Role of Ontological Semantics in Handling Privacy Policies
Olga Krachina

Understanding privacy policy (PP) is a key to prevent unsolicited marketing and disclosure of personal information. PP is written in natural language, hence need for a tool to convert natural language into formal machine language.

Current solutions
- disable cookies
- P3P

Limitations
- restricted transactions
- no validation of one’s policy in machine-language format is consistent with original;
- excludes web-sites that while don’t comply, may have privacy practices exceeding those of P3P;
- lacks means to enforce privacy policies;

CyberTrust Project aims at creating an innovative system with following results:
- An expressive language for specifying PP that has an intuitive and precise semantics based on a rich ontological resource;
- An advanced framework for authoring, enforcing and auditing PP;
- Tools to empower users with control of their PP through user-friendly and ontology-based interfaces;
- Tools for evaluating today’s privacy practices;

(CyberTrust proposal, 2004)
NLP: Ontological Semantics Approach

- ontology is the constructed-reality conceptual hierarchy of the domain, relating all the processes, objects and properties in it;
- the lexicon contains all the words and phrases of the domain, with their meanings defined in ontological terms;
- the processed sentence is expressed as a text meaning representation (TMR) in the formal ontology-based TMR knowledge representation language;
- analyzer takes the input sentence to its TMR, while the generator reverses the process; (CyberTrust proposal, 2004)

PP: Outline of Ontological Mapping

First Party (company)
  information-collect
    customer-provided
      personal information
    customer-not-provided
      click-stream information
  information-use
    provide-service
    disclose
      third party
        marketing
          solicit
          marketing
    web-site
      advertise
    user-interaction
      chat-room
      forum

Concept

(INFORMATION-SECURITY-ATTACK
  (DEFINITION (VALUE "the attempt to obtain, alter or erase information")))

(IS-A (VALUE (COMMUNICATIVE-EVENT CRIMINAL-ACTIVITY)))

(SUBCALESSES (VALUE
  COMMUNICATION-OBSTRUCT
  INFORMATION-ERASE
  INFORMATION MODIFY
  INFORMATION OBTAIN))

(AGENT (SEM INFORMATION-SECURITY-ATTACKER))

(BENEFICIARY (SEM HUMAN))

(INSTRUMENT (SEM COMMUNICATION-DEVICE NATURAL LANGUAGE))

(THEME (SEM INFORMATION))

(LEGALITY-ATTRIBUTE (VALUE NO))

(OPPOSITE (SEM SOCIAL-EVENT)))

Lexical Entry

(INTERFACE
  (INTERFACE-N1 (CAT N)
    (ANNO (DEF "point of connection between two systems, networks, or devices")
      (EX ")" (COMMENTS ""))
    (SYN-STRUC ((N ((ROOT $VAR1) (CAT N) (POSSESSIVE +) (OPT +)))
        (ROOT $VAR0) (CAT N)
        (PP-ADJUNCT ((ROOT WITH) (ROOT $VAR2) (CAT PREP)
            (OPT +) (OBJ ((ROOT $VAR3) (CAT N)))))))
    (SEM-STRUC (RELATION (DOMAIN (VALUE ^$VAR1))
        (RANGE (VALUE ^$VAR3))
        (^$VAR2 (NULL-SEM+ ))))))

(INTERFACE-V1 (CAT V)
  (ANNO (DEF "connect two otherwise possibly not communicating devices")
    (EX ")" (COMMENTS ""))
  (SYN-STRUC ((ROOT $VAR0) (CAT V)
      (SUBJECT ((ROOT $VAR1) (NP CASE NOMINATIVE))
        (DIRECTOBJECT ((ROOT $VAR2) (CAT N)
            (NP CASE ACCUSATIVE))
          (PP-ADJUNCT ((ROOT WITH) (CAT PREP)
              (OBJ ((ROOT $VAR3))))))))
  (SEM-STRUC (CONNECTS (DOMAIN (VALUE ^$VAR1))
      (RANGE (VALUE ^$VAR2)))))