Netanos - Named entity-based Text Anonymization for Open Science

Bennett Kleinberg[1,†] and Maximilian Mozes[2,†]

[1]University of Amsterdam, The Netherlands
[2]Technical University of Munich, Germany
[†]Both authors contributed equally to the development of this tool and are listed in alphabetical order.

June 4, 2017

Paper DOI: (later)
Software Repository: https://github.com/ben-aaron188/netanos
Software Archive: (later)

Summary

Netanos (Named Entity-based Text Anonymization for Open Science) is a natural language processing software that anonymizes texts by identifying and replacing named entities. The key feature of NETANOS is that the anonymization preserves critical context that allows for secondary linguistic analyses on anonymized texts.
Consider the example string "Max and Ben spent more than 1000 hours on writing the software. They started in August 2016 in Amsterdam." While coarse anonymization such as simple "XXX" replacement would suffice to mask the true content of the string, essential text properties are lost that are needed for secondary analyses. For example, content-based deception detection approaches rely on the number of specific times and dates to differentiate between deceptive and truthful texts [5].
The architecture of NETANOS relies on two software libraries capable of identifying named entities. (1) The Stanford Named Entity Recognizer [1] integrated with the ner Node.js package [3], and (2) the NLP-compromise JavaScript frontend-library [2]. Both libraries are used in a layered architecture to identify persons (e.g. "Max", "Ben"), locations (e.g. "Amsterdam", "Munich"), organizations (e.g. "Google"), dates (e.g. "August 2016"), and values (e.g. "42").
Specifically, the text anonymization is achieved with the following stepwise procedure: The input string is analyzed by Stanford's NER, identifying organizations, locations, persons, and dates. All identified entities are replaced with their context-preserving anonymized versions. NLP-compromise's named entity recognition tool is applied to identify potentially remaining, unrecognized entities.

Besides the key feature of context preserving text anonymization, Netanos also provides three alternative anonymization types.

- **Context-preserving replacement** (key feature)
  Identified named entity types are replaced with a composite string consisting of the entity type and the corresponding index of occurrence. "[PERSON_1] and [PERSON_2] spent more than [DATE/TIME_1] on writing the software. They started in [DATE/TIME_2] in [LOCATION_1]."

- **Named entity-based replacement**
  Identified entities are replaced with a different, randomly chosen named entity of the same

type. ”Barry and Rick spent more than 997 hours on writing the software. They started in January 14 2016 in Odessa.”

- **Non-context preserving replacement**
  This replacement type is inspired by the anonymization procedure suggested by the UK Data Service [4]. It replaces all strings having a capital first letter and all numeric values with XXX. ”XXX and XXX spent more than XXX hours on writing the software. XXX started in XXX XXX in XXX.”

- **Combined, non-context preserving anonymization**
  The context-preserving replacement is used to identify candidates for replacement that are then replaced with the procedure of the non-context preserving replacement ”XXX and XXX spent more than XXX XXX on writing the software. XXX started in XXX XXX in XXX.”

Note that all replacements are applied globally across the input string.

Technical Pipeline

The software architecture of NETANOS is illustrated in the following technical pipeline on FigShare: [LINK TO FIGSHARE](LINK HERE!!!)

Note

The software documentation for NETANOS with working examples and installation guidelines is available **here**.
The NETANOS tool has been experimentally validated on the potential re-identifiability of anonymized texts. A preprint to that paper is available on the **Open Science Framework preprint server**.

References

[1] Finkel, J. R., Grenager, T. and Manning, C. (2005, June). *Incorporating non-local information into information extraction systems by gibbs sampling*. In Proceedings of the 43rd annual meeting on association for computational linguistics (pp. 363-370). Association for Computational Linguistics.

[2] Kelly, S. (2016), *NLP Compromise: Natural language processing in javascript*, GitHub repository, https://github.com/nlpcompromise/compromise.

[3] Srivastava, N. (2016), *ner: Client for Stanford Named Entity Reconginiton*, GitHub repository, https://github.com/niksrc/ner.

[4] UK Data Service (no date) *ukds.tools.textAnonHelper / Home [BitBucket Wiki]*. Retrieved February 25, 2017, from https://bitbucket.org/ukda/ukds.tools.textanonhelper/wiki/Home.

[5] Warmelink, L., Vrij, A., Mann, S., and Granhag, P. A. (2013). *Spatial and temporal details in intentions: A cue to detecting deception.* Applied Cognitive Psychology, 27(1), 101-106.