Modeling Natural Language Policies into Controlled Natural Language: A Twitter Case Study

Special Session along with ICIMP 2020
The Fifteenth International Conference on
Internet Monitoring and Protection
September 27, 2020 to October 01, 2020
Lisbon, Portugal
https://www.iaria.org/conferences2020/ICIMP20.html

Authors:
Irfan Tanoli - irfan.khan.tanoli@ubi.pt
Sebastião Pais - sebastiao@di.ubi.pt

PRESENTER

- Irfan Khan Tanoli
  - Post-Doc Researcher at University of Beira Interior.
  - Receive PhD Degree at Gran Sasso Science Institute, L'Aquila, Italy.
  - Received the M.S degree in software engineering from Technical University of Madrid, Madrid, Spain.
  - Current Research Interest
    - Natural Language Processing.
    - Controlled Natural Processing.
    - Semantic Analysis.
    - Software Engineering.
    - Machine Learning.
  - Current Research Project
    - Moves Project (http://moves.di.ubi.pt/).

CO-PRESENTER

- Sebastião Pais
  - Professor at the Computer Science Department, the University of Beira Interior.
  - Researcher at NOVA-LINCS and GREYC Laboratory.
  - Current research and teaching interests:
    - Artificial Intelligence.
    - Statistical Natural Language Processing.
    - Lexical Semantics.
    - Machine Learning.
    - Unsupervised and Language-Independent Methodologies.
  - Current Research Project:
    - Moves (http://moves.di.ubi.pt/).
    - C4 - Cloud Computing Competence Centre (c4.ubi.pt).

OUTLINE

- Introduction
- Background
- Controlled Natural Languages
- Properties
- Expression in CNLs
- Evaluation
- Outcome
- Conclusion

IARIA

Introduction

- On social networks, the storage, usage, and sharing of users data is regulated by privacy policies, mainly written in natural language.
- To regulate diverse data operations, social network providers (Facebook, Google, Twitter etc.) publish appropriate data policies on their sites and invite users to accept them
- However,
  - Natural language (e.g., English) is readable and understandable, but:
    - Policies may be ambiguous.
    - Policies are not machine readable.
- No automatic control on how data are actually going to be used and processed

Background

- A systemic comparison of five machine-oriented, English-based Controlled Natural Languages (CNLs), originally designed within different contexts to identify a CNL as final target of our translation.
- Investigate their effectiveness in expressing 6 different data policies as specified on a popular SN site (i.e., Twitter).
- Evaluate against well-defined classification scheme known as `PENS (Precision, Expressiveness, Naturalness and Simplicity)‘ and a new property Policy Enforcement.
- Apart for the contribution, the result of this study can also be useful for other researchers to select a specific CNL for their research work.

Controlled Natural Languages

- Three different policy based controlled natural languages:

- **Attempto Controlled English (ACE)**
  - A subset of standard English with a restricted syntax and restricted semantics described by a small set of construction and interpretation rules.

  - Developed for an automatic and unambiguous translation into a first-order logic.

  - Later improved with time and focus more towards knowledge representation and application for the Semantic Web.

Controlled Natural Languages

- **Protune (Provisional Trust Negotiation) policy language)**
  - Based on logic programming and, like CNL4DSA.
  - Designed for policy evaluation, enforcement, and negotiation.

- **Logic Based Policy Analysis Framework (LBPAF)**
  - A logic-based policy analysis language for policy specifications.
  - It also consists a policy analyser providing diagnostic information about detected conflicts, separation of duty, coverage gaps, behavioral simulation and policy comparison.

Properties

- **Precision**
  - The degree to which the meaning of a text can be directly understood and recovered from its textual form in a particular language, i.e., the sequence of linguistic symbols.
- **Expressiveness**
  - The range of propositions that a language is capable of expressing.
  - PENS considers the following characteristics of expressiveness:
    - universal quantification over individuals
    - relations of arity greater than one, i.e., languages which functions/predicates taking as input more than one argument.
    - general rule structures, e.g., if-then-else conditions.
    - Negation (failure or strong negation).

Properties

- **Naturalness**
  - How a language is `natural', in terms of reading and understanding from the user standpoint.
- **Simplicity**
  - How much simple (resp., complex) is to describe the language in an exact and comprehensive manner, covering syntax and semantics.
- **Policy Enforcement**
  - A language is policy enforceable or not?

Expression in CNLs

Consider following Twitter policy

P1: You can choose to upload and sync your address book on Twitter so that we can help you find and connect with people.

P1 in ACE:

IF You can choose to upload and sync your address book on Twitter
THEN we can help you P2 in ACE: find and connect with people.

Expression in CNLs

- P1 in Protune:

allow (help(we,you,(FindandConnect(people))))
ChoosetoUpload (you,address book,Twitter),
ChoosetoSync (you,address book,Twitter)

- P1 in LBAPF:

permitted(W,Y,help(P,find,connect,T)
do(Y,C,AB,TW,ChoosetoUpload,T), do
(Y,AB,TW,sync,T)

Evaluation

	ACE	Protune	LBPAF
Precision	Yes	Yes	Yes
Expressiveness	Yes	Yes	Yes
Naturalness	No	No	No
Simplicity	No	Yes	Yes
Policy Enforcement	No	Yes	Yes

TABLE: COMPARISON OF CONTROLLED NATURAL LANGUAGES

Outcome

- The evaluation shows that Protune and LBPAF fulfil the highest number of properties.
- They are formal at the syntactic level and have an associated formal semantics.
- Description is concise.
- They were not defined with a specific vocabulary associated
- Both have a policy enforcement infrastructure associated.
- Protune and LBPAF enjoy the property of medium expressiveness.

Conclusion

- We considered three Controlled Natural Languages and we evaluated them according to a set of standard properties defined in the literature.
- The evaluation is carried out based on the translation of a Twitter policies into the analysed CNLs.
- Aiming at choosing a CNL as the target language to automatically translate NL social network(s) data policies, the outcome of our evaluation helps us towards Protune and LBPAF.

THANK YOU