An Exploratory Study on Vagueness in Website Privacy Policies

Nicole Fella
Department of Electrical & Computer Engineering
Manhattan College
nicolelfella@gmail.com

Kexin Liao
Department of Computer Science
University of Central Florida
ericaryo@knights.ucf.edu

Dr. Fei Liu
Department of Computer Science
University of Central Florida
feiliu@cs.ucf.edu

Abstract

Website privacy policies are often overlooked due to their length and difficulty to understand. The goal of this research is to create a framework to help define and identify characteristics of vagueness within a text document. Legal perspectives in relation to the "void for vagueness" doctrine and linguistic perspectives regarding comprehension were investigated to create this framework. Further analysis can lead to the use of machine learning and natural language processing to aid in identifying vagueness within these documents.

Motivation

"Void for Vagueness" Doctrine
- States that a law or regulation must be clearly specified for the average person to understand
- If a law or regulation does not satisfy this requirement, statutes can be considered unconstitutionally vague and unenforceable
- Existence of this doctrine implies vagueness can be reduced [11,17]

Why Website Privacy Policies?
- Privacy policies contain important information outlining the agreement between the user and the company
- These policies are often too long or too difficult to understand
- Policies are often overlooked and ignored
- Casual users have greater difficulty understanding policies than experts [2,14,16]

Vagueness is subjective [1], making it difficult to quantify vagueness and create an effective way of classification. Because privacy policies are not easy to understand, natural language processing becomes a relevant approach [13].

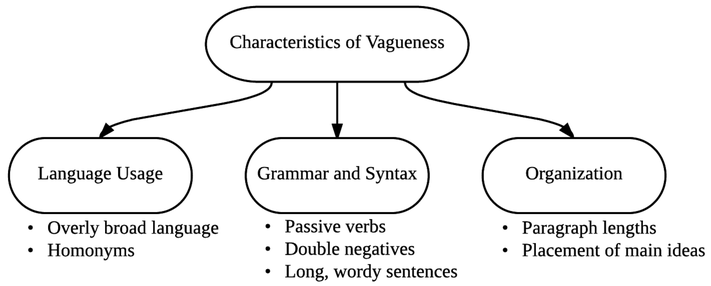
This study hopes to better define vagueness, identify clear characteristics, and establish where it can arise. With a better understanding of vagueness, there is potential to better understand natural language.

Investigation of Vagueness

Indeterminacy relates and describes cases involving vagueness, ambiguity and contestability [17]. While the three cases are distinct and create specific situations, indeterminacy and debate are involved [8,9,12].

Vagueness is introduced in borderline cases which require further interpretation and examination [1,6]. These borderline situations arise when there are not clearly specified conditions. Vagueness can arise when there are no specific words or terminology to best describe concepts [5].

Linguistic characteristics can be used to describe text vagueness [4,10]



Some terms associated with vagueness include: "may," "except," "aside from" [10]. The table below also includes an additional 40 vague key words and phrases [15].

Results from Applying Taxonomy to Privacy Policies

Category	Key Words and Phrases	Distribution*
Condition	depending, necessary, appropriate, inappropriate, as needed, as applicable, otherwise reasonably, sometimes, from time to time	7.20%
Generalization	generally, mostly, widely, general, commonly, usually, normally, typically, largely, often, primarily, among other things	3.63%
Modality (including modal verbs)	may, might, can, could, would, likely, possible, possibly	70.60%
Numeric quantifier	anyone, certain, everyone, numerous, some, most, few, much, many, various, including but not limited to	18.60%

Table 1: Table adopted from (Reidenberg et. al. 2016). It includes 40 vague key words and phrases divided, manually identified by experts, and organized into established categories.

Key Terms

Privacy policy excerpts that were marked "unclear" by Amazon Mechanical Turkers were extracted from the database in "Extracting Key Practices from Website Privacy Policies" [18].

Below is an example of an excerpt with highlighted vague expressions manually identified to be added to the list of 40 vague terms previously defined.

> We may match information collected from you through different means or at different times including both personal information and Service usage information and use such information along with offline and online information obtained from other sources (including third parties) such as demographic information and updated contact information (where that information has been lawfully disclosed to us) for the purposes of learning more about you so we can provide you with relevant content

Future Work

This study creates a framework to identify vagueness within text documents. To further this research, we hope to create a machine learning algorithm to classify vagueness.

Implementing a classifier, potentially Multinomial Naïve Bayes, Logistic Regression, or Support Vector Machine, will help classify a text as vague or not vague. Introducing the idea of saliency, the amount a word contributes to the final composed meaning, this classifier can focus on words with the highest significance [3,7].

Acknowledgements

The support for this work was provided by the National Science Foundation REU program under Award No. 1560302. Any opinions, findings, and conclusions and recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Thank you Dr. Liu, Dr. Turgut and Kexin for your constant support and guidance.

References

1. Alexopoulos, P., Pavlopoulos, J. 2014. *A Vague Sense Classifier for Detecting Vague Definitions in Ontologies*. EACL
2. Cranor, L., Guduru, P., Arjula, M. 2006. *User Information for Privacy Agents*. ACM Transactions on Computer-Human Interaction (TOCHI), Vol. 13, Issue 2
3. Deemter, K. 2000. *Generating Vague Descriptions*. Proceedings of the first international conference on Natural language generation, Vol 14
4. Hernacki, A. 2012. *A Vague Law in a Smartphone World: Limiting the Scope of Unauthorized access under the computer fraud and Abuse Act*. American University Law Review, Vol. 61, Issue 5.
5. Hunt, L. 2015. *What the Epistemic Account of Vagueness Means for Legal Interpretation*. Law and Philosophy, Vol. 35, Issue 1.
6. Jonsson, OP. 2009. *Vagueness, Interpretation and the Law*. Cambridge University Press, Vol. 15, Issue 3
7. Li, J., Chen, X., Hovy, E., Jurafsky, D. 2015. *Visualizing and Understanding Neural Models in NLP*. CoRR
8. Liebwald, D. 2013. *Law's Capacity for Vagueness*. International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique
9. Low, P., Johnson, J. 2015. *Changing the Vocabulary of the Vagueness Doctrine*. Virginia Law Review, Vol 101, No. 8.
10. Micheti, A., Burkell, J., Steeves, V. 2010. *Fixing Broken Doors: Strategies for Drafting Privacy Policies Young People Can Understand*. Bulletin of Science, Technology & Society, Vol. 30, Issue 2.
11. Post, R. 1994. *Reconceptualizing Vagueness: Legal Rules and Social Orders*. California Law Review, Vol. 82, Issue 3
12. Raffman, D. 2015. *Precis of Unruly Words: A Study of Vague Language*. Philosophy and Phenomenological Research, Vol. 90, Issue 2
13. Ramanath, R., Liu, F., Sadeh, N., Smith, N. 2014.*A Step Towards Usable Privacy Policy: Automatic Alignment of Privacy Statements.* In Proceedings of the 25th International Conference on Computational Linguistics
14. Reidenberg, J., Breaux, T., Carnor, L., French, B. 2014. *Disagreeable Privacy Policy: Mismatches Between Meaning and Users' Understanding*. Berkeley Technology Law Journal, Vol 30, Issue 1.
15. Reidenberg. J., Bhatia, J., Breaux, T., Norton, T. 2016. *Ambiguity in Privacy Policies and the Impact of Regulation.* Journal of Legal Studies
16. Vail, M., Earp, J., AntOn, A. 2008. *An Empirical Study of Consumer Perceptions and Comprehension of Web Site Privacy Policies.* IEEE Transactions on Engineering Management , Vol 55, Issue 3
17. Waldron, J. 1994. *Vagueness in Law and language: Some Philosophical Issues.* California Law Review, Vol. 82, Issue 3
18. http://www.nlp.cs.ucf.edu/research/