Question Answering for Privacy Policies: Combining Computational and Legal Perspectives

Abhilasha Ravichander¹, Alan W Black¹, Shomir Wilson³, Thomas Norton², Norman Sadeh¹

Motivation

• Privacy policies are the documents which disclose the ways in which a company gathers, uses, shares and manages a user’s data.
• Notice and Choice (FTC, 1998).
• Long, complex, difficult for users to read and understand.
• We would like to empower users to inform themselves about the issues that matter to them, and enable them to selectively explore these issues.
• We present PRIVACYQA, a corpus consisting of 1750 questions about the privacy policies of mobile applications, and over 3500 expert annotations of relevant answers.

Question Elicitation: Crowdworkers

Answer Construction: Legal Experts

Experimental Results

Dataset Statistics:
1750 Questions, 35 policies,
Average Length of Question: 8.4 words
Average Length of Document: 3237 words
Avg Length of answers : 139.6 words

Baselines:

Unanswerability

PrivacyQA
• Incomprehensibility, any difficulties to occupy the privacy assistant
• Relevance, Will the app consume much space
• Ill-formedness, How do I know the app is legit
• Silence, Is my data secure,
• Atypicality, Has this app had a data breach?

Natural Questions (Kwiatkowski et. al., 2019)
• Unresolvable coreference, How do i get to Missouri from here
• Answer does not exist, what guides Santa home after he has delivered presents?
• Queries are not questions, all I want for christmas is you mariah carey tour
• False presuppositions, what is the only fruit that does not have seeds

Baselines:

Code/Data: https://github.com/AbhilashaRavichander/PrivacyQA_EMNLP

Website: http://usableprivacy.org/
Email: aravicha@cs.cmu.edu