Representing Risky Behavior for Sexually Transmitted Diseases on the Semantic Web by Using Notation3

Herman Tolentino, MD; Catherine Pepper, MLIS, MPH; Abdou Salam Gueye, MD, MPH; Rachel Kachur, MPH; Mary MacFarlane, PhD

¹Public Health Informatics Fellowship Program, Career Development Division, Office of Workforce and Career Development; ²Strategic Science and Program Unit, Coordinating Center for Infectious Diseases; ³Division of Global Aids, National Center for HIV/AIDS, Viral Hepatitis, STD, and Tuberculosis Prevention; ⁴Division of Sexually Transmitted Disease Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and Tuberculosis Prevention; Centers for Disease Control and Prevention, Atlanta, Georgia

Note: The findings and conclusions in this presentation have not been formally disseminated by the Centers for Disease Control and Prevention and should not be construed to represent any agency determination or policy.

Background. Use of Internet technology for communication has become ubiquitous, and social networks which may be involved in transmission of sexually transmitted diseases (STDs) communicate with explicit language that implies risky behavior for STDs [1]. We can leverage the Internet and the technologies related to its use to enable detection of risky behavior that relates to disease transmission. Specifically, one method that facilitates automatic detection of risky behavior is to represent domain concepts and relationships as background knowledge by using ontologies. This representation can facilitate understanding of site content and enable detection of risky behaviors associated with STDs. This representation uses Semantic Web (SW) [2] technology called Notation3 (N3) [3] that enables easy annotation of concepts and relations in an ontology.

Methods. We used the following as inputs for the ontology: (1) samples of text obtained from web sites where related explicit language is available, (2) risk classification developed by our group by using the work of Sande and colleagues [4], and (3) inputs from local subject-matter experts. We created a conceptual diagram to facilitate conversion of concepts and relationships to N3 statements (Figure 1). We subsequently wrote the ontology in N3 and used a freely available tool to convert it to Resource Description Framework/Extensible Markup Language (RDF/XML) format for machine processing.

Results. We were able to represent key concepts and relationships that might be used to facilitate identification of STD-related risky behavior from text identified through the Internet by using N3. Certain concepts might already be available in existing ontologies (i.e., the disease ontology from the Open Biomedical Ontologies [5]). We visually represented these concepts and their relationships associations as a cyclic-directed graph (Figure 1). Examples of concepts represented include Act (i.e., "barebacking" and "fisting"), Gender Preference (i.e., "bisexual"), and Position Preference (i.e., "top" or "bottom").

Conclusion. N3 is a user-friendly SW language that can be used to represent ontologies. We represented concepts related to risky behavior associated with STDs as an ontology written in N3. This representation can be exported to machine-readable RDF format by using a free SW Tool called cwm [6] so it can be made available for reuse and merging with other related ontologies. Future use for this representation includes combining it with a custom-developed lexicon, containing concept-mapped terms and their variants, identifying risky behavior concepts from unknown text, and automatically performing markup annotation and highlighting of identified concepts for human visualization and interpretation.

References

Acknowledgements
This research was also supported in part by an appointment to the Public Health Informatics Fellowship Program (PHIFP) sponsored by the Office of Workforce and Career Development at the Centers for Disease Control and Prevention (CDC). PHIFP is administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and CDC.

Figure 1. Conceptual graph of risky behavior associated with sexually transmitted diseases

AMIA 2007 Symposium Proceedings Page - 1134