

PrecisionAndRecall

Precision and Recall

The traditional method of [MeasuringAccuracy](#) in the information-retrieval field is using a two-figure scheme of Precision and Recall.

- [a search-engine course](#)
- [Tim Bray comments](#)
- TODO: please add cites here if you've got 'em 🙄

Given the usual set of 4 numbers (see [FpFnPercentages](#)):

```
nspam = number of known-to-be-spam messages in the corpus
nham  = number of known-to-be-ham (nonsпам) messages in the corpus
fp    = number of ham messages incorrectly marked as spam
fn    = number of spam messages incorrectly marked as ham
```

Precision and Recall can be computed as follows:

```
nspamspam = nspam - fp

recall     = (nspamspam / nspam) * 100
precision  = ((nspamspam / (nspamspam + fn)) * 100
```

Again, Precision and Recall are part of the standard [SpamAssassin](#) statistics data reported in every release. The 'STATISTICS.txt' files distributed with [SpamAssassin](#) versions since about 2.30 include this data, measuring the ruleset's accuracy against a validation ruleset:

```
# SUMMARY for threshold 5.0:
# Correctly non-spam: 29443 99.97%
# Correctly spam:    27220 97.53%
# False positives:   9 0.03%
# False negatives:   688 2.47%
# TCR(l=50): 24.523726 SpamRecall: 97.535% SpamPrec: 99.967%
```

See also [MeasuringAccuracy](#) for other methods.