

## INDEX

By Tanja Bekhuis, PhD, TCB Research & Indexing LLC. TExtract® (v10 Pro) was used for semantic analysis and development.

This is a subject index for "The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions" by Patrick Juola, *Digital Scholarship in the Humanities*, Vol. 30, Supplement 1, 2015. Published by Oxford University Press on behalf of EADH: European Association for Digital Humanities. <https://doi.org/10.1093/llc/fqv040>.

- accuracy
  - assumption of, i109
  - authorship attribution and, i102
  - baseline, i105
  - chance and, i107
  - expected, i109–i110
- Ad-hoc Authorship Attribution Competition, i102
- authorship attribution
  - accuracy of, i105, i109
  - closed vs open class, i105
  - history of empirical testing, i101, i104
  - in the humanities, i100
  - implicit standards, i101
  - independence assumption, i109
  - interpretation of evidence, i102
  - number of authors, i105
  - standards, absence of explicit, i100
  - state of the art, i102
  - stylometry and, i100, i111
  - top candidate authors, i107–i108
  - See also protocol for authorship attribution
- average-rank test, formalization of, i108. See also statistics
- Baggins, Bilbo (pseudonym), i104–i106, i108–i109
  - asylum sought and granted, i104, i111
  - documents not in English, i109
  - See also US immigration courts
- baseline, analytic, i104
- Bitcoin, i104–i105
  - documents for case study, i104, i106, i109
  - protocol, i104
- blog posts, i106
- burden of proof, i110. See also evidence
- Burrow's Delta, i105
- case studies, i104–i105
- Casual Vacancy*, i105
- chance, i107–i109
- character 4-grams, i108. See also feature sets
- cherry-picking and credibility of analysis, i102–i103
- Cochrane Collaboration, standards for systematic reviews, i101
- Collins, Paul, i100, i108. See also Poe, Edgar Allan
- communities of practice, i102, i111
- corpus. See distractor corpus, ad hoc; documents; questioned documents (QD)
- courts (of law), i104–i105, i111
- cryptocurrency. See Bitcoin
- Cuckoo's Calling*, i104–i105, i110
- Daubert v. Merrell Dow Pharmaceuticals* (1993), i103
- Delta, Burrow's, i105
- digital humanities (DH)
  - community, i102, i104, i111
  - methods, i101
- digital libraries, i101
- digital scholarship, i101–i102
  - lack of explicit standards, i101
- distractor authors, i100, i106–i110
  - availability of, i106
  - number of, i109
- distractor corpus, ad hoc, i106–i107
- DNA, i110. See also evidence
- documents, i100–i102, i104–i106, i109
  - anonymous vs undisputed, i104
  - See also distractor corpus, ad hoc; questioned documents (QD)
- Dorian Satoshi Nakamoto (named by *Newsweek* as Bitcoin author), i104–i106, i109
- elimination tests, multiple independent, i100, i107. See also multistage filter
- errors
  - false acceptance or rejection, i107–i109
  - rate of, i103
- evaluation
  - competitions, i105
  - effect of feature sets, i107
  - TREC-style of, i102
- evidence
  - legal, in federal courts, i103
  - in medicine, i101
  - rules of, i102–i104
  - stylistic vs traditional, i110
- expert testimony, when admissible, i103
- false acceptance and rejection rates, i108–i109
- feature sets, i102, i106–i108
- Federal Rules of Evidence*, i103. See also evidence
- fingerprints, i110. See also evidence
- Forbes* (magazine), i104
- Frye v. United States* (1923), i103
- Galbraith, Robert (pen name for J. K. Rowling), i104, i108
  - Cuckoo's Calling*, i104–i105, i110
- ground truths, independent, i111
- Harry Potter* series (J. K. Rowling), i104
- immigration. See US immigration courts
- JGAAP (Java Graphical Authorship Attribution Program), i100, i106
- Juola, Patrick, i100–i102, i104–i106, i108–i109

King, Neal J., i106  
 as best match for Bitcoin author, i109

Koppel, Moshe, i100, i105–i107  
 harvesting documents, i106

Kuhn, Thomas S., i101, i111

language, i100, i102, i106, i109

multistage filter, i108. *See also* elimination tests,  
 multiple independent

nearest neighbors, i105–i106. *See also* statistics  
*Newsweek*, i104–i105  
*New York Times*, i106

Noecker, John, Jr., i106. *See also* Bitcoin  
*North American* (weekly newspaper), i100

PAN/CLEF, i105  
 conferences, i110  
 corpus of Dutch essays, i105

pass rate, by chance, i108–i109

Plagiarism Action Network/Conference and Labs of the  
 Evaluation Forum. *See* PAN/CLEF

Poe, Edgar Allan, i100, i104–i106, i108, i110  
 curated canon, i101  
 distractor authors for case study, i106, i110  
 documents for analysis, i109  
 "Dreams", i100  
 "The Happiest Day", i100  
 short stories, i104

Poe, Henry (brother of Edgar Allan Poe), i100, i106,  
 i108, i110

Poe, William Henry Leonard. *See* Poe, Henry

Popper, Karl R., and US Supreme Court, i103

practices  
 best, i101–i103, i106, i109 (*see also* evaluation)  
 good-but-not-best, i106

protocol for authorship attribution  
 acceptance in courts of law, i111  
 advantages and disadvantages, i100, i111  
 assumptions, i105–i106  
 case studies as subclass, i104  
 elements, i106–i109  
 formalization, i107–i108  
 potential flaw, i110  
*See also* authorship attribution

pseudoscholarship, i103

questioned documents (QD), i104–i105, i107–i108

random forests, i106. *See also* statistics

rank-order distributions, i108. *See also* statistics

reliability, i101–i104, i109

representativeness, i103

Rowling, J. K.  
 acknowledgment of authorship, i111  
*Casual Vacancy*, i105  
 corpus for case study, i106  
*Cuckoo's Calling*, i104–i105, i110  
 details of analysis, i108  
 distractor authors, i106  
*Harry Potter* series, i104  
 prepositions and verbs, i107

Rule 702 (in *Federal Rules of Evidence*), i103

Sagan, Carl, and extraordinary evidence, i110

sample size, i103

Satoshi Nakamoto and distractor authors for Bitcoin  
 case, i106, i109

scholars and scholarship, i101–i104, i110

science  
 as legal evidence, i101  
 maturation of, i101, i111  
 paradigms of, i101  
 as process, i103 (*see also* *Daubert v. Merrell Dow  
 Pharmaceuticals*)

Shakespeare, William  
*Hamlet*, i110  
 Marlowe as possible author of plays, i106

Siemens, Ray  
 Zampolli lecture, i102

standards  
 absence of, i100  
 adoption by other countries, i103  
 authorship attribution practice, i111  
 control of technique, i103  
 formalization of, i101  
 reliability of methods, i104  
 stakeholders interested in, i111

statistics  
 authorship attribution and, i101  
 significance, i108  
 style as evidence and, i110  
 stylometry and, i100  
 type of, i111

stylometry, i100, i110–i111  
*Sunday Times* (British national newspaper), i106

support vector machines, i106. *See also* statistics

text analysis, i100, i110

theory, i101, i103–i105

TREC (Text REtrieval Conference), i102

US immigration courts, i104–i105

US Supreme Court, i103

Victorian fiction and authorship studies, i108

words  
 distribution of length, i109  
 frequency, i106, i108  
 pairs of, i108  
 statistical methods and lack of independence, i106  
 in training samples of essays, i105