

Rough Analysis of marked_and_linked_test.json

General Information

- Num. Entries: 22
- Tokenizer used: dmis-lab/biobert-v1.1

Definitions:

- Words represent the number of entities that have been considered distinct during the process of assigning tags to part of the text, it is essentially the text split on spaces
- Tokens are the actual pieces of data going into the model, punctuation is split from words and some large words are split down into smaller parts
- Total tags is the total number of tags across all words (If two or more words are part of the same tag this counts as a tag for each word)
- Unique tags is the total number of tags that are unique (not part of same tag)
- Max tags is the most tags a single word has
- Tagged words is the number of words that have at least one tag
- Tagged words % is the percentage of words that are tagged, it's given in brackets for individual entries.
- Avg tags is the average number of tags per word that is tagged
- MC words is the number of words with tags across multiple categories
- Total links is the total number of links across all tags (If two or more tags are linked this counts as a link for each one)
- Unique links is the total number of links that are unique (not part of the same link)
- Max links is the most amount of links a single tag or word has associated with it
- Linked tags, words is the number of tags or words that have links
- Linked % tags, words is the percentage of tags or words with links
- Avg links per tag, word is the average number of links per tag or word
- The schema used is BIES as it is all encompassing and models can convert to a lower resolution if wanted (Beginning, Inner, End, Singleton)

Sizes

- Average words: 207.5
- Maximum words: 301
- Minimum words: 145
- Average tokens: 380.7
- Maximum tokens: 598
- Minimum tokens: 265
- Entries with over 512 tokens: 1/22, 4.55%

Tags

Maximums

- Total tags: 85
- Unique tags: 28

- Max tags: 4
- Tagged words: 78
- Tagged words %: 39.2%
- Avg tags: 1.32
- MC words: 0

Averages

- Total tags: 30.77
- Unique tags: 10.55
- Max tags: 1.64
- Tagged words: 28.5
- Tagged words %: 14.11%
- Avg tags: 1.05
- MC words: 0.0

Links

Maximums

- Total links: 143
- Unique links: 14
- Max links per tag, word: 10, 10
- Linked tags, words: 63, 57
- Linked % tags, words: 100.0%, 28.64%
- Avg links per tag, word: 3.29, 4.31

Averages

- Total links: 26.05
- Unique links: 2.86
- Max links per tag, word: 2.27, 2.27
- Linked tags, words: 17.05, 15.68
- Linked % tags, words: 61.65%, 7.91%
- Avg links per tag, word: 1.34, 1.45

Schema

- Maximums (B, I, E, S): 21, 28, 21, 18
- Averages (B, I, E, S): 8.27, 9.09, 8.27, 5.14

Labels

- Label: Total, Maximum per entry, Average per entry

Categories

- Perturbing_Action: 277, 33, 12.59
- Context: 366, 62, 16.64
- Effect: 434, 80, 19.73

Description

- Knock Out: 51, 17, 2.32
- Conditional Knock Out: 11, 7, 0.5
- Knock In: 0, 0, 0.0
- Conditional Knock In: 0, 0, 0.0
- Knock Down: 35, 13, 1.59
- Plasmid Vector: 0, 0, 0.0

- Viral Vector: 8, 8, 0.36
- Pharmacological Inhibition: 51, 10, 2.32
- Decreased Expression Level: 32, 15, 1.45
- Increased Expression Level: 22, 6, 1.0
- Other: 67, 26, 3.05

Experiment_Type

- Organism: 34, 5, 1.55
- Tumour: 16, 8, 0.73
- Cells: 57, 14, 2.59
- Cell Line: 3, 3, 0.14
- Transformed Cell Line: 45, 11, 2.05
- Primary Culture: 0, 0, 0.0
- Organoid Culture: 0, 0, 0.0
- Cell Transplantation: 0, 0, 0.0
- Xenotransplantation: 28, 14, 1.27
- Other: 0, 0, 0.0
- Not Stated: 0, 0, 0.0

Species

- Human: 9, 5, 0.41
- Mouse: 42, 9, 1.91
- Rat: 0, 0, 0.0
- Fish: 0, 0, 0.0
- Fly: 0, 0, 0.0
- Yeast: 0, 0, 0.0
- Bacterium: 0, 0, 0.0
- Other: 0, 0, 0.0
- Not Stated: 132, 26, 6.0

Phenotype

- Adhesion: 0, 0, 0.0
- Apoptosis: 23, 6, 1.05
- Anoikis: 0, 0, 0.0
- Autophagy: 15, 11, 0.68
- Cell Cycle Arrest: 12, 7, 0.55
- Cell Death: 3, 3, 0.14
- Cell Growth: 22, 5, 1.0
- Cell Survival: 2, 2, 0.09
- Colony Formation: 0, 0, 0.0
- Differentiation: 4, 4, 0.18
- Entosis: 0, 0, 0.0
- Epithelial-Mesenchymal Transition: 8, 8, 0.36
- Ferroptosis: 17, 12, 0.77
- Invasion: 4, 3, 0.18
- Metastasis: 11, 5, 0.5
- Migration: 7, 4, 0.32
- Mitophagy: 12, 7, 0.55
- Necroptosis: 0, 0, 0.0
- Necrosis: 0, 0, 0.0
- Netosis: 0, 0, 0.0
- Oncosis: 0, 0, 0.0
- Proliferation: 12, 4, 0.55
- Pyroptosis: 0, 0, 0.0

- Quiescence: 0, 0, 0.0
- Self-Renewal: 0, 0, 0.0
- Senescence: 5, 5, 0.23
- Transformation: 2, 2, 0.09
- Tumour Growth: 6, 3, 0.27
- Tumourigenesis: 52, 18, 2.36

Activity

- Causes: 28, 8, 1.27
- Inhibits: 63, 22, 2.86
- Increases: 58, 13, 2.64
- Decreases: 50, 12, 2.27
- Regulates: 0, 0, 0.0
- No Effect: 15, 8, 0.68
- Not Stated: 3, 3, 0.14

Information on each entry

Title: 481_PMID31570854.txt_CC.xml, Words: 195, Tokens: 382

Tag data: Total: 12, Unique: 5, Max tags: 1, Tagged words: 12 (6%), Avg tags: 1.0, MC words: 0

Link data: Total: 16, Unique: 2, Max links (tag, word): (2, 2), Linked (tags, words): (10 (83.33%), 10 (5.13%))

Avg links (tag, word): (1.6, 1.6), Schema counts (B, I, E, S): (5, 2, 5, 0)

Title: 444_PMID32231246.txt_CC.xml, Words: 237, Tokens: 447

Tag data: Total: 27, Unique: 13, Max tags: 1, Tagged words: 27 (11%), Avg tags: 1.0, MC words: 0

Link data: Total: 17, Unique: 3, Max links (tag, word): (3, 3), Linked (tags, words): (12 (44.44%), 12 (5.06%))

Avg links (tag, word): (1.42, 1.42), Schema counts (B, I, E, S): (5, 7, 5, 10)

Title: 19_PMID31177901.txt_CC.xml, Words: 232, Tokens: 453

Tag data: Total: 25, Unique: 8, Max tags: 1, Tagged words: 25 (11%), Avg tags: 1.0, MC words: 0

Link data: Total: 28, Unique: 2, Max links (tag, word): (2, 2), Linked (tags, words): (16 (64.0%), 16 (6.9%))

Avg links (tag, word): (1.75, 1.75), Schema counts (B, I, E, S): (6, 9, 6, 4)

Title: 360_PMID33257688.txt_CC.xml, Words: 199, Tokens: 387

Tag data: Total: 85, Unique: 28, Max tags: 4, Tagged words: 78 (39%), Avg tags: 1.09, MC words: 0

Link data: Total: 143, Unique: 14, Max links (tag, word): (10, 10), Linked (tags, words): (63 (74.12%), 57 (28.64%))

Avg links (tag, word): (2.27, 2.51), Schema counts (B, I, E, S): (21, 25, 21, 18)

Title: 591_PMID32901335.txt_CC.xml, Words: 145, Tokens: 265

Tag data: Total: 31, Unique: 9, Max tags: 2, Tagged words: 26 (18%), Avg tags: 1.19, MC words: 0

Link data: Total: 69, Unique: 6, Max links (tag, word): (6, 6), Linked (tags, words): (21 (67.74%), 16 (11.03%))

Avg links (tag, word): (3.29, 4.31), Schema counts (B, I, E, S): (6, 14, 6, 5)

Title: 395_PMID33243998.txt_CC.xml, Words: 237, Tokens: 443

Tag data: Total: 58, Unique: 17, Max tags: 3, Tagged words: 49 (21%), Avg tags: 1.18, MC words: 0

Link data: Total: 9, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (9 (15.52%), 9 (3.8%))

Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (15, 21, 15, 7)

Title: 321_PMID33188176.txt_CC.xml, Words: 179, Tokens: 411

Tag data: Total: 21, Unique: 8, Max tags: 1, Tagged words: 21 (12%), Avg tags: 1.0, MC words: 0

Link data: Total: 8, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (8 (38.1%), 8 (4.47%))

Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (8, 4, 8, 1)

Title: 399_PMID33203836.txt_CC.xml, Words: 198, Tokens: 311

Tag data: Total: 30, Unique: 15, Max tags: 2, Tagged words: 29 (15%), Avg tags: 1.03, MC words: 0

Link data: Total: 16, Unique: 2, Max links (tag, word): (2, 2), Linked (tags, words): (11 (36.67%), 11 (5.56%))

Avg links (tag, word): (1.45, 1.45), Schema counts (B, I, E, S): (8, 4, 8, 10)

Title: 589_PMD31177396.txt_CC.xml, Words: 201, Tokens: 336
Tag data: Total: 14, Unique: 5, Max tags: 1, Tagged words: 14 (7%), Avg tags: 1.0, MC words: 0
Link data: Total: 10, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (10 (71.43%), 10 (4.98%))
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (5, 2, 5, 2)

Title: 154_PMD32747363.txt_CC.xml, Words: 249, Tokens: 436
Tag data: Total: 28, Unique: 10, Max tags: 2, Tagged words: 25 (10%), Avg tags: 1.12, MC words: 0
Link data: Total: 19, Unique: 2, Max links (tag, word): (2, 2), Linked (tags, words): (15 (53.57%), 12 (4.82%))
Avg links (tag, word): (1.27, 1.58), Schema counts (B, I, E, S): (9, 6, 9, 4)

Title: 31_PMD31066324.txt_CC.xml, Words: 260, Tokens: 487
Tag data: Total: 38, Unique: 12, Max tags: 1, Tagged words: 38 (15%), Avg tags: 1.0, MC words: 0
Link data: Total: 5, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (5 (13.16%), 5 (1.92%))
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (12, 13, 12, 1)

Title: 339_PMD33288741.txt_CC.xml, Words: 157, Tokens: 267
Tag data: Total: 6, Unique: 3, Max tags: 1, Tagged words: 6 (4%), Avg tags: 1.0, MC words: 0
Link data: Total: 6, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (6 (100.0%), 6 (3.82%))
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (2, 0, 2, 2)

Title: 132_PMD32366477.txt_CC.xml, Words: 183, Tokens: 314
Tag data: Total: 36, Unique: 10, Max tags: 1, Tagged words: 36 (20%), Avg tags: 1.0, MC words: 0
Link data: Total: 23, Unique: 2, Max links (tag, word): (1, 1), Linked (tags, words): (23 (63.89%), 23 (12.57%))
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (9, 14, 9, 4)

Title: 142_PMD33277366.txt_CC.xml, Words: 185, Tokens: 326
Tag data: Total: 9, Unique: 4, Max tags: 1, Tagged words: 9 (5%), Avg tags: 1.0, MC words: 0
Link data: Total: 8, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (8 (88.89%), 8 (4.32%))
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (3, 1, 3, 2)

Title: 146_PMD32193285.txt_CC.xml, Words: 215, Tokens: 422
Tag data: Total: 45, Unique: 11, Max tags: 1, Tagged words: 45 (21%), Avg tags: 1.0, MC words: 0
Link data: Total: 20, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (20 (44.44%), 20 (9.3%))
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (11, 20, 11, 3)

Title: 226_PMD33303641.txt_CC.xml, Words: 152, Tokens: 276
Tag data: Total: 75, Unique: 21, Max tags: 4, Tagged words: 57 (38%), Avg tags: 1.32, MC words: 0
Link data: Total: 60, Unique: 8, Max links (tag, word): (5, 5), Linked (tags, words): (43 (57.33%), 32 (21.05%))
Avg links (tag, word): (1.4, 1.88), Schema counts (B, I, E, S): (17, 28, 17, 13)

Title: 143_PMD330223944.txt_CC.xml, Words: 183, Tokens: 300
Tag data: Total: 9, Unique: 3, Max tags: 1, Tagged words: 9 (5%), Avg tags: 1.0, MC words: 0
Link data: Total: 9, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (9 (100.0%), 9 (4.92%))
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (3, 3, 3, 0)

Title: 417_PMD31320750.txt_CC.xml, Words: 273, Tokens: 487
Tag data: Total: 35, Unique: 14, Max tags: 2, Tagged words: 32 (12%), Avg tags: 1.09, MC words: 0
Link data: Total: 26, Unique: 3, Max links (tag, word): (2, 2), Linked (tags, words): (21 (60.0%), 18 (6.59%))
Avg links (tag, word): (1.24, 1.44), Schema counts (B, I, E, S): (11, 5, 11, 8)

Title: 400_PMD33093470.txt_CC.xml, Words: 301, Tokens: 598
Tag data: Total: 18, Unique: 9, Max tags: 1, Tagged words: 18 (6%), Avg tags: 1.0, MC words: 0
Link data: Total: 11, Unique: 2, Max links (tag, word): (1, 1), Linked (tags, words): (11 (61.11%), 11 (3.65%))
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (5, 2, 5, 6)

Title: 228_PMD28698299.txt_CC.xml, Words: 229, Tokens: 400
Tag data: Total: 34, Unique: 12, Max tags: 2, Tagged words: 32 (14%), Avg tags: 1.06, MC words: 0
Link data: Total: 36, Unique: 4, Max links (tag, word): (2, 2), Linked (tags, words): (29 (85.29%), 27 (11.79%))
Avg links (tag, word): (1.24, 1.33), Schema counts (B, I, E, S): (9, 11, 9, 5)

Title: 195_PMD32605998.txt_CC.xml, Words: 205, Tokens: 345
Tag data: Total: 17, Unique: 7, Max tags: 1, Tagged words: 17 (8%), Avg tags: 1.0, MC words: 0

Link data: Total: 25, Unique: 3, Max links (tag, word): (2, 2), Linked (tags, words): (17 (100.0%), 17 (8.29%))
Avg links (tag, word): (1.47, 1.47), Schema counts (B, I, E, S): (6, 2, 6, 3)

Title: 734_PMID32860742.txt_CC.xml, Words: 150, Tokens: 282

Tag data: Total: 24, Unique: 8, Max tags: 2, Tagged words: 22 (15%), Avg tags: 1.09, MC words: 0

Link data: Total: 9, Unique: 2, Max links (tag, word): (2, 2), Linked (tags, words): (8 (33.33%), 8 (5.33%))

Avg links (tag, word): (1.12, 1.12), Schema counts (B, I, E, S): (6, 7, 6, 5)