response feature vector and assign each feature vector to the nearest centroid.

At the conclusion, adjust the centroid of each cluster.

**Pseudocode:**

while i < 10 or Converge:

- for J \( \in \) Response Feature Vector:
  - Assign Closest Centroid(j)
- for K \( \in \) Centroid
  - Adjust Centroid

**Vocabulary for Students**

- Extract all the content words from all RA_1...RA_n

List W = \{RA_1, RA_2, ..., RA_n\}

- for i \( \in \) List W
  - Synonym(i)
  - Antonym(i)

As students are responding to a specific question, append to List W the most frequently used words.

List W has a max of 5 x Number of Content Words in the current reference answer.

**SEEDING Development Team**

Dr. Rodney Nielsen,
Associate Professor, CSE
Director, HiLT Lab

UNT CSE Student Developers
Mingyu Lin
James Glenn

Frank Paiva

HiLT Lab

Human Language Technologies
Health Informatics Language Technologies
Dr. Rodney Nielsen

UNIVERSITY OF NORTH TEXAS®