Existing Metrics for Term-Similarity

- Information Content:
  - If G_c be set of molecules associated with concept c, then $IC(c) = - \log_2 (|G_c|/|G_r|)$
  - $\delta_R(c_i,c_j)$= max [$-\log_2$ (c)], c $\in$ A_i and c $\in$ A_j

    (c is common Ancestor)
  - Normalization:

    $\delta_L(c_i,c_j) = 2 * \delta_R(c_i,c_j) / (IC(c_i) + IC(c_j))$
  - Hybrid approach:

    $\delta_{JC}(c_i,c_j) = (1 - 2 * \delta_R(c_i,c_j) + IC(c_i) + IC(c_j))^{-1}$
- All three satisfy term-similarity properties