Semantic and Personalised Service Discovery

Phillip Lord, Chris Wroe, Robert Stevens, Carole Goble
University of Manchester

Simon Miles, Luc Moreau, Keith Decker, Terry Payne, Juri Papay
University of Southampton
myGrid

- A Grid middleware project for the bioinformatics domain
- Presentation of project as a whole on Thursday by Carole Goble
- Particular nuances of bioinformatics influence the process of Grid service discovery
Semantics

• Biologists apply their particular knowledge of the domain in judging service suitability which would not be automatically obvious from basic service advertisements in registries.

• For example, a service may be classified as a sequence alignment service with reference to an automatically parsable ontology.
Personalisation

• Variety of data sources and analysis tools for which individual biologists and their organisations have personal preferences

• For example, an organisation may recommend the use of particular database to its new members because it believes that database’s data to be most accurate
User Requirements - Semantics

• Discovery of services in a way most useful for bioinformaticians.

• The bioinformaticians:
  1. Would like to describe their desired functionality in one way, and discover applicable services that match when domain knowledge is taken into account.
  2. Know the type of data they wish to analyse, but don’t care about formats required by different services.
User Requirements - Personalisation

- The bioinformaticians:
  3. Would like their particular preferences regarding the suitability of one service over another taken into account
  4. Would like the recommendations of experienced bioinformaticians in the same organisation taken into account
Technical Requirements

• These user requirements must be met by providing extra information (metadata) with service descriptions.
  – Need to be able to publish, retrieve and reason over structured semantic description metadata associated with a given service (1)
  – Metadata relates to operation inputs and outputs as well as services as a whole (2)
  – Need to attach metadata personal and private to an individual or organisation (3, 4)
Architecture

Discovery by describing services required

User

Discovery Client

Semantic Find Component

Extract service descriptions to reason over

Personalised View Component

Pull service adverts from global registries

Service Registry

Service Registry

Service Registry

Personalised discovery using UDDI clients and publishing of personal metadata
Service Registries

• Examples include UDDI, LDAP, Jini…
• myGrid services is Web Services based at the moment, so UDDI used
• Multiple registries may be used by the community
Personalised Registry Views

- A *view* is a service that allows discovery over federated remote registries.
- Metadata can be attached to many parts of a service description such as the inputs and outputs of service operations.
- A personalised registry view can be public or deployed by and private to an individual or organisation.
View Information Model

• The data in a personalised registry view is composed into several information models supporting different specifications, e.g. UDDI, DAML-S, WSDL, stored as RDF allowing arbitrary extension with metadata.

• See Luc Moreau’s talk tomorrow for more details on views and their information models.
View Information Model Examples

- WSDL Document
  - WSDL Service
    - WSDL Operation
      - WSDL Message
  - WSDL Input
  - WSDL Output
- UDDI Business Entity
  - UDDI Business Service
  - UDDI Binding Template

Semantic Description of Service
May be references to ontology terms, or structured metadata in the form of RDF triples

Semantic Description of Output Type
Semantic Find Component

• A *semantic find component* provides discovery over domain-specific descriptions by reference to domain ontologies

• It uses the descriptive metadata stored in views to reason over the suitability of a service using a pre-constructed ontology
• If the user wishes to find services that will accept gene sequence data as input, the find component will reason that services accepting more general data of that form, e.g. all DNA sequences, will also be applicable.
Discovery Client

- User interface that is integrated into the myGrid WorkBench
- The discovery client allows the user to make use of the discovery capabilities of the views and the semantic find component through a unified portal
- At the moment, the discovery client can use the find component to reason over the services in a view and categorise them for browsing by the user, and can be used to find workflows by their input types
Discovery Client - Browser
Publishing Service Descriptions

- Services can be published in UDDI or as normal for Web Services
- Views pull these adverts into their own local stores
- Structured and personal metadata, including decomposed WSDL interfaces, can then be attached to the service descriptions in a view
Semantic Service Discovery

- The user presents the discovery client with queries, such as asking for tools that analyse a particular type of data and that are recommended by experts in their organisation
- The find component extracts descriptions from the organisation’s view, reasons over them using a domain ontology and determines which services are applicable
- The details of each service are extracted from the view and returned to the user for selection
Conclusions

• Using the myGrid architecture biologists will be able to perform sophisticated discovery that takes account of domain knowledge and personal preferences.

• We have developed personalised views over existent registries, semantic reasoning for sophisticated discovery and a client to allow a user or other services to browse or search.

• These service discovery mechanisms are used as part of the whole myGrid architecture for running personalised in-silico experiments.