



**Large Scale Distributed Information Systems Lab**  
<http://lsdis.cs.uga.edu/>



University of Georgia  
Computer Science Department

## **The LSDIS lab – a quick research overview**

The Semantic Web has emerged as the vision of the next generation of the web, in which meaning is associated with Web resources (data- documents and digital content as well as services) and represented in a machine processable form. LSDIS lab is one of the leading research groups in the world in the emerging area of Semantic Web and Semantic Web Services/Processes. Core faculty expertise includes distributed databases, information systems, knowledge representation, AI, decision theoretic planning, distributed systems, software engineering, workflows and Web processes, and algorithms. This has resulted in substantial body of the work in various areas, including

- development, management and visualization of populated ontologies
- automatic metadata extraction and semantic annotation (with associated challenges in entity identification/recognition and resolution /disambiguation) of all forms of (structured, semi-structured, and unstructured) textual data, digital media as well as scientific experimental data,
- scalable and high performance query processing and reasoning, including RDF query processing and analysis of large RDF graphs for discovery of complex relationships (called semantic associations), and
- specification of Semantic Web Services as well as use of semantics to specify the complete Web process life cycle

Semantic applications in the areas of biology, health care, national security, financial services, GIS and risk & compliance have been built, and several of these are deployed. Emerging research topics at LSDIS include (a) applying semantics to enable new capabilities at middleware, distributed systems and network levels, (b) development of highly scalable solutions with distributed algorithms on P2P and grid environments, (c) multi-paradigm reasoning spanning thematic, geospatial and temporal reasoning, and (d) virtual interactive interfaces for 3D visualization of digital content and data, metadata, ontological knowledge and results of reasoning.

Faculty members have extensive collaboration with industry (e.g., IBM and CISCO), and are involved with many international bodies and initiatives including W3C, OASIS, and Eclipse. They are also active members of multidisciplinary activities as members of UGA Faculty of Engineering, UGA Institute of Bioinformatics, UGA Biomedical and Health Science Institute, UCGIS, etc. LSDIS's strength is its vision and people; the latter (as of Fall 2005) consists of five faculty members, two associate faculty members, two research staff, at least sixteen funded students (majority pursuing PhD), and several non-

funded students. Their work has led to two commercial products and successful companies based on technology transfer, with over six millions spent on Athens' local economy and many high-tech employments of LSDIS and UGA graduates. LSDIS's PhD students routinely do internships at top places such as IBM Almaden, IBM Watson, Amazon, etc., LSDIS graduates are employed at top companies including IBM research, Yahoo!, Microsoft, Oracle, Amazon, SAP, and other top companies in the field, and some PhDs have chosen academic careers.

A top level summary of current projects in the LSDIS lab, funded by NSF, NIH, ARDA, IBM, and several other sponsors is as follows:

**SemDis** (Discovering Complex Relationships in Semantic Web): This NSF funded medium-ITR project involves modeling, discovering and reasoning about complex relationships on the Semantic Web that will transform the hunt for documents into a more automated analysis leading to insight and knowledge discovery. Among various output from the project are SWETO (Semantic Web Technology Evaluation Ontology), TOnToGen (Test Ontology Generation Tool), BRAHMS (A WorkBench RDF Store And High Performance Memory System for Semantic Association Discovery), and several algorithms for semantic association discovery over large RDF graphs, relationship based document ranking and ranking of complex relationships.

**METEOR-S** (Semantic Web Services and Processes): This project involves the creation and application of a broad variety ontologies related to data, function, non-functional/QoS and execution semantics to support the complete Web process lifecycle encompassing annotation, discovery, composition, optimization and execution. Key outputs of this project include semantic annotation of Web services using WSDL-S (with IBM), free open source graphical process designer for WS-BPEL, RADIANT (a graphical semantic web service annotation tool), ontological modeling of WS-Policy and WS-agreement, etc. Our current direction in METEOR-S is towards creating a framework for autonomic web processes by using involving multi-paradigm reasoning, semantic modeling and operations research based optimization (using DL reasoning, ILP, MDP to create Web processes with self-CHOP - self configuring, self healing, self optimizing and self healing properties).

**Bioinformatics for Glycan Expression**: This project is a part of a larger National Cancer Research Resource center, and involves substantial collaborations between LSDIS researchers and biologists at the Complex Carbohydrate Research Center. Key results include GLYDE (a representation standard that is being adopted by community of Glycomics researchers), GlycO (a very large and comprehensive 600+ class, 11 level deep), ProPreO (a large ontology that captures the process used in high throughput experiments), a tool for semantic search and browsing of large populated ontologies, development of bioinformatics semantic web services (using WSDL-S) and directory (semantic UDDI), semantic annotation of non-textual experimental data, etc. Recent work is investigating pathway development workbench for genomic researchers, with integrated access, analysis and discovery support covering experimental as well as textual data.

**SemGrid** (Semantic Discovery on Adaptive Services Grid): This early stage NSF funded project collaborates with large EU funded ASG project. It is investigating use of semantic associations in Web Service discovery and dynamic Web Process composition, and computing semantic Associations over the grid.

**Semantic Middleware**: This project is doing fundamental research and developing core technologies related to the challenging problems of entity and relationship identification/recognition, extraction/annotation, entity resolution/disambiguation, matching, mapping, and rule processing. An interim outcome is the concept of Active Semantic Documents, which is already deployed as an Active Semantic Electronic Medical Record application at the Athens Heart Center.

**SeNS** (Semantically Enabled Networking and Services): This new project seeks to take semantics to middleware and network level, starting with the definition and prototyping of semantic overlay network for scalable information dissemination.

October 26, 2005