

Toward the Interoperability of Language Resources

July 13-15, in conjunction with the 2007 LSA Summer Institute



Enhancing semantic interoperability for language resources

Jung-ran Park
Drexel University

In this position paper I would like to discuss some critical factors hindering semantic interoperability across ever-growing distributed digital collections and language resources.

As evinced through information sharing for traditional non-networked bibliographic collections such as books in libraries, successful resource discovery and exchange across distributed digital repositories demands semantic interoperability based on consistent and accurate resource description.

The flexibility and complex structure of natural language allows for the representation of a concept in various ways. Thus, consistent and accurate resource description is essential. Without it, semantic interoperability, even among digital repositories employing the identical vocabulary scheme (e.g., Dublin Core), metadata elements and digital collection software configuration will become increasingly problematic, resulting in a decrement in information sharing.

The salient characteristics of the Dublin Core (DC) metadata scheme, which was approved as an ANSI/NISO standard (Z39.85) in 2001, emerge from its simplicity, flexibility and interoperability. The functionalities of the DC metadata elements are easy to implement owing to this simplicity, which creates high compatibility across multiple repositories. However, despite such characteristics, there are inherent conceptual ambiguities and semantic overlaps in some of the DC metadata elements. That is, in addition to the lack of surrounding context in which a DC metadata element and corresponding usage occur, semantic overlap among certain DC metadata element names and their corresponding definitions create conceptual ambiguity. This consequently hinders accurate, consistent and complete application of the DC metadata scheme, in turn hindering semantic interoperability across the digital repositories utilizing DC metadata scheme.

In this position paper I will discuss the conceptual ambiguities and semantic overlaps inherent in the DC metadata elements, utilizing linguistic semantic analysis to examine the DC metadata element name and its corresponding definition. In this discussion, I would like to address the fact that the DC metadata scheme needs to be refined in order to disambiguate semantic relations of certain elements that present semantic overlaps and conceptual ambiguities between element names and their corresponding definitions.



As well, in this position paper I will discuss implications and issues drawn from the evaluation of the current status of metadata creation and mapping between heterogeneous vocabulary schemes. This is based on my previous studies (Park 2005, Park 2006, Bui and Park 2006), all of which concern semantic interoperability of concept representation across digital repositories. The findings of these studies suggest that the critical roadblock in achieving semantic interoperability is posed by the lack of a mediation mechanism. Thus, the development of mediation mechanisms such as concept networks that facilitate the metadata/vocabulary creation and mapping process between heterogeneous vocabulary schemes is critically needed for enhancing semantic interoperability.

