BattingAverage

Spam Filter Batting Average

John Graham-Cumming proposed this uniform measure of spam-filter effectiveness in his November 16, 2004 article entitled 'Understanding Spam Filter Accuracy'.

Essentially, it's a reformatting of the FalsePositive percentage and FalseNegative percentage, as 'spam hit rate / ham strike rate'. This can be computed from FP%/FN% as follows:

```
let fp = false positive percentage
let fn = false negative percentage
batting average hitrate = (1 - (fn / 100))
batting average strikerate = (fp / 100)
batting average = "hitrate/strikerate"
```

so if you have an FP% of 0.03%, and an FN% of 2.47%, the batting average is

```
(1 - (2.47 / 100)) "/" (0.03 / 100) =
.9753/.0003
```

That's actually the correct batting average for SpamAssassin 3.0.0's scoreset 3, measured against the validation corpus when we released it. 😉

See also MeasuringAccuracy for other schemes used, or FpFnPercentages for the main one we use in SpamAssassin.