CERIAS

Digital Identity Management in ontological semantics: methodology and practice of domain representation

Evguenia Malaia, ABD, CERIAS & NLP Lab, Purdue University

Research goals of ontology support for domain representation

Ontological support for the domain: The existing ontological semantic resources adjusted and supplemented for the entire domain of digital identities to provide conceptual and lexical framework for the research. This entails an estimated 500-700 concepts and 1,500 lexical items added or modified;

Ontological support for the representation formats: ontological properties, relations and attributes have to be developed to express identity management schemes in a precise, formally unified way, to facilitate automatic processing;

Ontological support for identity policy language: development of enriched semantic representation of identity management vocabulary for manipulation of the natural language texts dealing with identity policies.

Methodology for domain acquisition in ontological semantics: convergence of approaches	
Top-down approach (Conceptual hierarchies to concept end-nodes and lexicon)	**Bottom-up approach (Corpus to lexicon)**
Delimit the **body of the corpus**, splitting it in two parts for validity check, map out the sub-domains within it (technological, legal, social, etc.);	Run the first part of the **corpus through the available lexicon** and filter out lexemes which are not available yet (Java program);
Map out ontological tree for the most important concepts for each sub-domain; acquire **necessary properties** for the domain;	Filter out **lexemes by frequency**: frequency in one document; cross-documental frequency. Decide on what frequency is necessary for a lexical item to be a valid vocabulary member (decisions on trademark and proprietary terms – e.g. Mac as a definition of a certain computer architecture, delimiting its platform properties and available application);
Create ontological **sub-hierarchies** needed to support sub-domains; acquire **lexemes closest** to the ontological end-nodes;	**Sort the lexical items** as to whether they belong to Digital Identity Management domain;
Decide on **multi-word expressions** necessary for the vocabulary (if necessary, check for their occurrence frequency, to observe if they are sufficiently wide-spread);	**Acquire** lexical items which pass the two previous filters;
Check for **multiple meanings** of available items in the lexicon, so that they include DIM semantics (where appropriate, include all meanings as well);	Run the second part of the corpus through the lexicon. Repeat steps 2-4 if necessary; also, if necessary, expand the corpus for another **validity check**;
Result: ontological hierarchy and lexicon for Digital Identity Management domain.	

Relations:

Properties
- …object-relation
  - representational-information-relation
    - has-identifying-information
    - collects-identifying-information
    - stores-identifying-information
    - requests-identifying-information
    - matches-identifying-information
    - issues-identifying-information
    - identifies-what
  - inverse-representational-information-relation
    - identifying-information-issued-by
    - identified-by
    - furnishes-identifying-information
  - scalar-object-attribute
    - temporal-object-attribute (0 to 1 (constant))
    - authenticity (range 0 to 1 (likelihood of being unaltered self))

Objects:

Object…
- …representational-object
  - identifying-information
    - URL
    - timestamp
    - software-identifying-information
      - serial-number
      - digital-certificate
    - hardware-identifying-information
      - serial-number
      - IP-address
    - user-identifying-attribute
      - temporary-user-attributes
        - geographical-location
        - avatar
      - knowledge-token
        - PIN
        - username
        - password
    - document-identifying-information
      - document-number
      - document-expiration-date…

Attributes:

Attribute..
- …constant-user-attributes
  - biometrics (of varying reliability)
    - retina-scan
    - iris-scan
    - finger-print
    - voice-print
    - foot-print
    - DNA-print
    - hand-geometry
    - saliva-sample
    - signature
    - keystroke-dynamics
    - facial-geometry
    - gait-attribute
  - documentary-user-representations
    - passport
    - visa
    - state-ID
    - driver's-license
    - PGP-signature
    - insurance-number
    - social-security-number

PURDUE UNIVERSITY

CERIAS®

Discovery Park e-Enterprise Center

CERIAS

Digital Identity Management in ontological semantics: methodology and practice of domain representation

Evguenia Malaia, ABD, CERIAS & NLP Lab, Purdue University

Examples of ontological entries

```
(HARDWARE-IDENTIFYING-INFORMATION
(DEFINITION (VALUE ("a representational object identifying hardware")))
(IS-A (VALUE (IDENTIFYING-INFORMATION)))
(IDENTIFIES-WHAT (SEM (INDEPENDENT-DEVICE)))
(IDENTIFYING-INFORMATION-ISSUED-BY (ORGANIZATION))
(OWNED-BY (SEM (HUMAN ORGANIZATION)))
)

(INTERNET-SERVICE-PROVIDER
(DEFINITION (VALUE ("a communication-corporation that provides access to Internet service for the
        public")))
(IS-A (VALUE (COMMUNICATION-CORPORATION UTILITY-CORPORATION)))
(HAS-MERCHANDISE (SEM (*NOTHING*)))
(AREA-OF-BUSINESS-ACTIVITY (SEM (COMMUNICATION-SERVE)))
(OBJECT-INVOLVED (SEM (COMPUTER NETWORK-SERVER USER-COMPUTER)))
(CUSTOMER-OF (INV (TELECOMMUNICATION-EQUIPMENT-MANUFACTURING-
        CORPORATION)))
)

(INVARIABILITY-ATTRIBUTE
(IS-A (VALUE (SCALAR-OBJECT-ATTRIBUTE)))
(DEFINITION (VALUE ("used to indicate how long the given type of object typically exists, from 1 –
        constantly, to 0 – non-existent")))
(DOMAIN (RELAXABLE-TO (OBJECT)))
(RANGE (VALUE (<> 0 1))))
)

(KNOWLEDGE-TOKEN
(DEFINITION (VALUE ("a representational object which can be memorized for identifying a user")))
(IS-A (VALUE (TEMPORARY-USER-ATTRIBUTE)))
(SUBCLASSES (VALUE (USER-NAME PASSWORD PIN-NUMBER)))    //to be extended//
(IDENTIFIES-WHAT (SEM (DEAFULT (USER-COMPUTER) RELAXABLE-TO (HUMAN))))
(IDENTIFYING-INFORMATION-ISSUED-BY (DEFAULT (ORGANIZATION) RELAXABLE-TO (USER-
        COMPUTER)))
(OWNED-BY (SEM (DEFAULT (USER-COMPUTER) RELAXABLE-TO (ORGANIZATION))))
)

(VOICE-PRINT
(DEFINITION (VALUE ("a physical quality of voice that can be used to identify a human ")))
(IS-A (VALUE (BIOMETRICS))
(IDENTIFIES-WHAT (SEM (DEAFULT (HUMAN-VOICE) RELAXABLE-TO (HUMAN))))
(INVARIABILITY-ATTRIBUTE (VALUE (.6)))
(UNIQUENESS (VALUE (.9)))
)

(SEMICONDUCTOR-STORAGE-DEVICE
(IS-A (VALUE (STORAGE-DEVICE)))
(DEFINITION (VALUE ("computer storage device that reads and writes information magnetically")))
(SUBCLASSES (VALUE (FLASH-MEMORY BIOS-CHIP SECURE-DIGITAL-CARD)))
(MATERIAL (SEM (SILICON)))
(FUNCTIONS-LIKE (SEM (MEMORY)))
)
```

Examples of lexical entries

A noun with two distinct lexical meanings:

```
(AVATAR
(AVATAR-N1 (CAT N)
(ANNO
(DEF "an icon or representation of a user in a virtual reality. ")
(EX “I wanted to change my avatar, but forgot to resize the picture for it.") (COMMENTS ""))
(SYN-STRUC ((ROOT $VAR0) (CAT N)))
(SEM-STRUC (AVATAR)))
(AVATAR-N2 (CAT N)
(ANNO
(DEF "a type of robots or interactive computer programs, which was design to teach, or help to learn ")
   (EX “We are developing a signing avatar for a math program in ASL.") (COMMENTS ""))
(SYN-STRUC ((ROOT $VAR0) (CAT N)))
(SEM-STRUC (CHATTERBOT))
))
```

Adjective which enters in two different syntactic structures:

```
(UNIQUE
(UNIQUE-ADJ1 (CAT ADJ)
(SYN-STRUC
((ROOT $VAR1) (CAT N) (MODS ((ROOT $VAR0) (CAT ADJ)))))
(SEM-STRUC
(^$VAR1 (INSTANCE-OF (SEM OBJECT)) (UNIQUENESS (VALUE 1)))))
(UNIQUE-ADJ2 (CAT ADJ)
 (SYN-STRUC
(((ROOT $VAR1) (CAT N)) (ROOT BE) (CAT V)
((ROOT $VAR0) (CAT ADJ)))
(SEM-STRUC
(^$VAR1 (INSTANCE-OF (SEM OBJECT)) (UNIQUENESS (VALUE 1))))
))
```

Specific information pertaining to the lexical item is stored with it:

```
(BACHELOR-OF-SCIENCE
(BACHELOR-OF-SCIENCE-N1 (CAT N)
(SYNONYMS  “BS degree”)
 (ANNO
(DEF “an educational degree granted for the successful completion of a college program in science.")
(EX "") (COMMENTS ""))
(SYN-STRUC ((ROOT $VAR0) (CAT N)))
(SEM-STRUC (SEM (EDUCATIONAL-LEVEL
(FIELD-OF-STUDY (SEM SCIENCE))
(YEARS-OF-EDUCATION (VALUE > 3))))
))
```

A complex entry referring to a proposition:

```
(MALWARE
(MALWARE-N1 (CAT N)
(ANNO
(DEF “any software developed for the purpose of doing harm to a computer system. ")
(COMMENTS "Note for the future: Malware can be classified based on how it is executed, how it
        spreads, and/or what it does. ") (EX ""))
(SYN-STRUC ((ROOT $VAR0) (CAT N)))
(SEM-STRUC (SOFTWARE
(INSTRUMENT-OF (SEM REFSEM1))))
(REFSEM1 (INCUR-DAMAGE
(BENEFICIARY (COMPUTER SOFTWEAR USER-COMPUTER))))
))
```

PURDUE UNIVERSITY

CERIAS

Discovery Park e-Enterprise Center