Data Driven Ontology Alignment
Nigam Shah, Stanford University

In our recent work with annotations of tissue microarrays, we have automatically mapped approximately 80% of annotations for the samples in the Stanford Tissue Microarray Database to ontology terms. We have shown that a significant proportion of the diagnosis-related annotations map to terms from both the NCI thesaurus and SNOMED-CT. This mapping of a single record to terms from different ontologies presents a concrete data-driven mechanism for aligning related ontologies by using them for annotation. Such data-driven alignments have the potential to be complementary to existing alignment approaches, such as PROMPT, and may be used in tandem to better align related ontologies.