University of Stuttgart
Germany

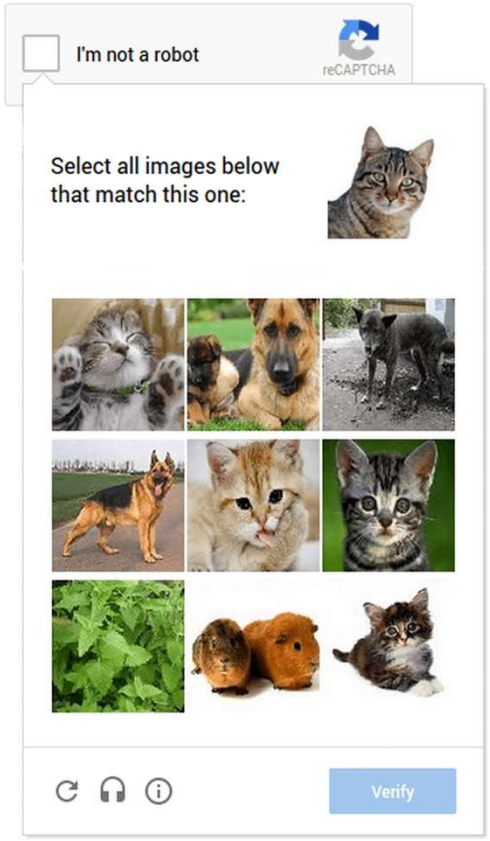
CAPTCHA mechanisms using semantic NLU tasks

Bachelor thesis

Marcel Wolkober

Everybody knows CAPTCHAs

Googleblog.com



Googleblog.com

CAPTCHA

- 2019: One-fourth of all internet traffic was made of malicious bots.
  → CAPTCHAs as one main countermeasure.
- Main CAPTCHA types:
  - Previously text-, image-, and sound-based
  - Now mostly behavior-based ones, with additional challenges
- Overall problem: advancement of these bots with the usage of artificial intelligence
  → Need for different kinds of challenges

Natural language understanding (NLU)

- Consists of hard to solve tasks.
- One such tasks is to rate the meaning similarity of a word in two contexts:

"**Banks** and credit-card firms are kept out of the picture."

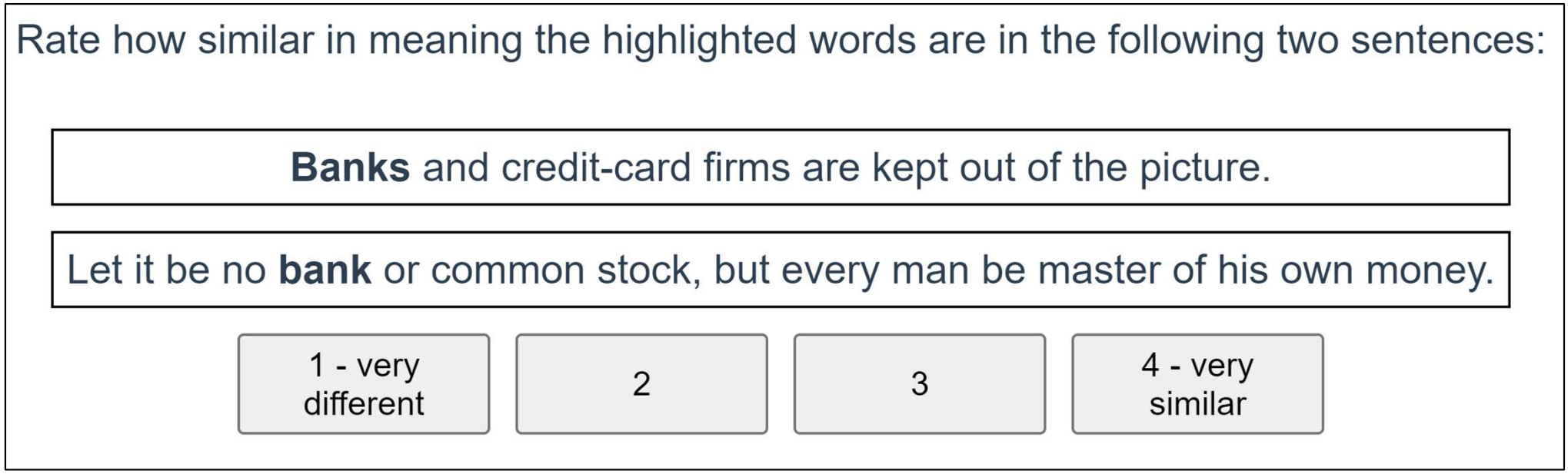
"Let it be no **bank** or common stock, but every man be master of his own money."

→ Same meaning

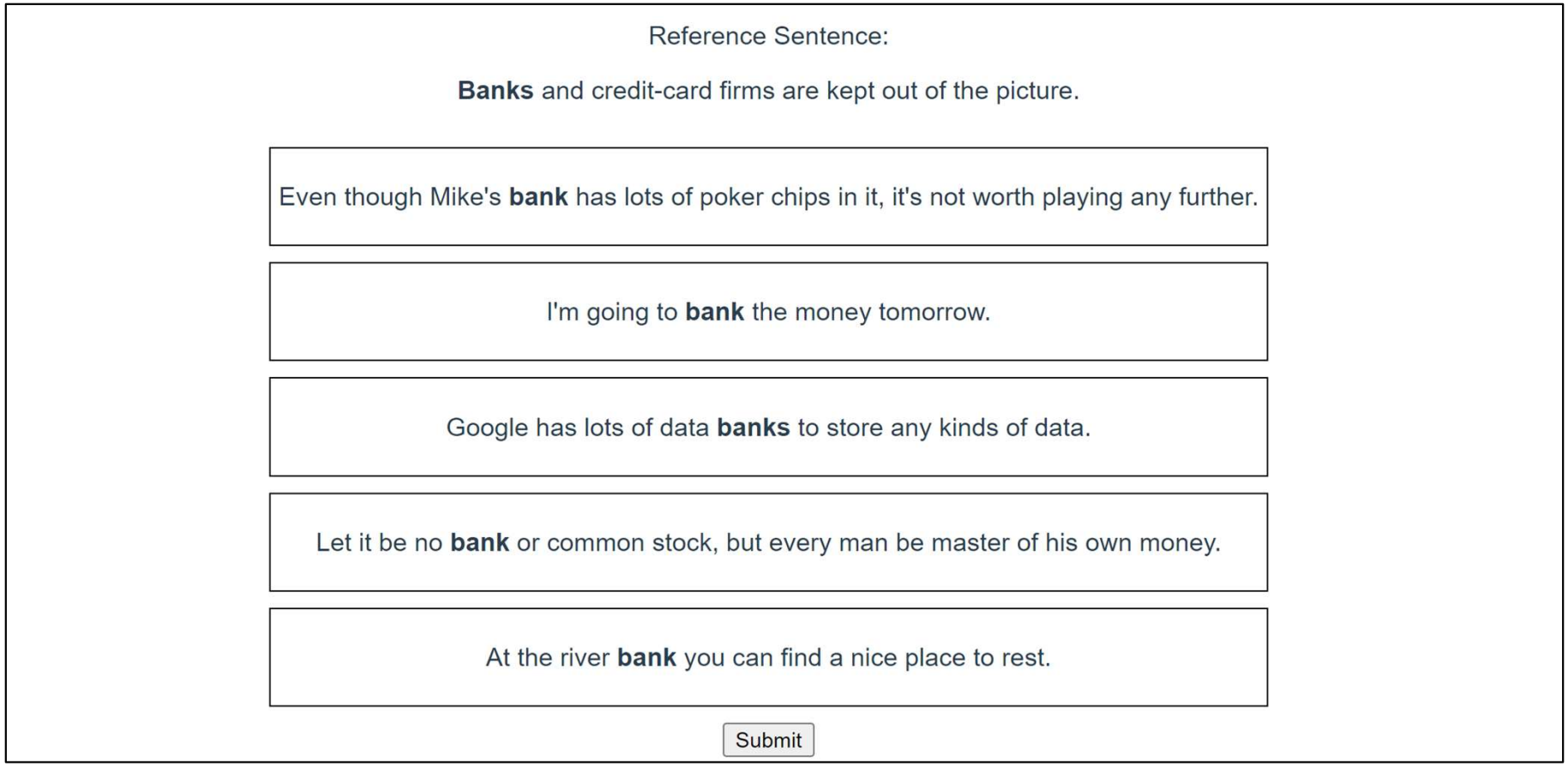
Devopedia.org

Pair challenge

Multiple of these usage pairs:



List challenge



Success criteria

- Pair challenge: Overall pairwise agreement of the chosen labels to the truth labels.

  →Krippendorff's Alpha coefficient

  Chosen: 1, 2, 2, 1, 4; Truth: 1, 3, 2, 2, 4; Krippendorff = 0.778

- List challenge: Order of the chosen list usages evaluated as their labels to the reference.

  →Spearman's rank correlation coefficient

  Chosen: 4, 2, 3, 3, 1; Truth: 4, 3, 3, 2, 1; Spearman = 0.808

→ Success rates will be looked at for different thresholds.

Study

- 8 different pair and list challenges with the same data
- 1 out of 16 challenges was asked to complete
- Feedback about the challenge was asked to be provided
- Collected:
  - Challenge results
  - Time to complete the challenge
  - Feedback

Results – time to complete

- 275 participants:

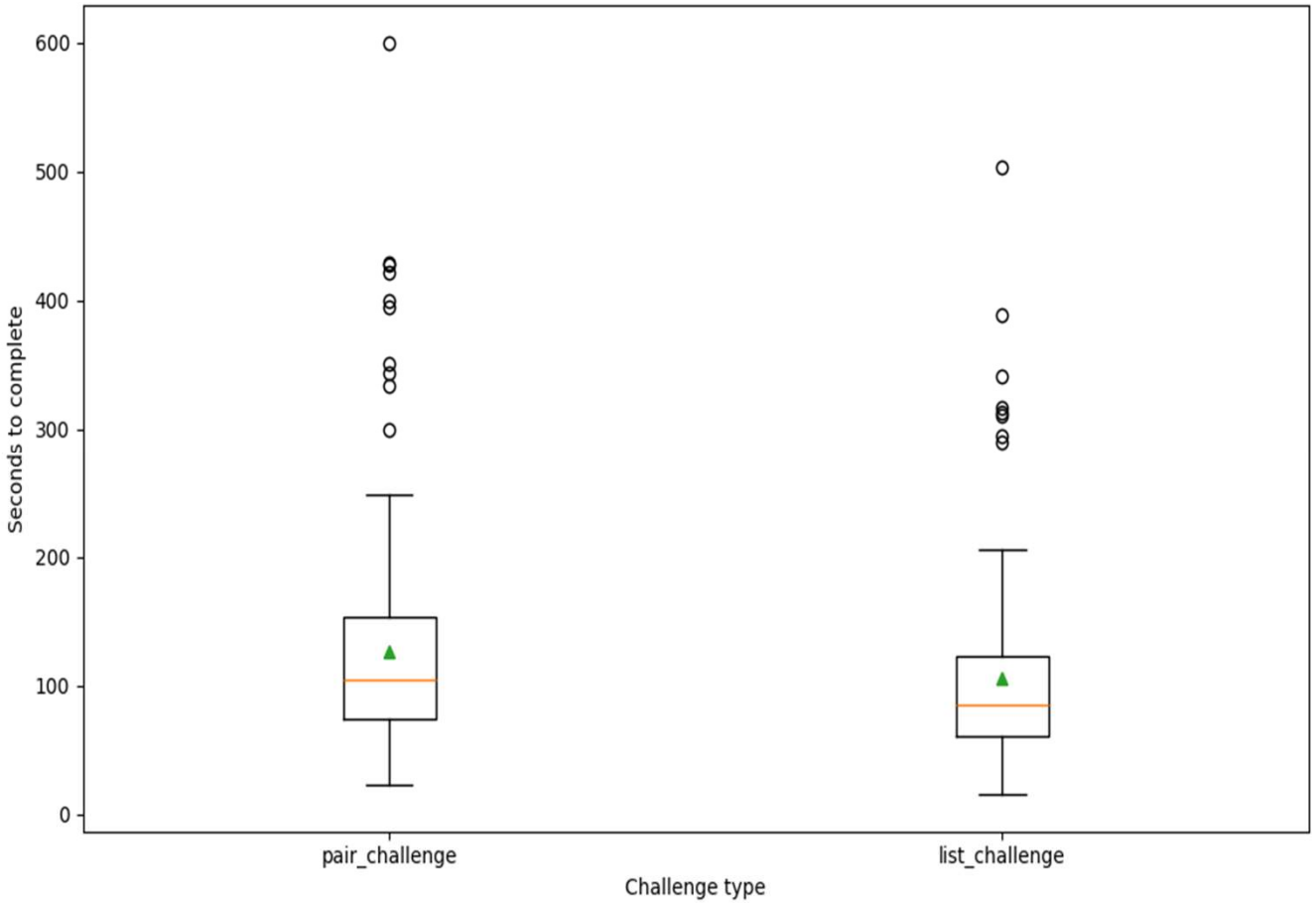
  On average 117 sec.

- 149 pair challenge:

  On average 126 sec.

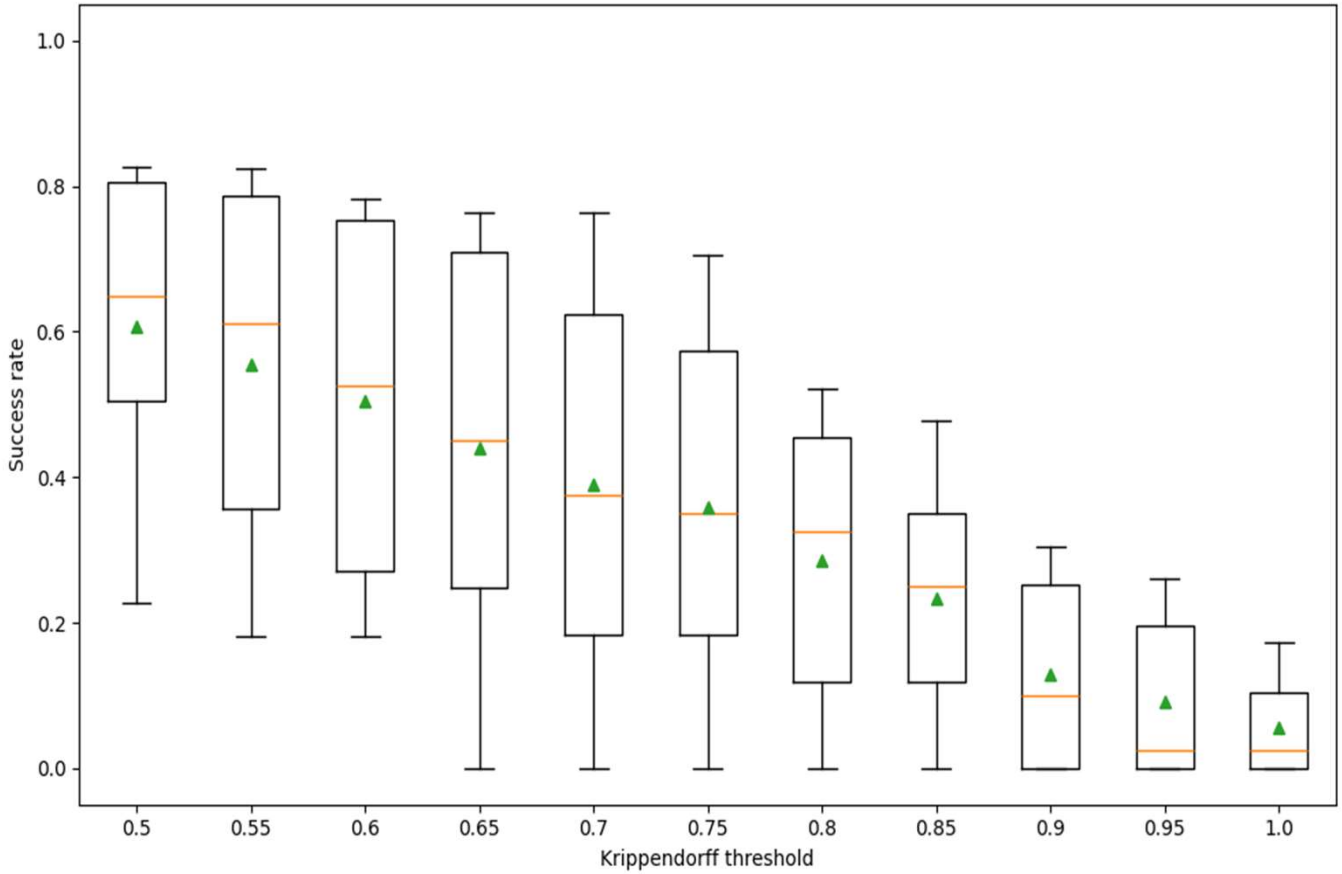
- 126 list challenge:

  On average 106 sec.



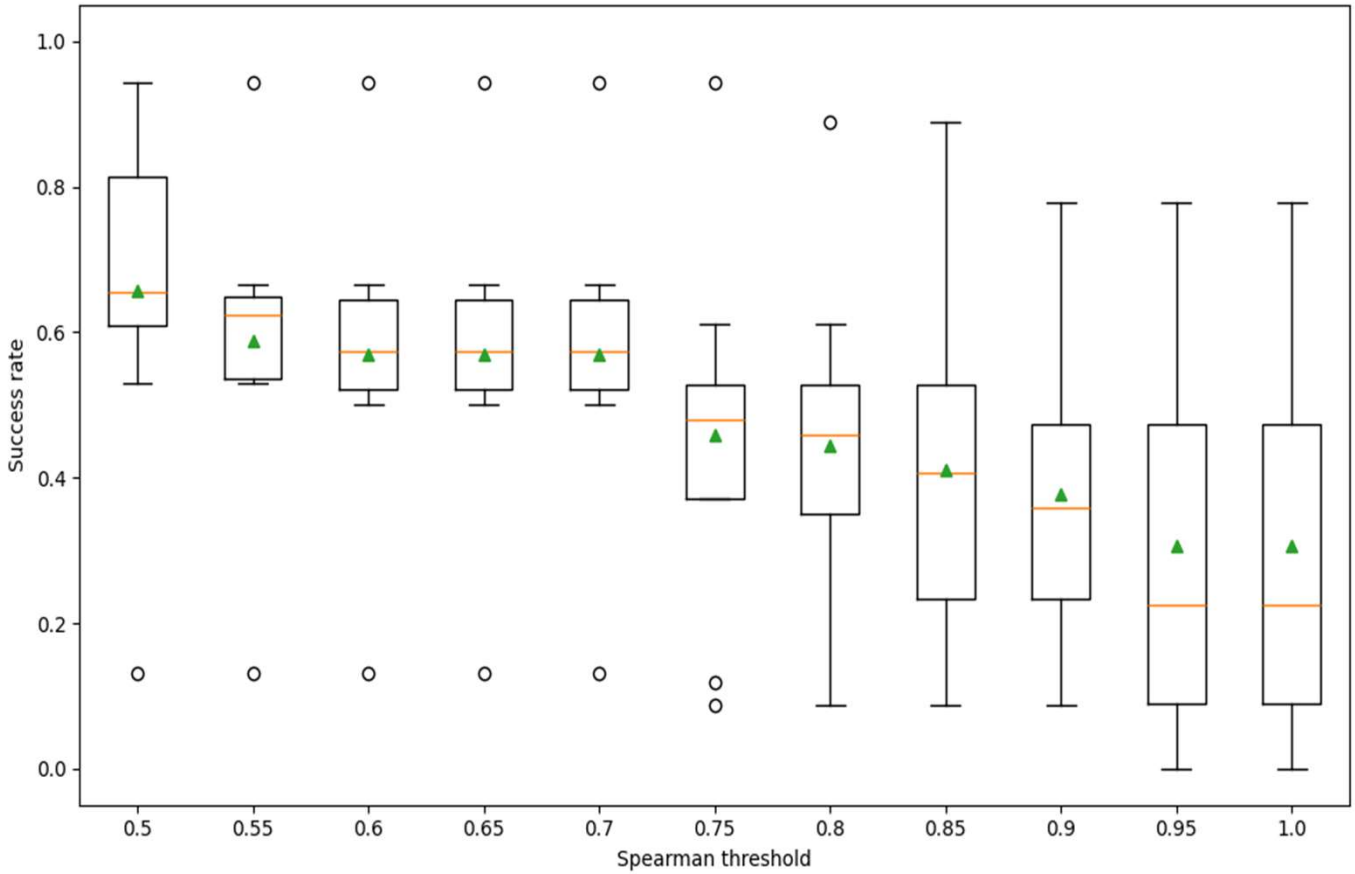
Results – pair challenges success rates

- Mean Krippendorff:
  0.53
- At threshold 0.5:
  Success rate of 62%
- At threshold 1.0:
  Success rate of 6%



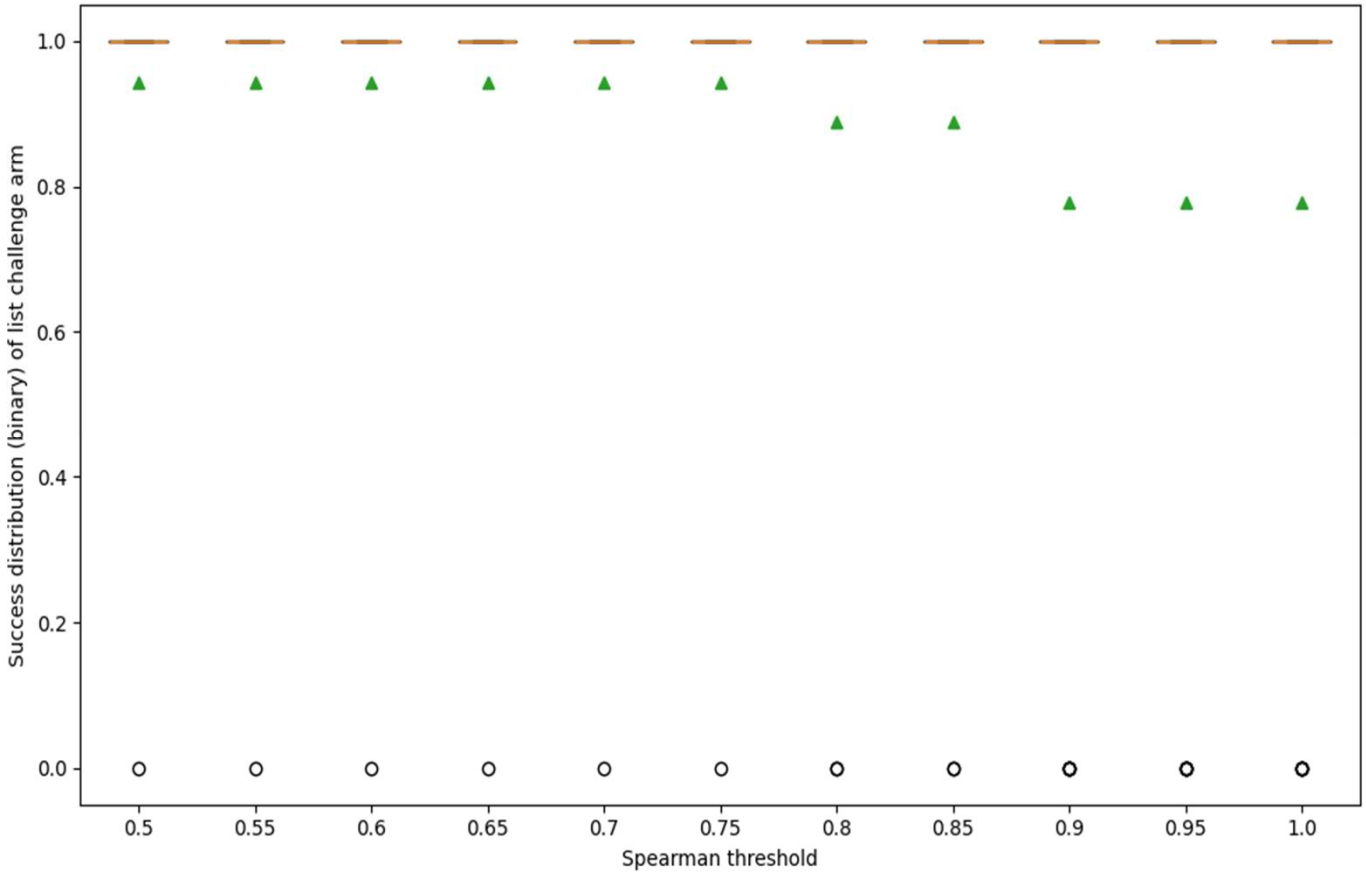
Results – list challenges success rates

- Mean Spearman:
  0.58
- At threshold 0.5:
  Success rate of 66%
- At threshold 1.0:
  Success rate of 32%



Results – list challenge *arm* success rates

- Mean Spearman:
  0.94
- At threshold 0.5:
  Success rate of 94%
- At threshold 1.0:
  Success rate of 78%



Results – participant feedback

- Most criticism in time and difficulty to complete:

  →Too much text to read

  →Tasks too complicated

- (Mostly) not preferred to other used CAPTCHAs:

  →A view participants liked either the pair or list challenge

Advanced attacker

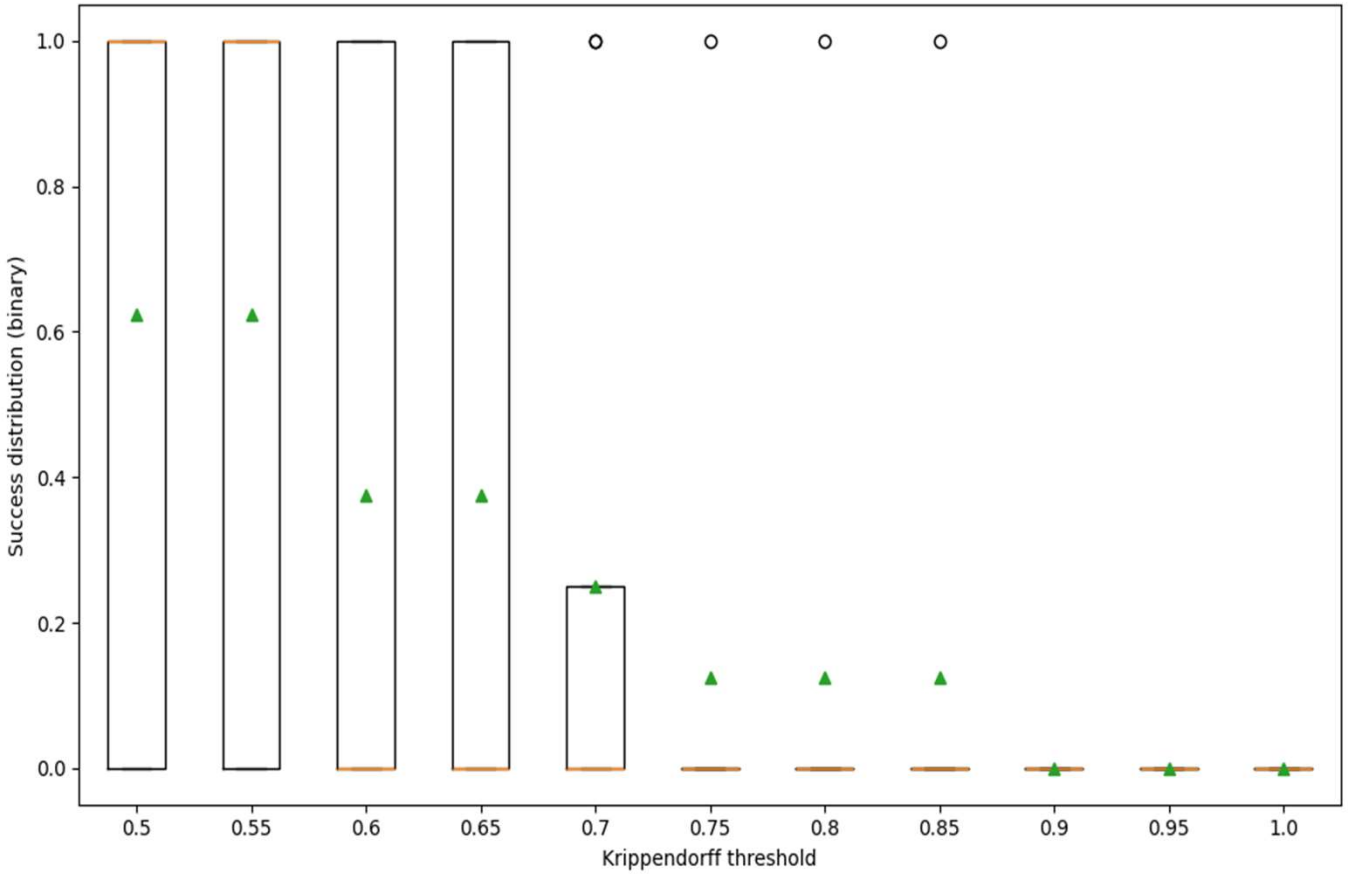
- Based on Word-in-Context pre-trained model XL-LEXEME.
  →Gives a cosine similarity for a target word in two contexts.

- Attacks pair challenges with a trained mapping of cosine similarity to label.

- Attacks list challenges by sorting the cosine similarity values.

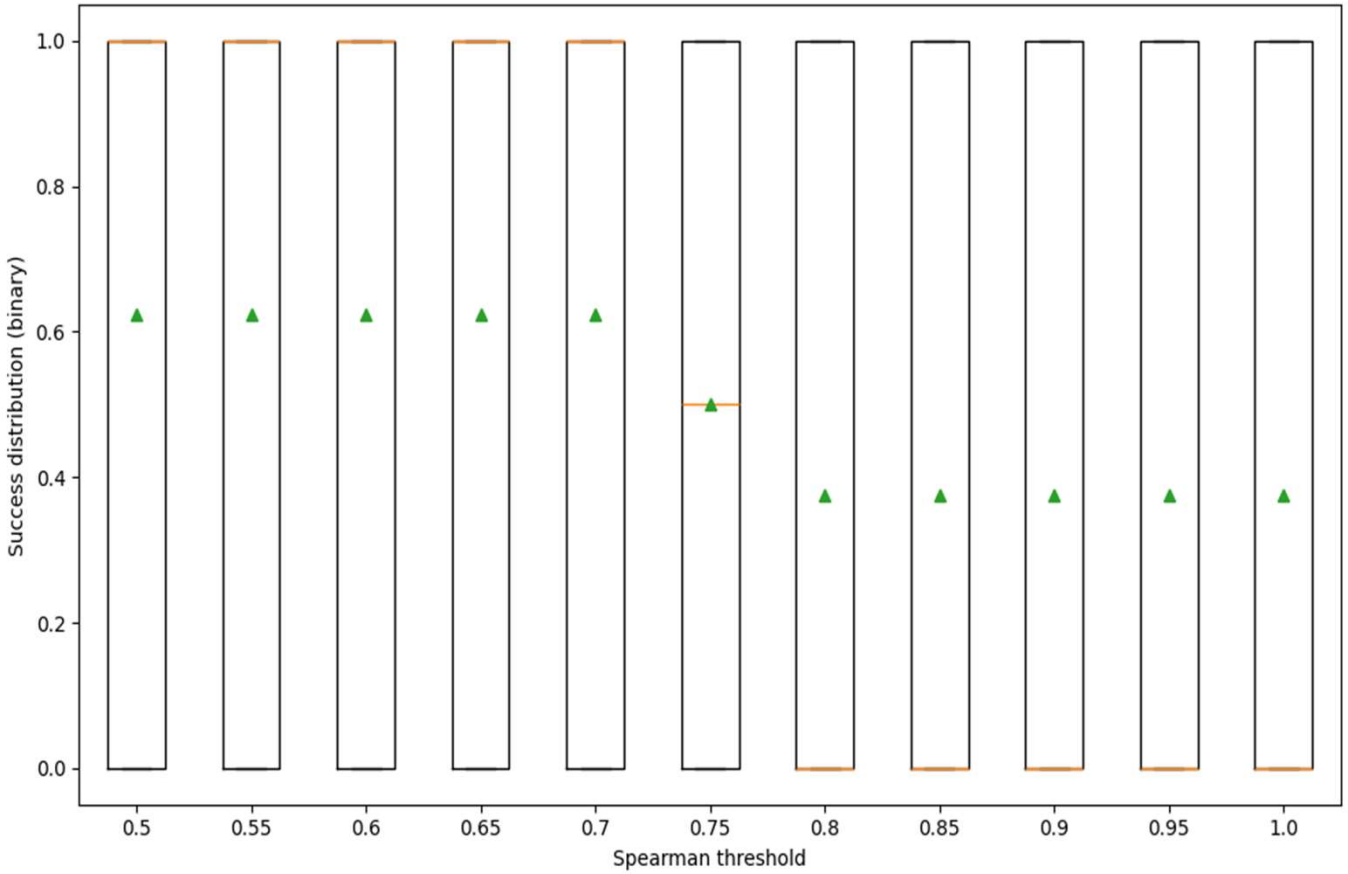
Attacker – pair challenges success rates

- Mean Krippendorff:
  0.44
- At threshold 0.5:
  Success rate of 63%
- Starting at threshold 0.9:
  Success rate of 0%



Attacker – list challenges success rates

- Mean Spearman:
  0.59
- At threshold 0.5:
  Success rate of 63%
- Starting at threshold 0.8:
  Success rate of 38%



Conclusion

- Overall human completion time and success rate do not improve in-use CAPTCHAs.

  →However, some individual challenge results show potential.

- Attacker’s success rate considerably too high considering a maximal success rate of 1%.

  →However, some challenges are attacker-proof and the challenge pool may not be large enough.

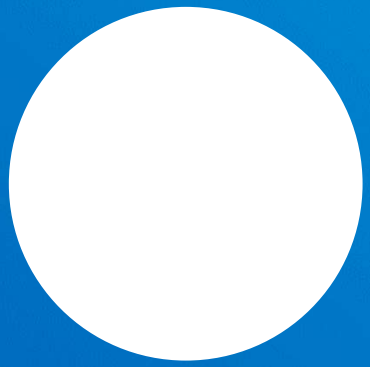
Future work

- Improvement of challenges by introducing human behavior analysis.
  →Possibly less complicated for humans, while increasing security.
- Finding / Creating more suitable data.

- Need for advanced attacker proof challenges remains.
  →Maybe even more than expected.
- The research done in this thesis might provide a basis for further challenges using tasks of NLU.

University of Stuttgart
Germany

Thank you!



Marcel Wolkober

e-mail st163937@stud.uni-stuttgart.de

Image sources

- https://security.googleblog.com/2014/12/are-you-robot-introducing-no-captcha.html
- Waldron, Mike. "Structured vs Unstructured Data: Exploring an Untapped Data Reserve." *AYLIEN, April* 15 (2015). Of website: https://devopedia.org/natural-language-understanding