SIBIOS Ontology: A Robust Package for the Integration and Pipelining of Bioinformatics Services

M. Mahoui, Z.B. Miled, S. Srinivasan, M. Dippold, B. Yang, N. Li

IUPUI, Indianapolis, USA

DILS 2006 **07-21-2006**

Outline

- Introduction & Background
- SIBIOS Architecture
- Ontology Design
- Ontology Deployment in SIBIOS
- Conclusions & Future Work

Introduction

- Biology increasingly becoming information driven
- Large amount of data generated need to be analyzed through a series of pipelines "*in-silico experiments*"
- Substantial involvement from researchers is required to work on the logistics of the workflow

Introduction

- Need for a workflow infrastructure to assist researchers in:
  - Selecting appropriate biological databases and analytical tools – *bioinformatics services*
  - Composing workflows
  - Enacting workflows

Introduction

- Challenges
  - Large number of available bioinformatics services – absence for a service discovery solution
  - Distributed and mostly http-based bioinformatics services
  - Data Heterogeneity of bioinformatics services
    - Different data structures, formats, entity labeling, etc.

Introduction

- Dimensions for Data integration approaches include:
  - Aim of integration
    - Portals (NCBI Entrez)
    - Complex querying systems (TAMBIS)
    - Workflow systems (MyGrid)
  - Integration techniques include:
    - Warehousing systems (NCBI Entrez)
    - Wrapper-Mediator systems (MyGrid)

Indiana University School of informatics @ IUPUI

Introduction

- Data Integration techniques
  - Warehouse approach
    - Global schema to reconcile heterogeneity between services
    - Simple solution but require a local copy of integrated services
  - Wrapper mediator approach
    - Wrappers specific to each service are applied to query services and extract data from remote service to the integrated system
    - Scale well with addition of new services but more sensitive to service changes
    - Often rely on *ontologies* as basis for the integration solution

Introduction

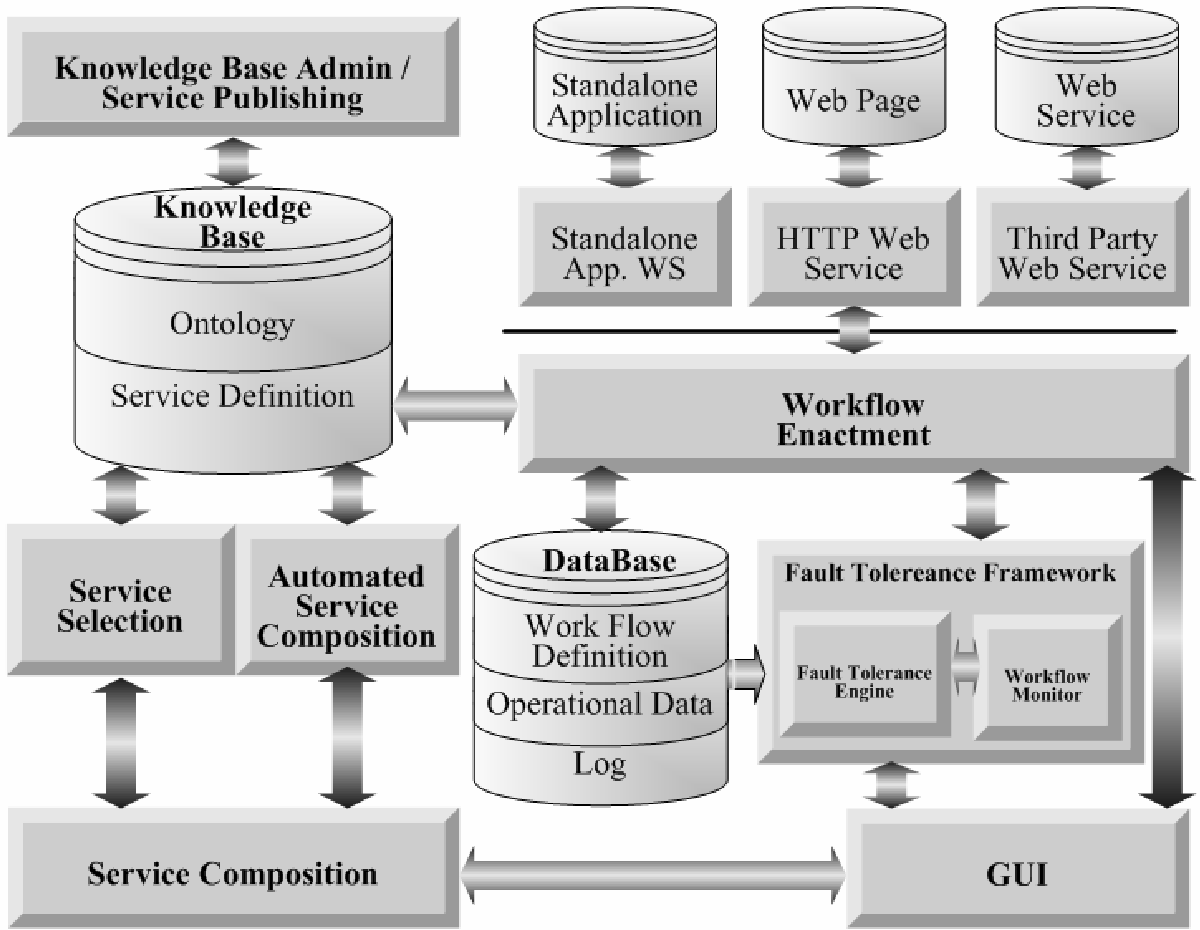
- Example of workflow environment for bioinformatics using a wrapper approach with an ontology: *SIBIOS*

The System for the Integration of BIOinformtics Servcies

Outline

- Introduction & Background
- SIBIOS Architecture
- Ontology Design
- Ontology Deployment in SIBIOS
- Conclusions & Future Work

SIBIOS Architecture



Outline

- Introduction & Background
- SIBIOS Architecture
- Ontology Design
- Ontology Deployment in SIBIOS
- Conclusions & Future Work

Ontology Design

- What's an ontology?
  - *The specification of conceptualizations, used to help programs and humans share knowledge."* *- Thomas R. Gruber*
  - A simplified and well defined view of a specific area of interest or domain
  - Provide semantic capabilities
  - Human and machine-readable

Ontology Design

- Service centric approach
  - Allow the definition of services at different levels
    - Service composition
    - Service invocation

Ontology Design

- Ontology Role at the service invocation
  - Provide a common terminology to describe input and output parameters to facilitate pipelining of services
  - E.g. EC (output of SwissProt) and accession (input of Genbank)

Ontology Design

- Ontology Role at the service composition
  - Provide service description that cater to researchers with different expertise in bioinformatics
  - Serve as a mapping model to ensure correct composition of services

Ontology Design

- Service description
  - Properties are most common way for describing concepts within an ontology
  - In SIBIOS a service can be described by five properties
    - Input (e.g. accession number)
    - Output (e.g. DNA sequence)
    - Task performed (e.g. sequence alignment)
    - Resources (e.g. database used for a tool)
    - Function/algorithm (e.g. Smith-Waterman algorithm)

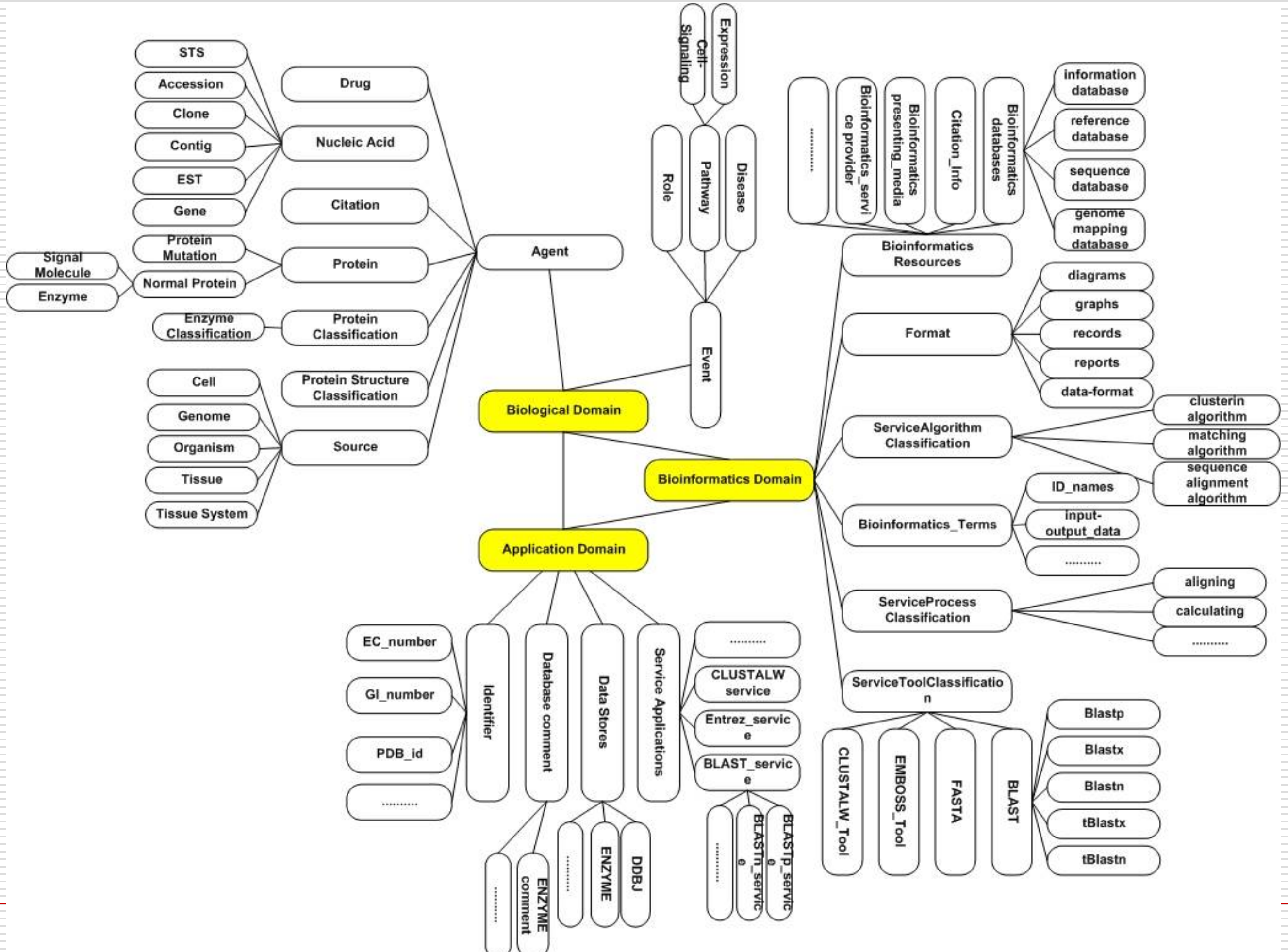
Ontology Design

- Design criteria
  - Scalable and easy to maintain ontology- by proposing a structure that can easily adapt to new concepts and services localizing projects that are likely to change together
  - Provide a hierarchical structure for concepts to enhance service selection composition capabilities

Ontology Design

- SIBIOS ontology serialization:
  DAML + OIL → OWL DL → OWL DL + OWL-S
- Using tools such as Protégé, WonderWeb, and RacerPro

Ontology Design

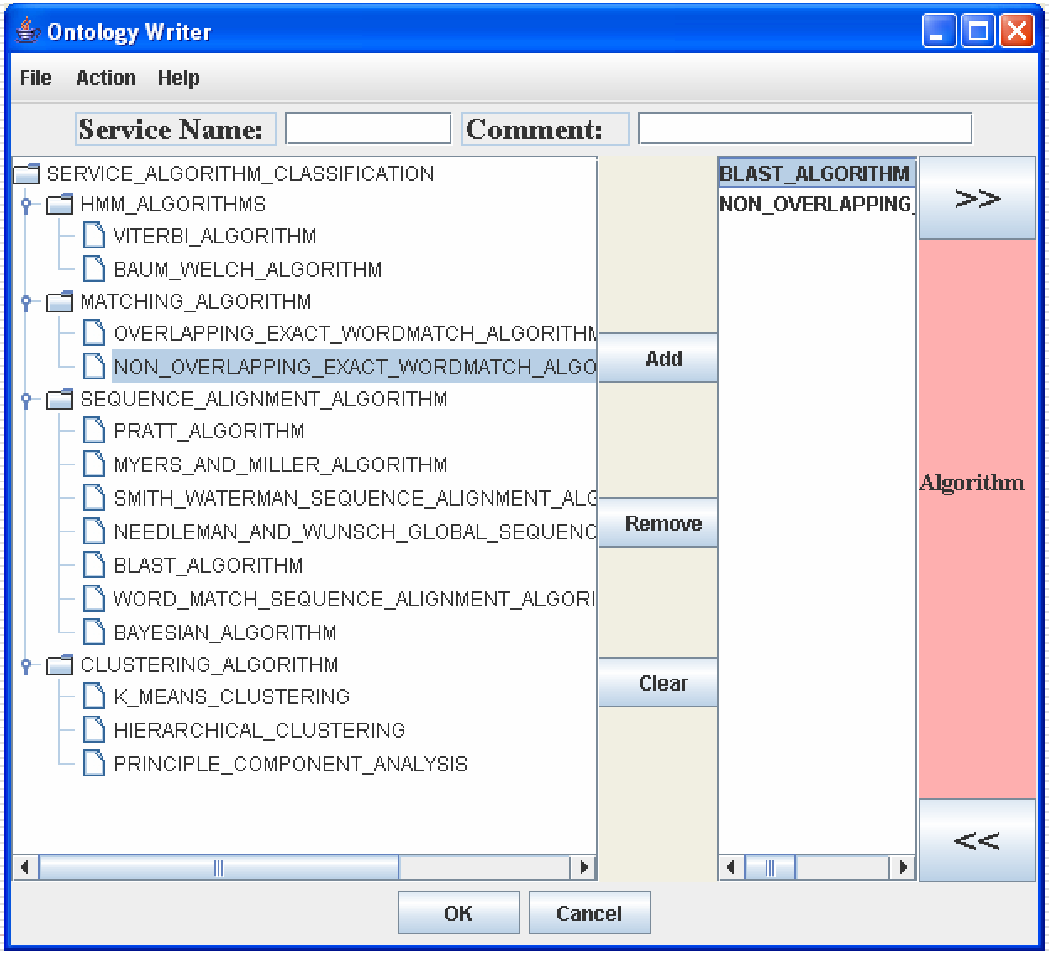


Ontology Maintenance

- SIBIOS ontology is designed as an open ontology
- Despite availability of editing tools, ontology update is still a complex process
- E.g. adding new service necessitates placing the service at the right level in the ontology, defining restrictions (properties)
- Need for user-friendly tools to facilitate adding/deletion of services

Ontology Maintenance

- *The ontology writer*: allow researchers to add/delete users without having to be exposed to low level operations



Outline

- Introduction & Background
- SIBIOS Architecture
- Ontology Design
- Ontology Deployment in SIBIOS
- Conclusions & Future Work

Ontology Deployment in SIBIOS

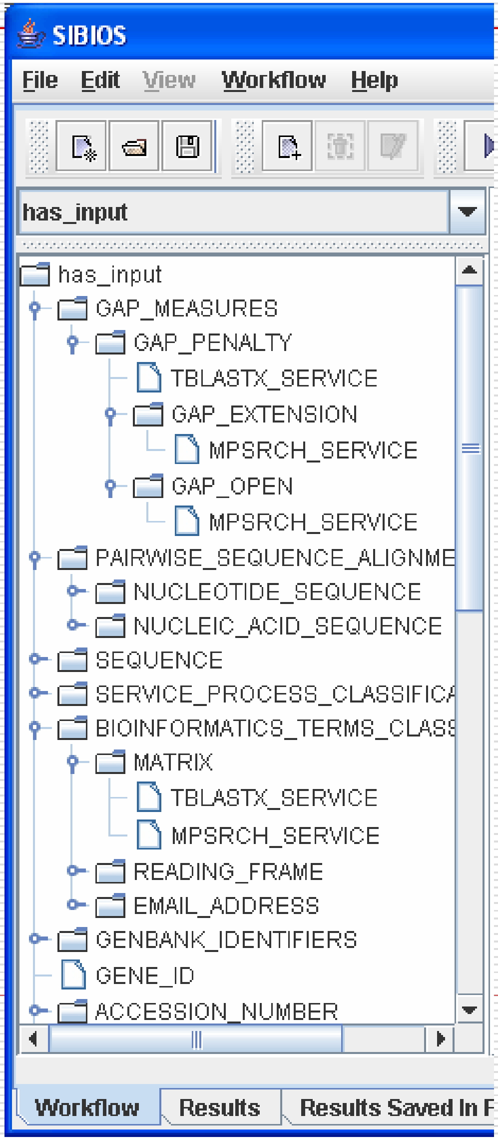
- Service selection
- Service composition
- Fault tolerance framework

Ontology Deployment in SIBIOS

- Service Selection
  - Provide a flexible system for service selection that tackle to researchers with different expertise
  - Two approaches:
    - Service browsing
    - Service discovery

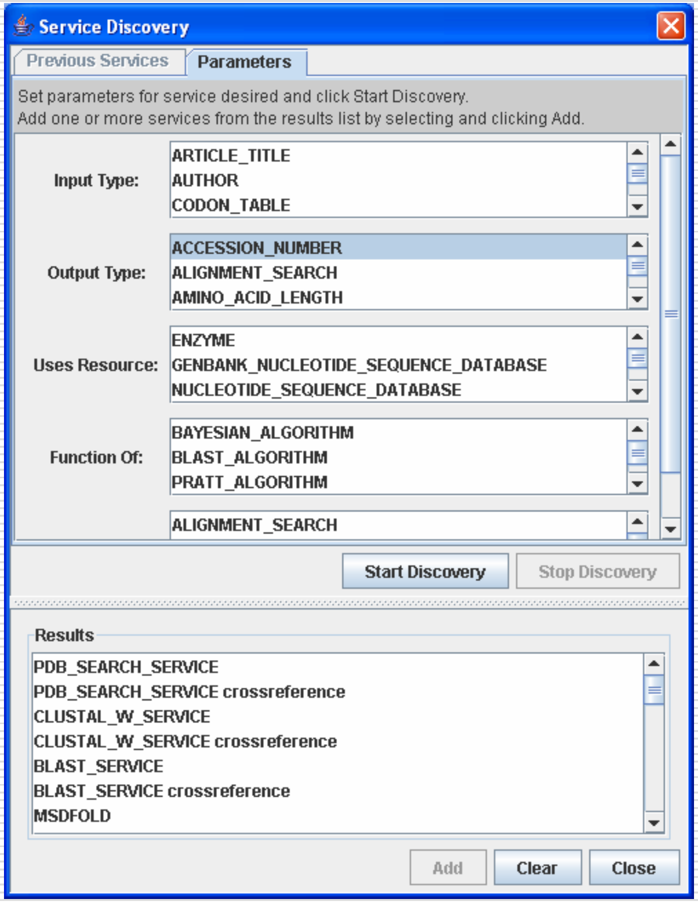
Ontology Deployment in SIBIOS

- Service browsing
  - Supported by classification of services based on each of their properties
  - Hierarchical structure of each property is available for browsing
  - Leaf nodes correspond to services
  - Property based classification is built dynamically using RacePro reasoning capabilities



Ontology Deployment in SIBIOS

- Service discovery
  - Advanced searching
  - Combine service properties (e.g. input and task)
  - For non starting services, the output of previous services replace the input parameter
  - RacerPro is used to infer services that match combined properties

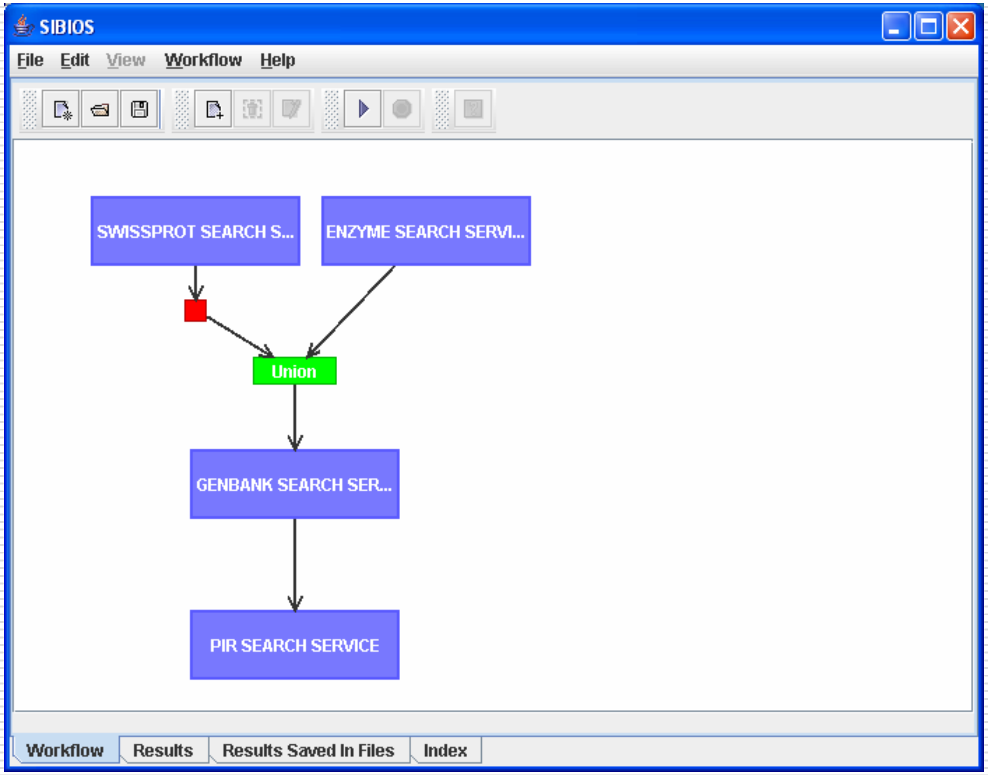


Ontology Deployment in SIBIOS

- Service Composition
  - Process of connecting services into meaningful workflow
  - Using service browsing together with service connectors as well as service discovery
  - Five connectors: primitive, UNION, INTERSECT, MINUS, CROSS

Ontology Deployment in SIBIOS

- Service Composition
  - Based on matching between input/output parameters
  - SIBIOS uses inclusive matching
  - E.g. service with sequence as output can be connected to a service with protein sequence as output – likewise for reverse situation
  - Rely on RacerPro reasoning capabilities to perform matching



Ontology Deployment in SIBIOS

- Fault tolerance framework
  - Increase the reliability of workflow enactment in SIBIOS
  - By decreasing the sensitivity of the workflow enactment due to service failure
  - Three main strategies for service failure:
    - Mirror service
    - Service replacement
    - Nested workflow

Ontology Deployment in SIBIOS

- Service replacement for service failure
  - Hypothesis: to replace a failed service, we only consider the invoked parameters for each property (vs. formal parameters)
  - Aim:
    - Increase the likelihood of finding replacement services using service properties
    - By defining different levels for matching replacement services

Ontology Deployment in SIBIOS

- Service replacement for service failure
  - Criteria for matching services
    - **Service level**: assign different weights for service properties with input and output having the highest weight
    - **Service property level**: consider the number of parameters for each property of the failed service that are present in the replacement service
    - **Service parameter level**: consider whether the subsumption relationship is used at the property parameter level
  - RacePro is used to infer information needed to check each criteria

Summary & future work

- SIBIOS co-exist with *very* interesting workflows management systems (e.g. Kepler, BioMOBY, InforSense, Taverna, Ubertool)
- SIBIOS project main distinguishing features:
  - Bioinformatics services include both http-based ones but also web services
  - Simple workflow infrastructure that cater to researchers with different expertise
  - Contribute to solve data integration issues (e.g. service replacement, automation support for wrapper generation)

Summary & future work

- Future work includes:
  - Description of services as semantic web services using OWL-S
  - Automated service composition
  - SIBIOS deployment to specific applications such as biomarker discovery using micro-array data
  - Include other properties for service description such as data quality

SIBIOS link:

- Please visit - http://sibios.engr.iupui.edu/

Acknowledgements

This project was supported in part by NSF CAREER DBI-DBI-0133946 and NSF DBI-0110854.

Thank You!

Questions

&

Comments