

Research Statement

Jie Bao

My current research interests are in *semantic web* and *knowledge representation*, with a focus on the representation, scalable reasoning algorithms and collaborative construction of modular ontologies and privacy protection in knowledge reuse.

Current Research

Representation, Reasoning Algorithms and Collaborative Construction of Modular Ontologies

The success of the World Wide Web can be partially attributed to the *network effect*: the power of the Web is realized by *interlinked* web pages contributed by *independent* actors. Recent efforts to extend the web into a semantic web are aimed at enriching the web with machine interpretable content using ontologies. We expect the network effect will also play an important role in realizing the full potential of the semantic web: instead of a single, centralized ontology, it is much more natural to have multiple, interlinked, and distributed ontologies. Such ontologies represent the contextual, local knowledge of the ontology designers. Ideally, ontology languages will support localized and contextualized semantics, partial and selective reuse of ontology modules, and structured approaches to the collaborative construction of large ontologies.

To address these issues, I developed a novel approach, namely Package-based Description Logics (**P-DL**), to support modular ontologies, distributed inferences, and collaborative construction of ontologies. In particular, I have identified and investigated the following problems:

- **Modular Ontology Representation.** I have identified the minimal requirements for ensuring well-defined semantics in a modular ontology language. This allows combining multiple autonomously developed, context-dependent ontologies into an internally consistent knowledge base using the *semantic importing* mechanism offered by P-DL. The proposed language compares favorably to existing approaches in terms of expressivity and reasoning soundness for many inference tasks that are common in practice (e.g., on knowledge transitive reusability).
- **Distributed Reasoning Algorithms.** Due to the scale and distributed nature of web ontologies, reasoning with such ontologies differs from classic reasoning in that the combination of all ontology modules into a centralized knowledge base is costly, and in some cases (e.g., in peer-to-peer applications), simply impossible. I have developed approaches to dividing the reasoning process with P-DL into many local reasoning processes associated with each ontology module, thereby avoiding the need to construct a centralized ontology. My dissertation illustrates the application of this strategy to devise several federated, tableau-based algorithms for reasoning with different languages of the P-DL family. Such algorithms can potentially help to scale up reasoning in the case of large ontologies while respecting the relative independence or autonomy of individual ontology modules.
- **Collaborative Ontology Building.** Construction of a complex ontology is necessarily a collaborative process that involves cooperation among diverse and relatively autonomous ontology engineers. My dissertation offers a general strategy for building ontologies collaboratively, exploiting the notion of modularity. I have shown how to control conflicts among ontology engineers and inconsistencies between ontology modules. I have developed WikiOnt (a wiki-based ontology editor) and COB Editor, to support sharing, reuse, and collaborative editing of domain-independent ontologies or domain-specific ontologies (e.g., biological ontologies in COB Editor).

Privacy-Preserving Reasoning on the Semantic Web

Many semantic web applications require selective sharing of ontologies between autonomous entities in part due to privacy, copyright or security concerns. In such cases, an agent might want to hide a part of its ontology while sharing the rest. I have investigated a framework for privacy preserving reasoning. In the framework, an agent can safely answer queries against its knowledge base using both the hidden and visible parts of the knowledge base, without revealing the hidden knowledge. The research is aimed to:

- verify the correctness and consistency of web ontology security policies;
- avoid overly restrictive protection on data or knowledge;
- allow flexible safe usage of the same knowledge base to multiple users.

I have shown an application of this framework in the cases of the commonly used hierarchical ontologies and the description logic SHIQ. This approach is particularly well suited for applications (e.g., semantic web services and medical information systems) that call for privacy-preserving reasoning.

Research Agenda

I plan to continue my efforts on the following problems:

1. **Scalable inference with very large knowledge bases.** In particular, I will investigate distributed, incremental and approximate reasoning algorithms that may answer queries against web-scale knowledge bases in timely fashion. One of the immediate goals is to develop a scalable query answering engine for (potentially distributed) large triple databases, which is useful in web data integration, semantic web search engine and data mining from the web.
2. **Collaborative ontology construction and usage** by putting together the power of social web and semantic web. Social tagging (e.g., in flickr) and wiki-like environment provide practical tools to link knowledge contributed by millions of web users. I plan to investigate several critical problems in the coupling of social web and semantic web, including: What fragment of web ontology language is suitable to be used by common web users? How to create semantic links between ontologies using social web tools (e.g., wiki) and resources (e.g., Del.icio.us)? How to optimize ontology storage and query tools to meet the scalability challenge presented in such a setting? How semantic tagging may improve everyday activities (e.g., email)? I will also seek an active cooperation with industrial colleagues for the delivery of this research into practice.
3. **Privacy protection on semantic web.** I will extend my current work on privacy-preserving reasoning on the following problems: How to design a policy language that can address privacy in RDF graphs? What are formal methods to debug inconsistent security policies for web ontologies? How to handle heterogeneous policy expressions in different policy languages, e.g., KAoS and xACML? How can we ensure privacy protection in multi-agent setting in that information will not be used inappropriately and an agent that misuses information can be tracked and punished? Targeted applications include medical information system, social network and homeland security.

In summary, working with modular ontologies and privacy-preserving reasoning with ontologies are exciting and active research areas that offer solutions to many challenges spanning the areas of semantic web, knowledge representation, database, information integration, information security and bioinformatics. As one of the active researchers in this area, I believe my early experience puts me in a strong position to tackle some of those challenges.