Knowledge modeling

Tom Rochette <tom.rochette@coreteks.org>

August 30, 2025 — 861fb9d0

0.1 Context

0.2 Learned in this study

0.3 Things to explore

- Vocabulary
  - Size of vocabulary vs another entity
  - Overlapping/Intersecting vocabulary with another entity

1 Overview

$$Vocabulary\ of\ entity\ 1 = V_1 Cardinality\ of\ the\ vocabulary\ of\ entity\ 1 = |V_1|$$

$$Vocabulary\ of\ entity\ 2 = V_2 Cardinality\ of\ the\ vocabulary\ of\ entity\ 2 = |V_1|$$

$$Common\ vocabulary = V_1 \cap V_2 Shared\ vocabulary = V_1 \cup V_2 Vocabulary\ only\ in\ V_1 = V_1 - V_2 Vocabulary\ only\ in\ V_2 = V_2 - V_1 U$$

$$Jaccard\ index = J(V_1, V_2) = \frac{|V_1 \cap V_2|}{|V_1 \cup V_2|} Size\ relative\ Jaccard\ index = |V_1 \cap V_2| \times J(V_1, V_2) = 2\frac{|V_1 \cap V_2|}{|V_1 \cup V_2|}$$

$$Overlap\ coefficient = overlap(V_1, V_2) = \frac{|V_1 \cap V_2|}{\min(|V_1|, |V_2|)}$$

2 See also

3 References

- https://en.wikipedia.org/wiki/Jaccard_index