**DESCRIPTION**

Enron Email Dataset converted to tabular format: From, To, Subject, and Content. Some records labeled by CMU students.

**SUMMARY**

The Enron email dataset contains approximately 500,000 emails generated by employees of the Enron Corporation. It was obtained by the Federal Energy Regulatory Commission during its investigation of Enron's collapse.

This is the May 7, 2015 Version of dataset, as published at <https://www.cs.cmu.edu/~./enron/>

Those 1,700 records that were labeled by students are marked with labeled set were labelled by CMU students.

There are up to 12 categories per email:

* Cat\_[1-12]\_level\_1 = top-level category
* Cat\_[1-12]\_level\_2 = second-level category
* Cat\_[1-12]\_level\_weight = frequency with which this category was assigned to this message

Here are the categories:

* 1 Coarse genre
  + 1.1 Company Business, Strategy, etc. (elaborate in Section 3 [Topics]) (855 cnt.)
  + 1.2 Purely Personal (49 cnt.)
  + 1.3 Personal but in professional context (e.g., it was good working with you) (165 cnt.)
  + 1.4 Logistic Arrangements (meeting scheduling, technical support, etc) (533 cnt.)
  + 1.5 Employment arrangements (job seeking, hiring, recommendations, etc) (96 cnt.)
  + 1.6 Document editing/checking (collaboration) (176 cnt.)
  + 1.7 Empty message (due to missing attachment) (25 cnt.)
  + 1.8 Empty message (26 cnt.)
* 2 Included/forwarded information
  + 2.1 Includes new text in addition to forwarded material (680 cnt.)
  + 2.2 Forwarded email(s) including replies (913 cnt.)
  + 2.3 Business letter(s) / document(s) (42 cnt.)
  + 2.4 News article(s) (124 cnt.)
  + 2.5 Government / academic report(s) (15 cnt.)
  + 2.6 Government action(s) (such as results of a hearing, etc) (25 cnt.)
  + 2.7 Press release(s) (13 cnt.)
  + 2.8 Legal documents (complaints, lawsuits, advice) (22 cnt.)
  + 2.9 Pointers to url(s) (55 cnt.)
  + 2.10 Newsletters (18 cnt.)
  + 2.11 Jokes, humor (related to business) (7 cnt.)
  + 2.12 Jokes, humor (unrelated to business) (6 cnt.)
  + 2.13 Attachment(s) (assumed missing) (311 cnt.)
* 3 Primary topics (if coarse genre 1.1 is selected)
  + 3.1 regulations and regulators (includes price caps) (203 cnt.)
  + 3.2 internal projects -- progress and strategy (125 cnt.)
  + 3.3 company image -- current (71 cnt.)
  + 3.4 company image -- changing / influencing (63 cnt.)
  + 3.5 political influence / contributions / contacts (108 cnt.)
  + 3.6 california energy crisis / california politics (249 cnt.)
  + 3.7 internal company policy (83 cnt.)
  + 3.8 internal company operations (107 cnt.)
  + 3.9 alliances / partnerships (64 cnt.)
  + 3.10 legal advice (77 cnt.)
  + 3.11 talking points (26 cnt.)
  + 3.12 meeting minutes (33 cnt.)
  + 3.13 trip reports (7 cnt.)
* 4 Emotional tone (if not neutral)
  + 4.1 jubilation (12 cnt.)
  + 4.2 hope / anticipation (20 cnt.)
  + 4.3 humor (22 cnt.)
  + 4.4 camaraderie (20 cnt.)
  + 4.5 admiration (13 cnt.)
  + 4.6 gratitude (21 cnt.)
  + 4.7 friendship / affection (13 cnt.)
  + 4.8 sympathy / support (7 cnt.)
  + 4.9 sarcasm (18 cnt.)
  + 4.10 secrecy / confidentiality (130 cnt.)
  + 4.11 worry / anxiety (28 cnt.)
  + 4.12 concern (38 cnt.)
  + 4.13 competitiveness / aggressiveness (8 cnt.)
  + 4.14 triumph / gloating (3 cnt.)
  + 4.15 pride (3 cnt.)
  + 4.16 anger / agitation (10 cnt.)
  + 4.17 sadness / despair (2 cnt.)
  + 4.18 shame (1 cnt.)
  + 4.19 dislike / scorn (9 cnt.)