Evaluation criteria:

1. Highest F-score for longest threshold
2. Highest break even point (BEP) of P&R
3. Highest Precision in general
4. Largest area under the curve (AUC) of P&R (same as in [1])
5. Always better to simply add words?

Evaluation:

- Best performance was achieved by filtering methods (case1 and case2)
- Seed word filtering increases performance in general
- But too much was no good
  (Only secondary filtering was better than Primary + Secondary)
- Single filtering was also more time efficient
- Simply adding more words does not increase performance

Automatic Acquisition of Seed Words

- Primary Filtering (cleaning)
  - Nitta's [1] seed words
  - Ishizaka's [2] seed words
  - Non-harmful words

- Secondary Filtering (optimizing)
  - Filtered seed words
  - Harmful word candidates

Evaluation Diagram

- McNemar test
  - winner

Possible reasons:

1. Page rankings change
2. Net-patrol movement
3. Usage policies tightening

Seed word candidates:

- case1: 7 words obtained after Primary Filtering with [2] 17 words
- case2: 12 words obtained after Primary Filtering with [1] 9 words
- case3: 16 words 7 words from case1 + original 9 words [1]
- case4: 21 words 12 words from case2 + original 9 words [1]
- case5: 5 words (baseline 1) originally used by [2]
- case6: 9 words (baseline 2) originally used by [1]

REFERENCES