Licensing and Technology Transfer Opportunity

Title of Technology:

A System and Method for Linked Data Process to Explore Interdisciplinary Research Collaborations across Enterprises.

Intellectual Property Status: "Application Published". Filed at the Indian Patent Office, with application number as 201741019854 on 6-June-2017.

Brief Description of Invention

The general purpose of the invention is to provide a linked data process that facilitates the identification, interconnection and exploration of interdisciplinary research collaborations:
(a) Within an enterprise, among its constituent units
(b) Between enterprises and
(c) Between enterprises and applications in the World Wide Web

The present invention provides a method and system for linking data processes in interdisciplinary research collaborations. The non-semantic data is transformed into semantic data using a data transformation module. The data is cleaned up by removing inconsistencies and represented in standard representation format. Resources are identified in the cleaned up data. The cleaned up data is mapped with a resource representation using semantic ontologies and the mapped data is converted into triple format. The triplicated data is enriched with semantic vocabularies and is linked with other data sources on the web using a link reconciliation module. The interlinked data available in machine-readable RDF format is deployed into a database. The visualization module allows visualization of machine readable RDF format data in human interpretable format using SPARQL protocol. The clients access the human interpretable format data through SPARQL protocol, custom applications or linked data applications.

A. Claims of the Invention

1. **Linked data process** - Represents a novel process that involves publishing entity data as RDF dataset, which includes enrichment and interlinking with other enterprises and

applications of the World Wide Web. The linked data process prescribes a "federated" architecture, based on the semantic Web, wherein all the data required for exploring interdisciplinary research collaborations, need not be stored centrally. Hence, the process is flexible and decentralized, since the tools used to implement different stages of the publishing process are disparate. Moreover, since the interdisciplinary data is in RDF format, any human or System (such as Web agent, crawler or semantic browser) can query the data using SPARQL and analyze the results.

2. **SEEC (Semantic Exploration of Enterprise Collaboration) application** - Represents a novel consuming process for discovering interdisciplinary research collaborations in enterprises. The novelty lies in the design of input filter widgets and output visualization widgets, which offer intuitive ways for the end-users to explore the collaborations.

B. Advantages of our approach over other known alternatives

1. The linked data process advocates a *self-service* approach in both publishing and consuming phases.

a. *Publishing Phase*: The linked data process is a "self-service" approach wherein the end-users can execute different stages of the process without requiring much technical assistance. The end-users have "full control" over the publication process, since every stage is designed to be manual or semi-automatic in nature, requiring human intervention. Second, all the tools used for executing different stages of the process have a GUI and are "easy-to-use" for non-expert end-users. Third, the process is "reproducible", as substantiated by the working example. Fourth, the publishing process is "flexible and decentralized", since every tool used in implementing different stages of the process are independent of each other, thus provides a decentralized setting. Lastly, the results of the published RDF dataset can be "shared via visualization" in a multitude of formats, enabling any human or system (Web agent, crawler or browser) to query and reuse the data.

b. *Consuming Phase*: In the SEEC application, the input filter widgets are intuitive and offer a "self-service" approach for the end-user to design queries for discovering interdisciplinary research collaborations in enterprises. Furthermore, the results of the queries can be visualized graphically as charts through a multitude of output visualization

widgets.

2. The linked data process scores over its competitors in Economic Viability and Time to Market factors.
   a. Economic Viability (cost) factor - From a cost benefit perspective, all the tools prescribed in the patent application are open-source. This advantage contrasts with the existing applications for discovering enterprise collaborations which are commercial (such as SciVal, PURE and VIVO).
   b. Time to Market factor - The SEEC application is a ready-to-use application that can be reused by any enterprise for exploring the interdisciplinary research collaborations: within their enterprise, with other enterprises, with applications on the Web, without much changes. The only prerequisite is that the enterprise has deployed its entity dataset as a SPARQL 1.1 compliant endpoint or as a RDF dump, by following all the steps of the publishing process.