

Approach to evaluate short answers

Two major steps:

- Extract content words
 - Get rid of stop words, retain only content tokens.
 - Identify semantic related words using Wordnet.
 - Wordnet matches categorized into seven categories, rate these matches from 0 to 6 to assign scores.

Input: Rubric text, top-scoring answers, and prompt
and stimulus texts (if available)

Output: Patterns containing unordered content
words.

Algorithm: for each sentence in the rubric text do

/ Rubric text: "size or type of container to use" */*

1. Remove stopwords or relatively common words.

/ Output: size type container use*/*

*2. Rank tokens in top-scoring answers, and prompt and stimulus texts based
on their frequency, and select the top most frequent tokens.*

/ size container type*/*

*3. Identify classes of alternate tokens, for each rubric token, from among
most frequent tokens (from Step 2).*

/ {size, large, mass, thing, volume} {container, cup, measure} {type, kind} */*

4. Stem words and use the suffix as an alternative

/ container → (stem: contain, suffix: er) → contain(er)?/*

5. Generate the pattern by AND-ing each of the classes of words.

/ (?=.*(large|mass|size|thing|volume).*)(?=.*(contain(er)?|cup|measure).*(
?=.*(kind|type).*)*

End

- Structure of sentences information
 - Get context text in word order (subject-verb, verb-object etc) form.
 - Generate word order graphs by using POS Tagger to tag parts of speech and then club all the **consecutive** nouns components to a noun vertex, verb components to a verb entity and various other components.
 - Group the components to form a graph ex: verb vertex looks for noun vertex to form verb-noun edge. Ordering matters here noun-verb is different from verb-noun.(Refer Ramachandran and Gehring (2012).)
 - Substitute stop words with regular expressions for **gramatically incorrect sentences**.
 - Semantic alternatives(words from Wordnet) also to be taken care of.

Input: Rubric text, top-scoring answers, and prompt and stimulus texts (if available)

Output: Patterns containing ordered word phrases.

Algorithm: for each sentence in the rubric text do

/ Rubric text: ". . .particles like sodium, potassium ions into membranes. . ."*

1. Generate word-order graphs from the text, and extract edges from the word-order graph.

/ The extracted segment: particles like—sodium potassium--ions into membranes. Graph edges are connected with a "--"*

2. Replace stopwords or function words with $\backslash w\{0,4\}$.

/ The segment becomes: particles($\backslash s\backslash w\{0,4\}\backslash s$){0,1} sodium potassium ions($\backslash s\backslash w\{0,4\}\backslash s$){0,1}membranes*

3. Rank tokens in top-scoring answers and prompt and stimulus texts based on their frequency, and select the top most frequent tokens.

4. Identify class of alternate tokens, for each rubric token, from among most frequent tokens.

5. Add all synonyms of the rubric token from WordNet to the class of alternatives.

/ E.g. class of alternate tokens for sodium: {potassium, bismuth, zinc, cobalt}, for potassium: {tungsten, zinc, calcium, iron, aluminum, tin}, for membrane: {film, sheet}*

6. Stem words and generate pattern by AND-ing all classes of words.

End