Ontological Semantic (OS) Support for Digital Identity Management (DIM): Expanding the Ever-Expanding Domain

John M. Spartz and Evguenia Malaia

References: Malaia, E. (2005) "Digital identity management for ontological semantics : domain acquisition methodology and practice." Doctoral dissertation, Purdue University

Abstract

Digital identity management (DIM) has emerged as a critical foundation for supporting successful interaction in today's globally interconnected society. It is crucial for not only conducting business and the government, but also for a large and growing body of electronic or online social interactions. In its broadest sense, identity management encompasses definitions and life-cycle management for digital identities and profiles, and the environments for exchanging and validating such information, including anonymous and pseudonymous representations. Although the basic tools underlying identity management have existed for a long time, we still lack comprehensive, dependable and flexible solutions for supporting multiple and partial identities. Moreover, support for anonymity, a key requirement for digital identity systems, should not undermine the dependability of the system and the accountability of the interacting parties. DIM systems should thus enforce good audit practices and support forensic analysis consistent with the criticality of the underlying system.

Because digital identities have such varied uses and meanings, and because of the far-reaching implications of DIM policies, which extend to free speech, privacy, and online accountability, it is essential to develop a universal vocabulary for developing a digital identity framework and policy language. A firm understanding and facility with DIM vocabulary, which includes and ever-expanding and prevalent list of acronyms and DIM products, is compulsory for any and all work in identity management. We propose the framework of ontological semantic processing and text meaning representation through ontology as a possible and probable means through which one can come to possess a working knowledge of DIM and associated vocabulary.

Goals of OS Support for DIM Domain

To continue to adjust, supplement, and provide a conceptual and lexical framework for the domain of digital identities and their management, which entails acquiring an estimated 500-700 ontological concepts and some 1,500 new or modified lexical items.

To expand the existing work in DIM, to include lexical items and any necessary ontological concepts, with special attention to both products and acronyms and their prose synonyms in the DIM domain.

Methodology for Domain Acquisition in Ontological Semantics: A Convergence of Approaches (Malaia, 2005)

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 yet available (Java program/by hand)
Map out an 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 DIM 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 (Malaia, 2005)

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
    - 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

DIM Acronym Lexical Entry

```
(SAML
(SYNONYMS "SECURITY-ASSERTION-MARKUP-LANGUAGE")
(SAML -NI (CAT N)
(ANNO (DEF "an XML-based framework for exchanging security information.")
      (EX "The SAML protocol consists of four main components.")
      (COMMENTS ""))
      (SYN-STRUC (ROOT $VAR0) (CAT N)))
      (SEM-STRUC (MARKUP-LANGUAGE (SEM COMPUTING-
      PROTOCOL (INSTRUMENT-OF (SEM EXCHANGE (THEME
      (SEM INFORMATION (HAS-OBJECT-AS-PART (SEM
      COMPUTER-PROGRAM-FUNCTION)))))))))
```

DIM Acronym Prose Synonym Lexical Entry

```
(SECURITY-ASSERTION-MARKUP-LANGUAGE
(SYNONYMS "SAML")
(SECURITY ASSERTION MARKUP LANGUAGE -NI (CAT N)
(ANNO (DEF "an XML-based framework for exchanging security information.")
      (EX "The SAML protocol consists of four main components.")
      (COMMENTS ""))
      (SYN-STRUC (ROOT $VAR0) (CAT N)))
      (SEM-STRUC (MARKUP-LANGUAGE (SEM COMPUTING-
      PROTOCOL (INSTRUMENT-OF (SEM EXCHANGE (THEME
      (SEM INFORMATION (HAS-OBJECT-AS-PART (SEM
      COMPUTER-PROGRAM-FUNCTION)))))))))
```

DIM Product Lexical Entry

```
(eToken
(eToken-N1 (CAT N)
(ANNO (DEF "an Aladdin company USB based smart card device)
      (EX "eToken's strong authentication helps ensure that only authorized
      individuals access your organization's sensitive information.")
      (COMMENTS ""))
      (SYN-STRUC (ROOT $VAR0) (CAT N)))
      (SEM-STRUC (STORAGE-DEVICE (INSTRUMENT-OF (SEM
      STORE-COMPUTER-DATA (THEME (SEM COMPUTER-DATA
      (THEME-OF (SEM AUTHENTICATE)))))))))
```

*Entries were acquired using the Bottom-Up Approach

PURDUE UNIVERSITY · PURDUE HOMELAND SECURITY INSTITUTE · CERIAS · REGENSTRIEF CENTER FOR HEALTHCARE ENGINEERING · Discovery Park e-Enterprise Center