

# Analysis of Similarity Between 2 Tag Sets

## File information

**File 1:** ../json\_data/cc2\_comparison.json, 50 entries

**File 2:** ../json\_data/hz\_comparison.json, 50 entries

Matching entry titles: 50

## How similarity is calculated

Each token (word) in the text it is given a score from 0-1 unless neither annotator has tagged it, in which case it is ignored. The score is based on what fraction of the tags for that word match, regardless of linking.

**Note:** If words that are not tagged are counted as matches, %'s go up to 90-100% range.

## Overall

Total similarity: 45.47%

Best match: 564\_PMID30879166.txt.xml, 78.12%

## Analysis of each matching entry

Title: 570\_PMID29858716.txt.xml, Total Score: 13.0/30, Match Percentage: 43.33%

Title: 582\_PMID30778709.txt.xml, Total Score: 7.0/20, Match Percentage: 35.0%

Title: 573\_PMID30515612.txt.xml, Total Score: 26.0/39, Match Percentage: 66.67%

Title: 557\_PMID32632545.txt.xml, Total Score: 12.0/16, Match Percentage: 75.0%

Title: 556\_PMID29435687.txt.xml, Total Score: 7.0/10, Match Percentage: 70.0%

Title: 553\_PMID29236198.txt.xml, Total Score: 13.0/20, Match Percentage: 65.0%

Title: 586\_PMID31342239.txt.xml, Total Score: 26.0/47, Match Percentage: 55.32%

Title: 572\_PMID30430397.txt.xml, Total Score: 17.0/29, Match Percentage: 58.62%

Title: 590\_PMID26386572.txt.xml, Total Score: 6.0/14, Match Percentage: 42.86%

Title: 566\_PMID30421279.txt.xml, Total Score: 0/0, Match Percentage: -%

Title: 583\_PMID32409930.txt.xml, Total Score: 9.0/18, Match Percentage: 50.0%

Title: 576\_PMID30610505.txt.xml, Total Score: 10.0/28, Match Percentage: 35.71%

Title: 554\_PMID31489517.txt.xml, Total Score: 0.0/4, Match Percentage: 0.0%

Title: 559\_PMID30911960.txt.xml, Total Score: 7.0/20, Match Percentage: 35.0%

Title: 561\_PMID32737651.txt.xml, Total Score: 22.0/43, Match Percentage: 51.16%

Title: 562\_PMID31641960.txt.xml, Total Score: 2.0/28, Match Percentage: 7.14%

Title: 563\_PMid29978434.txt.xml, Total Score: 5.0/20, Match Percentage: 25.0%  
Title: 592\_PMid30426280.txt.xml, Total Score: 9.0/19, Match Percentage: 47.37%  
Title: 595\_PMid30578463.txt.xml, Total Score: 18.0/49, Match Percentage: 36.73%  
Title: 588\_PMid26801321.txt.xml, Total Score: 8.0/13, Match Percentage: 61.54%  
Title: 574\_PMid25953318.txt.xml, Total Score: 12.0/28, Match Percentage: 42.86%  
Title: 598\_PMid25218423.txt.xml, Total Score: 17.0/29, Match Percentage: 58.62%  
Title: 567\_PMid27770267.txt.xml, Total Score: 15.0/44, Match Percentage: 34.09%  
Title: 591\_PMid32901335.txt.xml, Total Score: 14.0/34, Match Percentage: 41.18%  
Title: 569\_PMid28012059.txt.xml, Total Score: 22.0/40, Match Percentage: 55.0%  
Title: 597\_PMid30680481.txt.xml, Total Score: 9.0/24, Match Percentage: 37.5%  
Title: 560\_PMid32591959.txt.xml, Total Score: 11.0/16, Match Percentage: 68.75%  
Title: 584\_PMid32026186.txt.xml, Total Score: 0/0, Match Percentage: -%  
Title: 571\_PMid31583496.txt.xml, Total Score: 0.0/50, Match Percentage: 0.0%  
Title: 599\_PMid32036474.txt.xml, Total Score: 24.0/36, Match Percentage: 66.67%  
Title: 575\_PMid28879567.txt.xml, Total Score: 20.0/30, Match Percentage: 66.67%  
Title: 568\_PMid32405891.txt.xml, Total Score: 27.0/35, Match Percentage: 77.14%  
Title: 577\_PMid31867678.txt.xml, Total Score: 1.0/8, Match Percentage: 12.5%  
Title: 589\_PMid31177396.txt.xml, Total Score: 6.0/21, Match Percentage: 28.57%  
Title: 565\_PMid31768842.txt.xml, Total Score: 12.0/17, Match Percentage: 70.59%  
Title: 552\_PMid29058102.txt.xml, Total Score: 4.0/9, Match Percentage: 44.44%  
Title: 581\_PMid32468177.txt.xml, Total Score: 13.0/18, Match Percentage: 72.22%  
Title: 564\_PMid30879166.txt.xml, Total Score: 25.0/32, Match Percentage: 78.12%  
Title: 578\_PMid31538267.txt.xml, Total Score: 15.0/24, Match Percentage: 62.5%  
Title: 551\_PMid31605257.txt.xml, Total Score: 9.0/14, Match Percentage: 64.29%  
Title: 558\_PMid31691131.txt.xml, Total Score: 3.0/19, Match Percentage: 15.79%  
Title: 587\_PMid28176146.txt.xml, Total Score: 9.0/27, Match Percentage: 33.33%  
Title: 600\_PMid26467923.txt.xml, Total Score: 30.0/48, Match Percentage: 62.5%  
Title: 555\_PMid33230593.txt.xml, Total Score: 7.0/23, Match Percentage: 30.43%  
Title: 594\_PMid28168327.txt.xml, Total Score: 10.0/25, Match Percentage: 40.0%  
Title: 580\_PMid30084053.txt.xml, Total Score: 13.0/17, Match Percentage: 76.47%  
Title: 585\_PMid30879165.txt.xml, Total Score: 5.0/10, Match Percentage: 50.0%  
Title: 593\_PMid32583319.txt.xml, Total Score: 0/0, Match Percentage: -%  
Title: 579\_PMid30788651.txt.xml, Total Score: 0.0/19, Match Percentage: 0.0%  
Title: 596\_PMid26456506.txt.xml, Total Score: 12.0/72, Match Percentage: 16.67%