Ontologies for Provision of HIV/Aids Information Using Telephony Infrastructure

Mathe Maema
Department of Computer Science
Rhodes University
P.O. Box 94
Grahamstown 6140
mathe@rucus.ru.ac.za

Alfredo Terzoli
Department of Computer Science
Rhodes University
P.O. Box 94
Grahamstown 6140
a.terzoli@ru.ac.za

Lorenzo Dalvit
Department of Education
Rhodes University
P.O. Box 94
Grahamstown 6140
ldalvit@gmail.com

ABSTRACT

Over the years, the number of new telephony services has increased drastically. This increase can, in part, be attributed to the public switched telephone network (PSTN) converging with computer networks into a single network, based on the Internet Protocol (IP), as well as the advent of mobile telephony. Although innovative, many of these services are database driven and consequently lack intrinsic reasoning capabilities. Hence, they cannot be used with ease to support rapid creation of new services, which are dependent on existing data. In light of this, we explore the feasibility of creating knowledge driven services by creating a service that provides information related to HIV and AIDS using the converged telephony infrastructure. The proposed service will use an ontology to represent knowledge. Ontology refers to a formal explicit description of concepts within a specific domain, which in the context of this research will be HIV and AIDS information. This ontology will interface with a voice application that will allow users to call and receive requested information about HIV and/or AIDS. The role of the ontology will be to infer through use of reasoning techniques what information should be returned to the user.

Categories and Subject Descriptors

C.2.1 [**Computer Systems Organization**]: --Computer-Communication NetworksNetwork Architecture and Design Design, Human Factors

Keywords

Ontology, information retrieval, telephony

1. INTRODUCTION

Over the years, the Department of Computer Science at Rhodes University has engaged in numerous projects that involve telephony services on a converged network. Many of these projects are based on Asterisk, an open source telephony engine and toolkit that supports both traditional and voice over IP (VoIP) telephony services [11]. As an extension to the work done, we explore service creation using ontologies on an Asterisk-based platform called iLanga, a system developed at Rhodes University to leverage on the convergence of telephony and data networks [12]. Consequently, a telephony service that moves beyond the use of traditional database system for the storage and processing of data will be created. This service will be knowledge-driven since ontologies are formal knowledge representation techniques that explicitly specify a shared conceptualisation [5, 8].

In this paper, we will present current work towards building the conceptual model of our ontology, which will capture knowledge on HIV and AIDS that is specifically tailored for laypersons. The objective is to develop a telephony service that will deliver this knowledge in order to allow users to make informed decisions about their health and well-being. This objective will contribute directly to ongoing efforts by Rhodes University under the Higher Education HIV/AIDS (HEAIDS) Programme funded by the European Union to reduce rates of infection and increase better management of the disease.

The paper is organised into 5 Sections. Section 2 provides a brief overview of the proposed service. Related work pertaining to construction of an ontology is examined in Section 3. The process for modelling the conceptual framework for our ontology is discussed in Section 4. Finally, in Section 5 conclusions and work yet to be done are discussed.

2. PROPOSED SERVICE

This section describes the proposed service and briefly provides a motivation for its selection. In a nutshell, the aim of the research is to create a knowledge-driven telephony service that will allow people to call and receive answers to questions about HIV and AIDS. The service will be instrumented to inform its own development by, for example, keeping a record of the answers given to callers. This will enable us to see what content is relevant to the needs of the callers. Further, this will help to inform decision-making with regards to updating of content at a later time.

2.1 Motivation for Proposed Service

In the previous section, it was explained that an ontology refers to explicit specification of a *shared* conceptualisation. *Shared* suggests that the conceptualisation is not that of an individual but that of a collective. Thus, an ontology is not just about specifying concepts and their relation to each other using an ontological language and/or encoding such as Web Ontology Language (OWL). An ontology is about capturing consensual knowledge that is accepted by a collective or community [4]. According to Dragan *et al.* [4] this makes any ontology a better medium for facilitat-

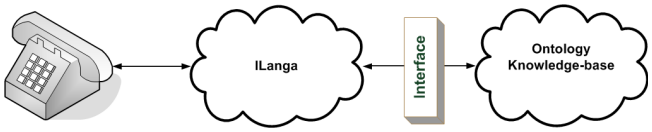

Figure 1: Overall Architecture

ing knowledge sharing and automated knowledge reuse by applications.

The above, compounded with the increased use of ontologies for knowledge representation (due to adoption of the Semantic Web), was central in the selection of the proposed service. By focusing on the dissemination of HIV and AIDS information, the service allowed for a strategic synergy to be formed with the team engaged in tailoring knowledge related to HIV and AIDS within the Rhodes University's HEAIDS Programme.

Another reason for the selection was the growing evidence that telephone-based applications such as interactive voice response (IVR) systems can yield satisfactory health outcomes particularly for targeting low-income individuals. According to Migneault *et al.* [9], who claim to have two decades of experience in developing and evaluating such systems, this can be attributed to the fact that "telephones are ubiquitous" as well as familiar to most people. Consequently, information delivered telephonically is easily accessed.

2.2 Service Architecture

Figure 1 depicts the core architecture for the proposed service. The architecture presents iLanga and the ontology knowledge-base as abstractions that interact with each other through an interface. On the basis of this architecture, the proposed service will operate as follows:

- A user will place a call to request information. The details of processing the call which involve interactions with PSTN, the voice application, VoiceXML browser, speech recognition and text-to-speech applications will be handled by iLanga.
- The request from iLanga will be forwarded to the ontology-based back-end for information retrieval via an interface that would facilitate querying of the ontology.
- The retrieved information might be presented to the text-to-speech application or played directly to the user if it was an audio recording.

3. METHODS AND METHODOLOGIES FOR ONTOLOGY CONSTRUCTION

There are a number of methods and methodologies for ontology construction. Examples include the Cyc method, the Uschold and King's method, the Grüninger and Fox's methodology, METHONTOLOGY and the 101 Method. The suitability of each method and methodology is dependent on several factors. Some of these factors, as identified by Gómez-Pérez *et al* [7], include the development life cycle and strategies employed for identification of concepts.

In comparing the above-mentioned (except the 101 Method), Gómez-Pérez *et al* [7] conclude that each approach has some drawbacks. This is in alignment with the position of various authors, who assert that there is no single method or methodology that is deemed as best for ontology construction [4, 10]. Hence, as noted by Brusa *et al* [3], it is common to merge different methods and methodologies in order to capitalise on the strengths of each.

Ontology construction is an iterative process that demands careful analysis, feedback and redesign throughout the entire process [4, 10]. This process as described by Brusa *et al* [3] has two main phases: specification and conceptualisation. The objective of the specification phase is to clarify the use and scope of the ontology through an informal process of knowledge acquisition about the domain to be modelled. In contrast, the objective of the conceptualisation phase is to organise and structure the acquired knowledge using representation that is independent of the tools that will be used to formalise the ontology [3, 7].

The next section discusses both the specification and conceptualisation of our ontology using a hybrid of some of the above mentioned methods and methodologies. That is, we will describe the process we are following to construct the HIV and AIDS ontology for knowledge dissemination.

4. BUILDING THE CONCEPTUAL MODEL OF HIV AND AIDS ONTOLOGY

Conceptualisation is an important activity that "determines the rest of the ontology construction" [7]. This process is preceded by the specification of the ontology. In this section, we will focus on these two major phases of ontology construction. That is, we will not delve into the formalisation and implementation of the ontology since this phase is yet to be completed. Protégé has been selected for implementation since it is regarded as one of the leading ontological engineering tool [4, 10].

4.1 Specification: Determining the goal and scope of the ontology

For all methods and methodologies, the first step is to identify the goal or purpose of the ontology together with its scope. In our case, this step required a series of meetings with members of the HEAIDS programme team. These meetings served to delineate the goals of the telephony service and consequently that of the ontology. Figure 2 captures the main goal of our ontology which is to provide knowledge about HIV and AIDS as well as the delineated sub-goals. According to the figure, the knowledge will focus on 5 key areas: basic facts, counselling and testing, prevention and awareness, rights and responsibilities, lastly medication and treatment.

The meetings also provided a platform for determining the types of questions that can be answered by the ontology. Grüninger and Fox [6] refer to these questions as *competency questions* since they provide the scope of the ontology by expressing the capability of what knowledge can be availed.

4.1.1 Competency questions

Competency questions by definition are representation of queries in a form of questions that will be used to test the validity of the ontology in serving its intended purpose [6]. Some of these questions that will be used to determine the competency of our ontology include:

- What's the difference between HIV and AIDS?
- What is the relationship between tuberculosis and HIV?

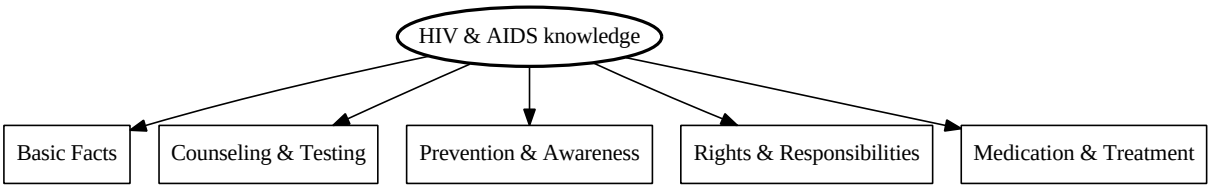

Figure 2: Specification for HIV and AIDS information dissemination ontology

- How does one get infected with HIV?
- Does male circumcision prevent HIV transmission?
- Can an employer ask an employee to have an HIV test?
- Are test results confidential?
- Where can one go for testing?
- When on antiretroviral therapy, can one transmit the virus to others?

4.1.2 Identifying candidate ontologies for reuse

Subsequent to the listing of competency questions and clarifying the scope of the ontology, the next step was to investigate availability of other similar ontologies, which could be reused. With respect to representing knowledge about places for accessing services for HIV and AIDS, ontologies such as the travel ontology may be used. However, for representing many of the other concepts, no specific ontologies were identified for communicating HIV and AIDS knowledge to a layperson.

4.1.3 Enumeration of important terms

Noy and McGuinness [10] propose enumerating all terms that may be important to the ontology as a means of defining the vocabulary. In our case, we used Termine, a tool by the National Centre for Text Mining [2] for extracting candidate terms from a text. The text was collated mainly from the knowledge centre of the UNAIDS [1] website. After the extraction, we manually selected terms that are either concepts, instances, concept attributes or terms that express relations between concepts.

Example of extracted terms include: anonymous HIV test, antiretroviral drug, HIV test, community home-based counselling, confidential HIV test, female condom, health care provider, HIV status, immune system, male circumcision, mother-to-child transmission, pretest counselling, safer sex, sexual act, window period, etc.

4.2 Conceptualisation

As Gómez-Pérez *et al* [7] note, conceptualisation is the most important activity and deserves special attention since it has impact on the rest of the ontology construction. Its goal is to organise and structure knowledge acquired in the specification phase into taxonomies and vocabulary that is representative of the ontology. The taxonomy together with the vocabulary provide a conceptual framework for the retrieval of information [4].

Due to the elaborateness of METHONTOLOGY in guiding the process of conceptualisation of an ontology, we will follow this methodology for modelling the conceptual model of our ontology. This methodology defines a set of ordered tasks to be performed. However, the order is not strict "as in waterfall life cycle" [7] but needs to be maintained to ensure consistency and completeness of the process. The required tasks as cited by Gómez-Pérez *et al* [7] are as follows:

Table 1: Sample of glossary of terms from HIV and AIDS ontology

Name	Synonym	Acronym	Description	Type
Acquired immunodeficiency syndrome	–	AIDS	An advanced stage of HIV infection in which the immune system begins to fail	Concept
Antiretroviral drug	Antiretroviral medication	ARV	Drug or medication for managing the spread of HIV	Concept
Human immunodeficiency virus	–	HIV	A virus that causes AIDS	Concept
HIV Positive	Seropositive	HIV +	HIV status showing that a person has been infected with HIV	Instance attribute

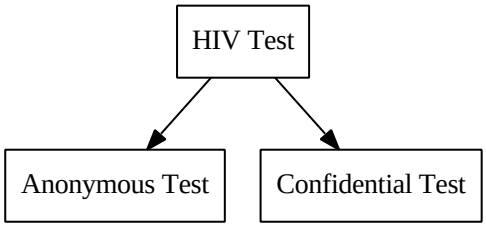

Figure 3: Sample classification for HIV test concept

1. Build glossary of terms by tabulating the extracted terms such that we have a description, a synonym and acronym. Table 1 provides a small example from the HIV and AIDS ontology.
2. Build concept taxonomies by classifying concepts listed in the glossary of terms (see figure 3, for an example).
3. Build ad hoc binary relation diagrams that establish relationships between concepts of the ontology with concepts of other ontologies.
4. Build concept dictionary to capture ad hoc relations of each concept as well as concept instance and class attributes.
5. Describe ad hoc binary relations in detail so as to reflect the relations in the concept dictionary.
6. Describe instance and class attributes in detail from the concept dictionary by means of a table that includes fields that act as descriptors of the attributes.
7. Describe in detail any constants that may be defined in the glossary of terms.
8. Define formal axioms, which are logical expressions for specifying constraints.

9. Describe rules that may be used to infer knowledge within the ontology.

Currently, we are rigorously building concept taxonomies whilst also engaging in the process of refining the glossary of terms.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we presented the work towards building a conceptual model of the ontology that will form the core of the service created to offer HIV and AIDS information over the telephone. Although the ontology is not yet implemented, we hope that conceptualisation of a domain is recognised as an important activity for constructing an ontology. Hence, investing time and effort in the conceptualisation may result in easier implementation of the ontology to be constructed.

As part of work to be done in the future, once the conceptual model is complete and has been translated into a fully functional ontology, a verification process will follow. This will involve testing for any inconsistencies and confirming that identified competency questions are answered by the ontology. Finally, the voice application and interface for facilitating the retrieval of information will be built.

6. ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of Telkom SA, Bright ideas Projects 39, Comverse SA, Stortech, Tellabs, Amatole, Mars Technologies, openVOICE and THRIP through the Telkom Centre of Excellence in the Department of Computer Science at Rhodes University. Further, the authors would like to thank members of the HEAIDS programme.

7. REFERENCES

[1] The Joint United Nations Programme on HIV/AIDS. Online: `http://www.unaids.org/en/KnowledgeCentre/Resources/FastFacts/default.asp` [Last accessed: 31 July 2008].

[2] The National Centre for Text Mining (NaCTeM). Online: `http://www.nactem.ac.uk/software/termine/` [Last accessed: 31 July 2008].

[3] Brusa, G., Caliusco, M. L., and Chiotti, O. A Process for Building a Domain Ontology: an Experience in Developing a Government Budgetary Ontology. In *AOW '06: Proceedings of the second Australasian workshop on Advances in ontologies* (Darlinghurst, Australia, Australia, 2006), Australian Computer Society, Inc., pp. 7–15.

[4] Dragan, G., Dragan, D., and Vladan, D. *Model Driven Architecture and Ontology Development.* Springer, 2006.

[5] Gruber, T. R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In *Formal Ontology in Conceptual Analysis and Knowledge Representation* (Deventer, The Netherlands, 1993), N. Guarino and R. Poli, Eds., Kluwer Academic Publishers.

[6] Grüninger, M., and Fox, M. S. Methodology for the Design and Evaluation of Ontologies. In *IJCAI95 Workshop on Basic Ontological issues in Knowledge Management Sharing* (1995), D. Skuce, Ed.

[7] Gómez-Pérez, A., Fernández-López, M., and Corcho, O. *Ontological Engineering: With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web.* Springer, 2004.

[8] Lam, T. H. W., and Lee, R. S. T. iJADE FreeWalker-An Intelligent Ontology Agent-based Tourist Guiding System. *Studies in Computational Intelligence (SCI) 72* (2007), 103–125.

[9] Migneault, J. P., Farzanfara, R., Wrighta, J. A., and Friedmana, R. H. How to Write Health Dialog for a Talking Computer. *Journal of Biomedical Informatics 35*, 5 (October 2006), 468–481.

[10] Noy, N., and McGuinness, D. Ontology Development 101: A Guide to Creating Your First Ontology. Tech. rep., Stanford University, 2001.

[11] Penton, J., and Terzoli, A. Asterisk: A converged TDM and Packet-based Communications System. In *South African Telecommunications Networks and Application Conference (SATNAC)* (2003).

[12] Penton, J., and Terzoli, A. iLanga: A Next Generation VOIP-based,TDM-enabled PBX. In *Southern African Telecommunications Networks and Applications Conference (SATNAC)* (2004).