Inferring ontological boundaries based on implied social networks: creating an architecture for participation in Life Sciences.

TIMOTHY M. KUNAU, MS, (kunau@umn.edu)

The difficulty of sharing and comparing Life Sciences data has reached a critical level. The National Center for Biotechnology Information (NCBI) Genbank database currently lists 86 billion base pairs and is doubling every six to ten months. Researchers struggle to interact with current data as the flood of new data swells around them. This paper addresses mapping knowledge domains from NCBI public data to aid in the creation of social connections. Topologies of interest are created between researchers to foster further collaborative research and data sharing. Standards are beginning to emerge that might be used to produce the functional and semantic interoperability necessary to support the examination of these data. These standards include: common information models, data elements, ontologies, data structures, and transport standards.

The relevant information for this study is collected from NCBI Genbank entries using published XML application programmer interfaces (APIs) to create attention profiles using APML. We investigate stable clustering algorithms and comment on their applicability to the creation of social networks. We contrast our effort with cluster studies using publication abstracts found in NCBI PubMed and the NEC CiteSeer database. We evaluate existing open standard social networking platforms, Google Open Social, FaceBook, and Ning, and their appropriate application to academic Life Sciences research communities.

In this paper, we explain the details of the method, in particular the application of implied social networks to disparate research communities and the creation of new tools based on emerging standards to support the exchange of data and Life Sciences research collaboration. Using these tools, a research group might find a group working in a similar domain that would be hard to discover using other means.