SeaLife: A Semantic Grid Browser for the Life Sciences Applied to the Study of Infectious Diseases

From Bioinformatics Services to a Semantic Web for the Life Sciences

Albert Burger

MRC Human Genetics Unit

Department of Computer Science, Heriot-Watt University

Overview

- Project participants;
- Objectives;
- Use Cases;
- Main components of Sealife;
- Existing technology;
- Conclusion;

Project Reference

- **A Semantic Grid Browser for the Life Sciences Applied to the Study of Infectious Diseases;**
- IST-2006-027269
- Call: IST-2004-2.4.11 Integrated biomedical information for better health;
- Priority: Sixth Framework Programme, Priority 2, Information Society Technologies
- Type: Specific targeted research project;
- Start date: April 2006;
- End date: March 2009;

Sealife Partners

Sealife Partners

- Michael Schroeder, Biotec, TU Dresden, Germany (Coordinator)
- Albert Burger, MRC Human Genetics Unit; Dept. of Computer Science, Heriot-Watt University, Edinburgh, UK
- Patty Kostkova, City eHealth Research Center, City University, London, UK
- Robert Stevens, BioHealth Informatics Group, University of Manchester, UK
- Bianca Habermann, Scionics, Dresden, Germany
- Rose Dieng-Kuntz, Acacia research group, Inria Sophia-Antipolis, France

Edinburgh SeaLife People

- Karen Sutherland (RA);
- Kenneth McLeod (RA);
- Panagiotis Papakos (MSc 2006);
- Albert Burger (PI Sealife);
- Richard Baldock (PI Mouse Atlas and Sealife Advisory Board);

Objectives

- Conception and realisation of a **Semantic Grid Browser**, which **links** the **current Web to** the emerging **eScience infrastructure**
  - **Many grids, few users**: make Web servers and services accessible to end users
  - **Semantic Hyperlinks**: use ontologies and background knowledge to map web contents to services
  - **Shopping cart**: service composition and enactment module
- **Applications: from cells via tissue to patients**
  - Evidence-based medicine
  - Patent and literature mining
  - Molecular biology

Literature and Patent Mining

- user is browsing a patent database;
- comes across statement: "*An improved infant formula is described which includes a phospholipid metabolism*."
- SeaLife browser identifies the term '*phospholipid metabolism'* and offers a definition: "*The chemical reactions and physical changes involving phospholipids,* ...";
- It also identifies 'human' in its taxonomy and offers to show all human proteins in the UniProt database, which are involved in phospholipid metabolism;

Evidence-based Medicine

- clinician consults the national electronic library of infections (NeLI);
- clinician is guided by SeaLife browser to further relevant information sources, e.g. ENSEMBL and the Protein Databank;
- for example, the browser might recognise ‘hepatitis’ as a disease and offer to query ENSEMBL to find out more about the genetics relating to hepatitis;

Molecular Biology

- Consider statement: “*Rabaptin-5 interacts with the small GTPase Rab5 and is an essential component of the fusion machinery for targeting endocytic vesicles to early endosomes*”;
- SeaLife browser identifies ``*Rabaptin-5*'' and ``*Rab5*'' as protein names, ``*endocytosis*'' as biological process, and ``*early endosome''* as cellular component;
- Browser might offer sequence search for Rab5;
- If other sequences have been “collected”, browser might offer a domain search of these sequences, displaying a multiple sequence alignment;

Mouse Atlas

- first Mouse Atlas use case meeting: 16 June 2006;
- to form part of larger molecular biology use case;
- some initial ideas:
  - browser identifies mouse gene in web page (e.g from PubMed) and offers to retrieve tissues from MA that express gene (and vice versa);
  - browser identifies gene expression in mouse and offers to find human tissues that express homologous genes;
- suggestions made:
  - items in shopping cart should have provenance data attached;
  - user may want multiple shopping carts (for different investigations);

SeaLife Main Components

- Ontologies:
  - background knowledge for browser;
- Text Mining and Concept Mapping:
  - bridging gap between free text on current web and ontology-based mark-up for Semantic Web;
  - developing automated mark-up for free text, based on text-mining and natural language processing techniques;
- Service Composition:
  - bridging the gap between ontologies of the Semantic Web and Web/Grid services in the Life Sciences;
  - “shopping cart” management;

SeaLife Components Interaction

User

SeaLife

SeaLife: Shopping Cart

SeaLife: Concept Mapping

SeaLife: Browser

SeaLife: Task Composition

SeaLife: Ontologies

Current Web

Web Server | ... | Web Server

eScience Infrastructure

Web/Grid Service | ... | Web/Grid Service

SeaLife Components Interaction

User

Step 1:
User loads web page into browser.

SeaLife:
Shopping Cart

SeaLife:
Concept Mapping

SeaLife:
Browser

SeaLife:
Task Composition

SeaLife:
Ontologies

①

Current Web

Web Server | ... | Web Server

eScience Infrastructure

Web/Grid Service | ... | Web/Grid Service

SeaLife Components Interaction

User

Step 2:
Browser uses concept mapper to identify ontological concepts and highlights terms in page.

SeaLife: Shopping Cart

SeaLife: Concept Mapping

②

SeaLife: Browser

SeaLife: Task Composition

SeaLife: Ontologies

Current Web

Web Server | ... | Web Server

eScience Infrastructure

Web/Grid Service | ... | Web/Grid Service

SeaLife Components Interaction

User

Step 3:
User selects terms/concepts and puts them in shopping cart.

③

SeaLife:
Shopping Cart

SeaLife:
Concept Mapping

SeaLife:
Browser

SeaLife:
Task Composition

SeaLife:
Ontologies

Current Web

Web Server | ... | Web Server

eScience Infrastructure

Web/Grid Service | ... | Web/Grid Service

SeaLife Components Interaction

User

Step 4:
Task Composition Manager (TCM) determines available (combinations of) services based on shopping cart.

SeaLife: Shopping Cart

④

SeaLife: Concept Mapping

SeaLife: Browser

SeaLife: Task Composition

SeaLife: Ontologies

Current Web

Web Server | ... | Web Server

eScience Infrastructure

Web/Grid Service | ... | Web/Grid Service

SeaLife Components Interaction

User

Step 5:
User selects services of interest.
Browser asks TCM to execute service(s).

5a

5b

SeaLife: Shopping Cart

SeaLife: Concept Mapping

SeaLife: Browser

SeaLife: Task Composition

SeaLife: Ontologies

Current Web

Web Server | ... | Web Server

eScience Infrastructure

Web/Grid Service | ... | Web/Grid Service

SeaLife Components Interaction

User

Step 6:
TCM executes services and returns result to browser.

SeaLife: Shopping Cart

SeaLife: Concept Mapping

SeaLife: Browser

SeaLife: Task Composition

6b

SeaLife: Ontologies

6a

Current Web

Web Server | ... | Web Server

eScience Infrastructure

Web/Grid Service | ... | Web/Grid Service

SeaLife Components Interaction

User

Step 7:
Browser displays output to user.

7

SeaLife:
Shopping Cart

SeaLife:
Concept Mapping

SeaLife:
Browser

SeaLife:
Task Composition

SeaLife:
Ontologies

Current Web

Web Server | ... | Web Server

eScience Infrastructure

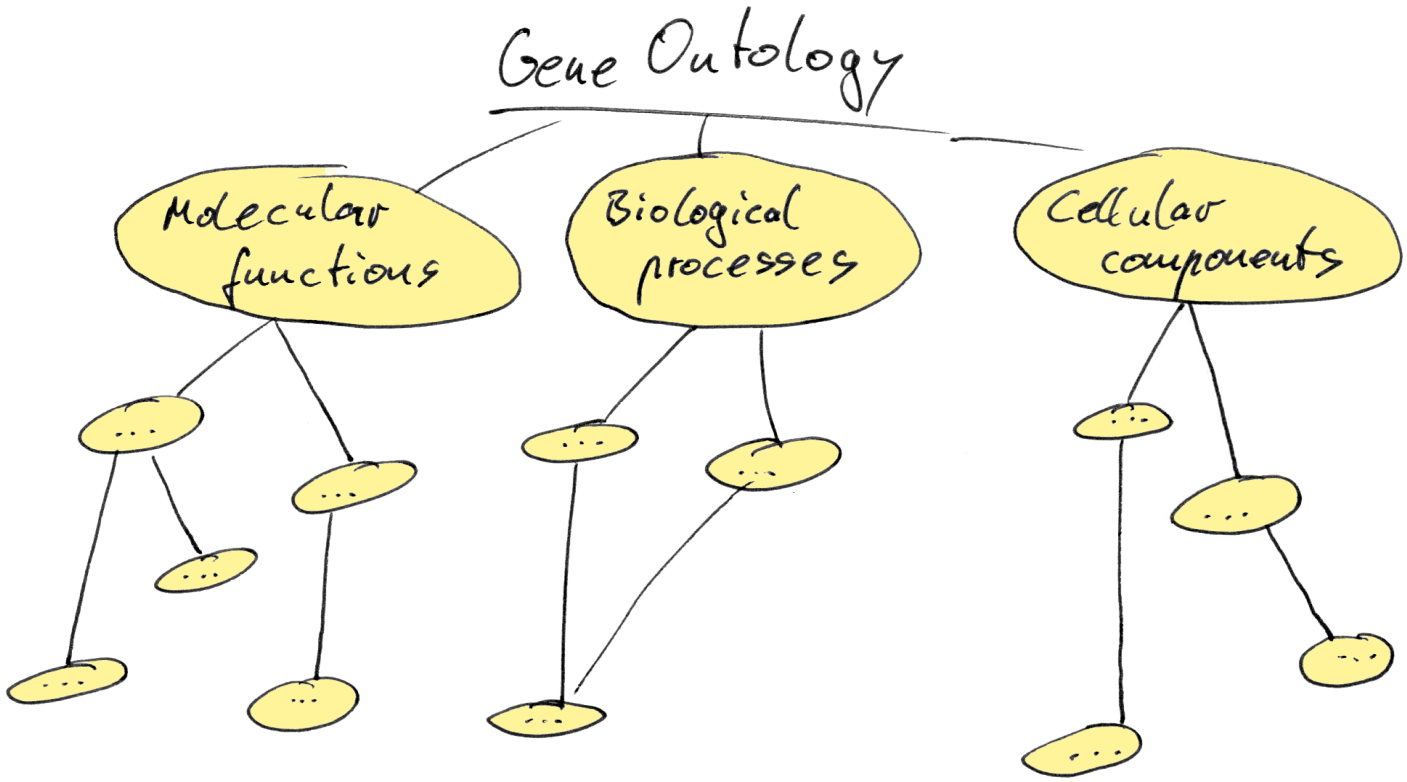
Web/Grid Service | ... | Web/Grid Service

Using existing Technology

SeaLife is building on existing technology:

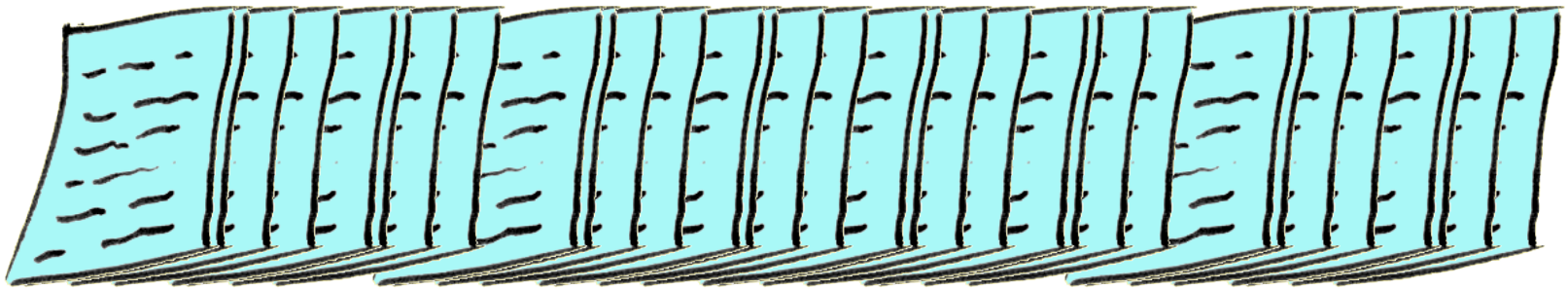
- GoPubMed: ontology-based literature search (Dresden University);
- COHSE: Conceptual Open Hypermedia Services Environment (Manchester University);
- myGrid: UK e-Science project for distributed middleware services in bioinformatics (Manchester University);

GoPubMed uses the GeneOntology as background knowledge.

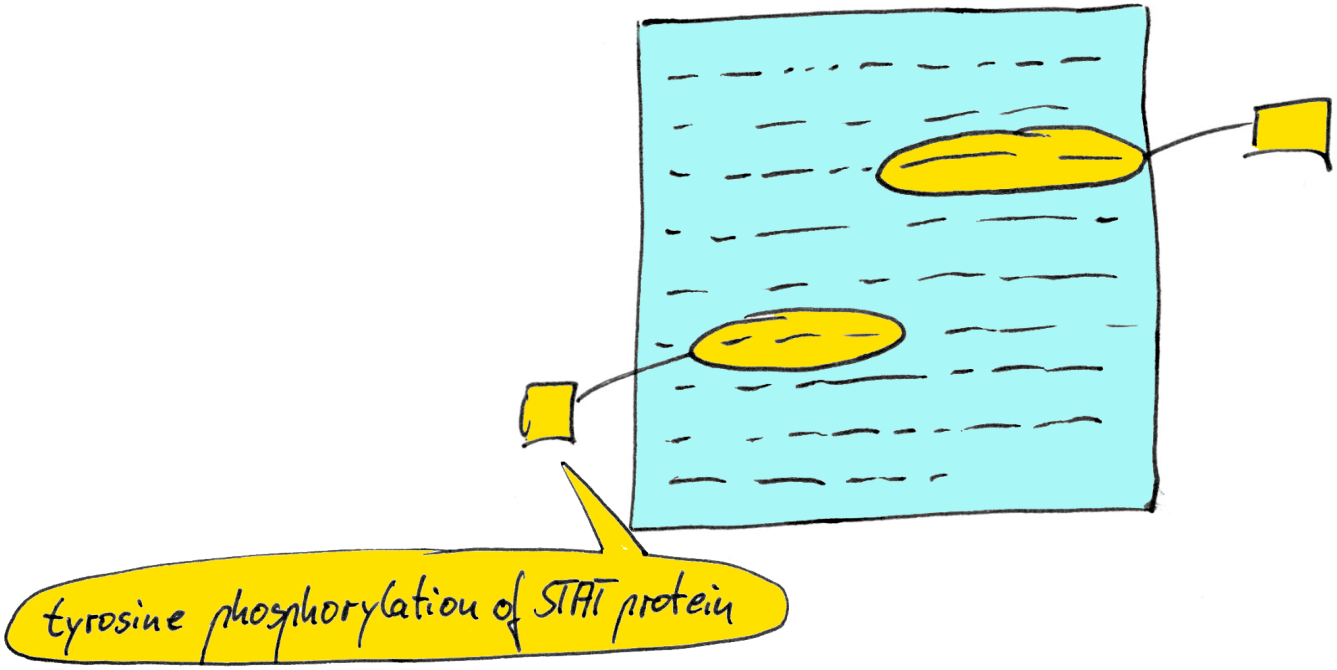


Your PubMed search: „Levamisole inhibitor"

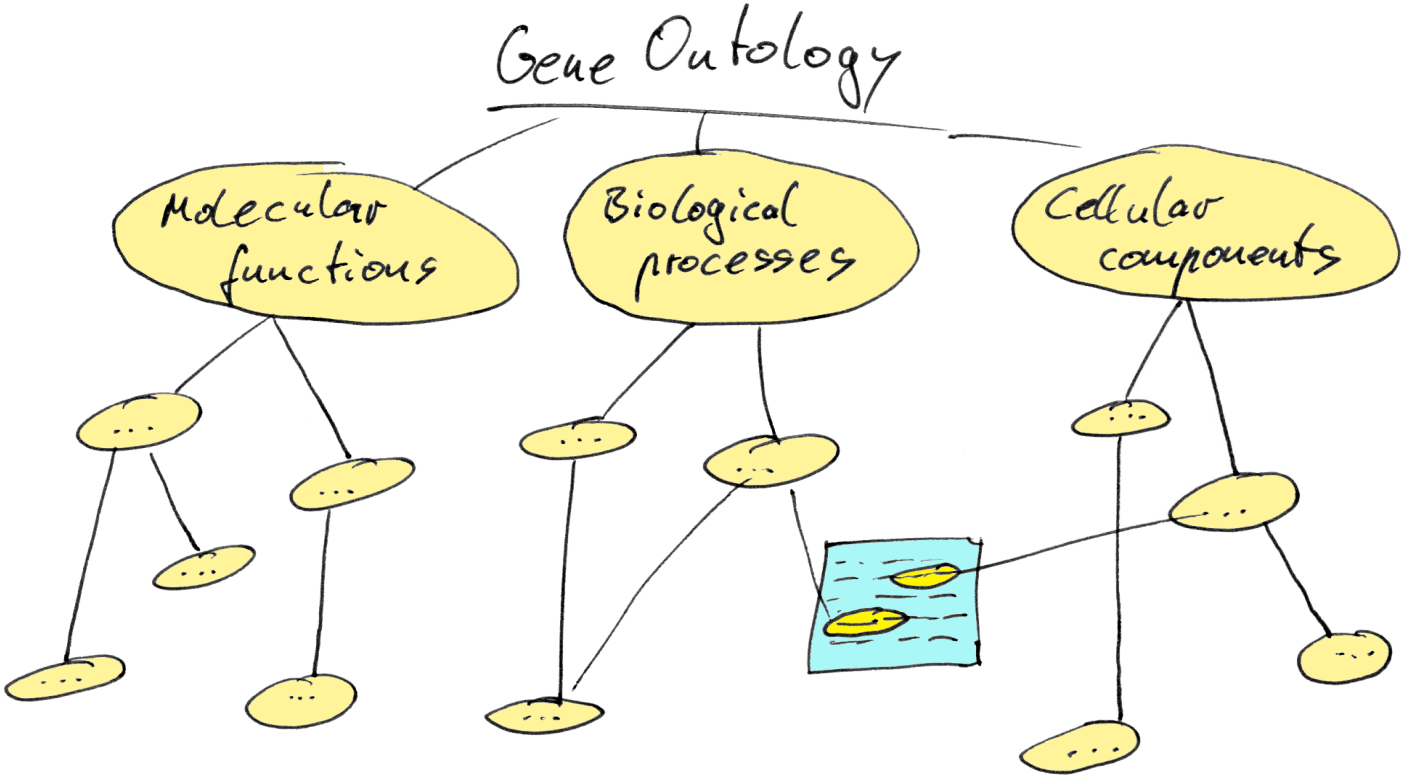
long result list in PubMed



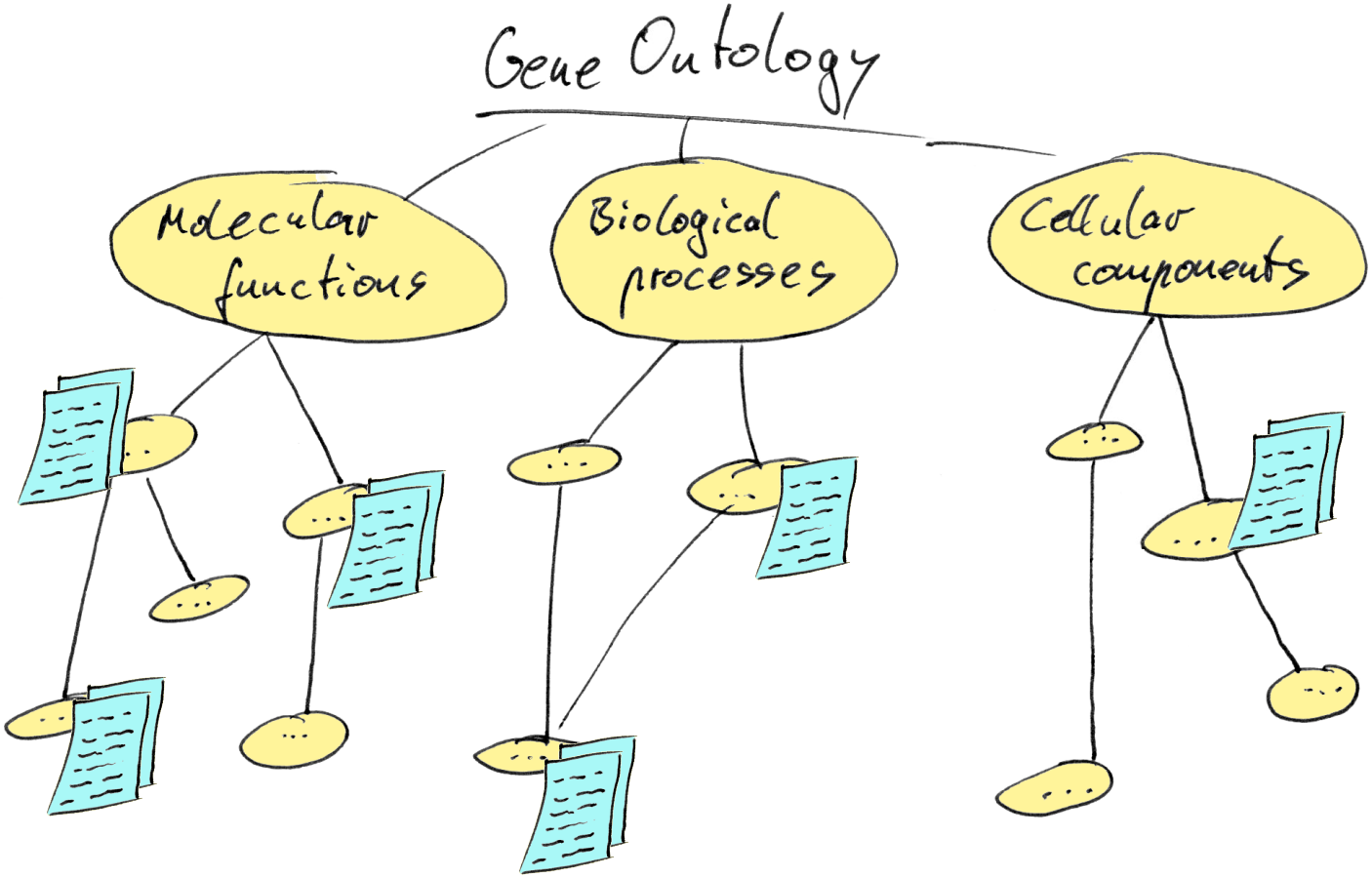
Term extraction takes place



One document may contain several ontology terms

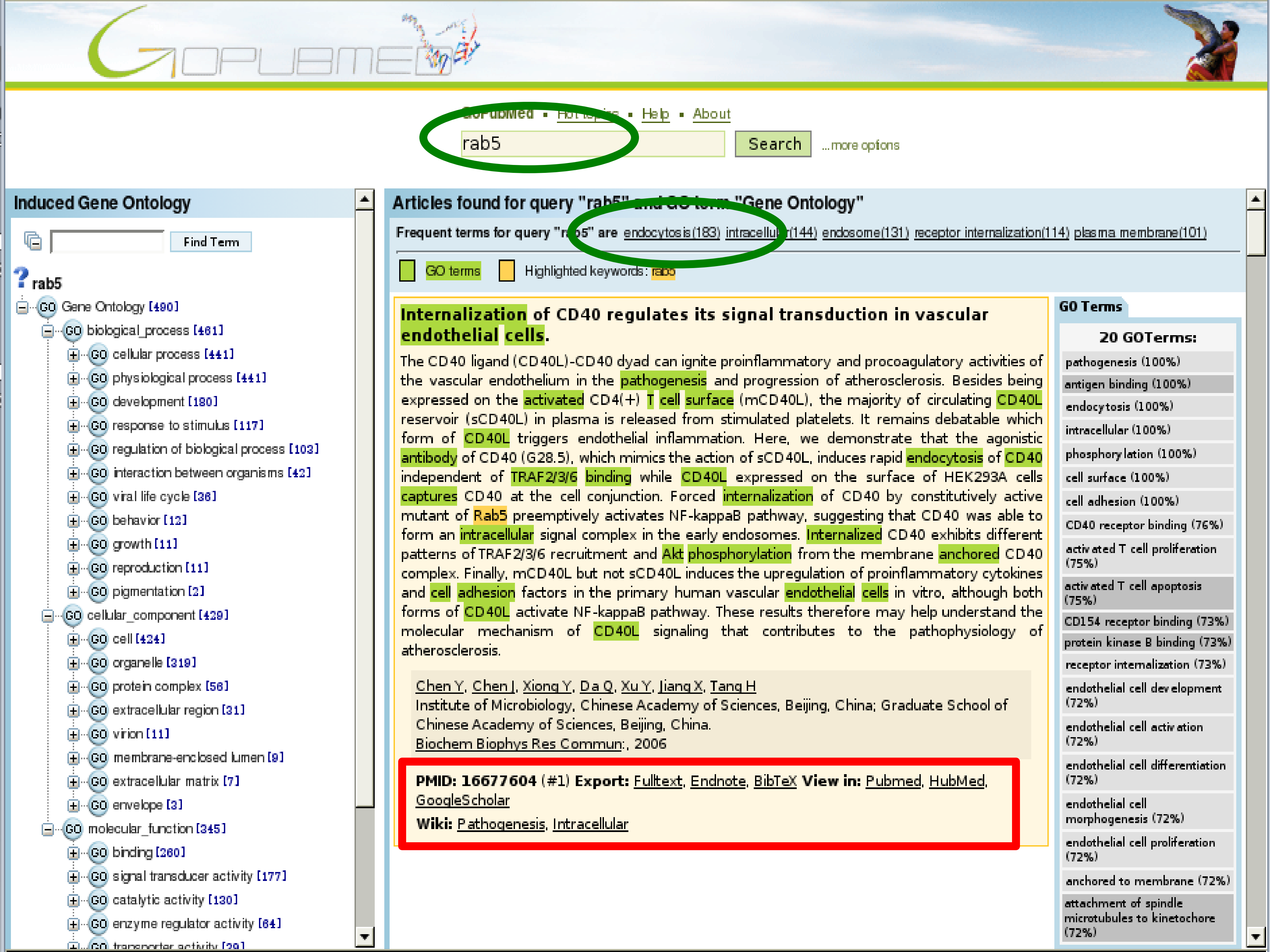


Documents are assigned to annotated concepts in the ontology



GoPubMed Example

Which cellular component and process is Rab5 associated with?



GOPUBMED
Help • About
rab5
Search
...more options
Induced Gene Ontology
Find Term
rab5
GO Gene Ontology [490]
GO biological_process [461]
GO cellular process [441]
GO physiological process [441]
GO development [180]
GO response to stimulus [117]
GO regulation of biological process [103]
GO interaction between organisms [42]
GO viral life cycle [36]
GO behavior [12]
GO growth [11]
GO reproduction [11]
GO pigmentation [2]
GO cellular_component [429]
GO cell [424]
GO organelle [319]
GO protein complex [56]
GO extracellular region [31]
GO virion [11]
GO membrane-enclosed lumen [9]
GO extracellular matrix [7]
GO envelope [3]
GO molecular_function [345]
GO binding [260]
GO signal transducer activity [177]
GO catalytic activity [130]
GO enzyme regulator activity [64]
Frequent terms for query "rab5" are endocytosis(183) endosome(131) receptor internalization(114) plasma membrane(101)
GO terms
Highlighted keywords: rab5
Internalization of CD40 regulates its signal transduction in vascular endothelial cells.
The CD40 ligand (CD40L)-CD40 dyad can ignite proinflammatory and procoagulatory activities of the vascular endothelium in the pathogenesis and progression of atherosclerosis. Besides being expressed on the activated CD4(+) T cell surface (mCD40L), the majority of circulating CD40L reservoir (sCD40L) in plasma is released from stimulated platelets. It remains debatable which form of CD40L triggers endothelial inflammation. Here, we demonstrate that the agonistic antibody of CD40 (G28.5), which mimics the action of sCD40L, induces rapid endocytosis of CD40 independent of TRAF2/3/6 binding while CD40L expressed on the surface of HEK293A cells captures CD40 at the cell conjunction. Forced internalization of CD40 by constitutively active mutant of Rab5 preemptively activates NF-kappaB pathway, suggesting that CD40 was able to form an intracellular signal complex in the early endosomes. Internalized CD40 exhibits different patterns of TRAF2/3/6 recruitment and Akt phosphorylation from the membrane anchored CD40 complex. Finally, mCD40L but not sCD40L induces the upregulation of proinflammatory cytokines and cell adhesion factors in the primary human vascular endothelial cells in vitro, although both forms of CD40L activate NF-kappaB pathway. These results therefore may help understand the molecular mechanism of CD40L signaling that contributes to the pathophysiology of atherosclerosis.
Chen Y, Chen J, Xiong Y, Da Q, Xu Y, Jiang X, Tang H
Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; Graduate School of Chinese Academy of Sciences, Beijing, China.
Biochem Biophys Res Commun:, 2006
PMID: 16677604 (#1) Export: Fulltext, Endnote, BibTeX View in: Pubmed, HubMed, GoogleScholar
Wiki: Pathogenesis, Intracellular
GO Terms
20 GOTerms:
pathogenesis (100%)
antigen binding (100%)
endocytosis (100%)
intracellular (100%)
phosphorylation (100%)
cell surface (100%)
cell adhesion (100%)
CD40 receptor binding (76%)
activated T cell proliferation (75%)
activated T cell apoptosis (75%)
CD154 receptor binding (73%)
protein kinase B binding (73%)
receptor internalization (73%)
endothelial cell development (72%)
endothelial cell activation (72%)
endothelial cell differentiation (72%)
endothelial cell morphogenesis (72%)
endothelial cell proliferation (72%)
anchored to membrane (72%)
attachment of spindle microtubules to kinetochore (72%)

The COHSE Project

Web Infrastructure

Open Hypermedia + Semantic Web

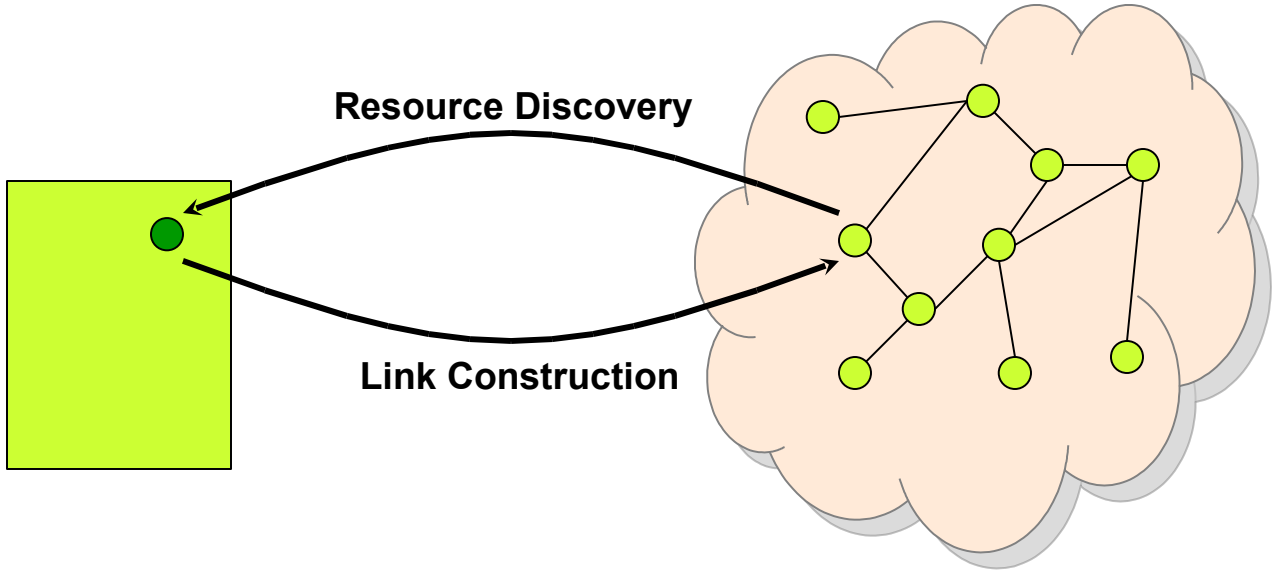
COHSE

- The *separation* of links from document;
- The reader is given the same access as the author;
- Integration of *third-party* documents and information.

- Explicit modelling of the world (e.g., ontologies);
- Knowledge *representation* and *reasoning*;
- Machine-processable semantics for improving navigation;

COHSE Philosophy

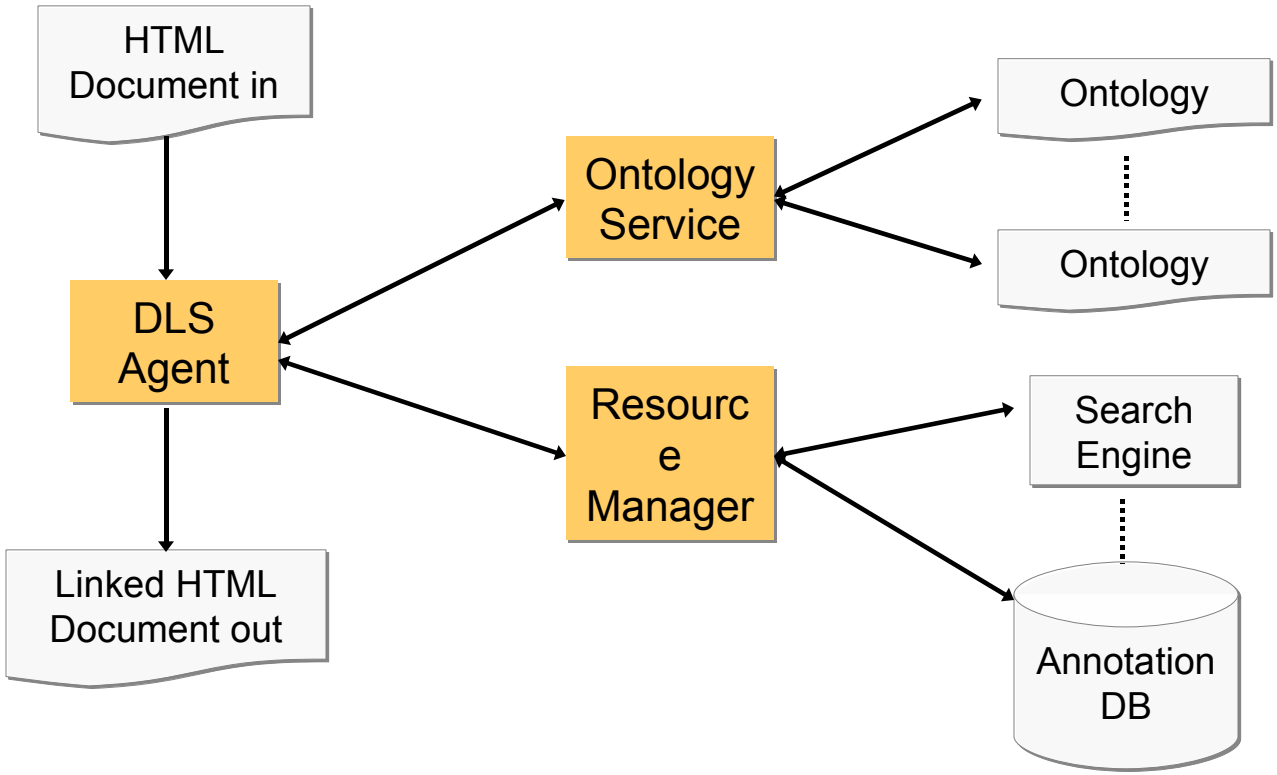
- Metadata can provide a mechanism not only for the support of resource discovery, but also for the provision of source anchors.
- Metadata allows both linking **into** and **out of** a resource.



COHSE System

- A software agent that generates and presents links on behalf of both an author and a reader.
- Used at browsing time, provides just-in-time hypertext.
- Used at authoring time, supports the authoring task.
- Link Creation, not just Resource Discovery.
- Different domain ontologies can be used to create different hypertexts (different view of the world).
- Creates an open Web - author has the same access as the reader.

COHSE's Architecture



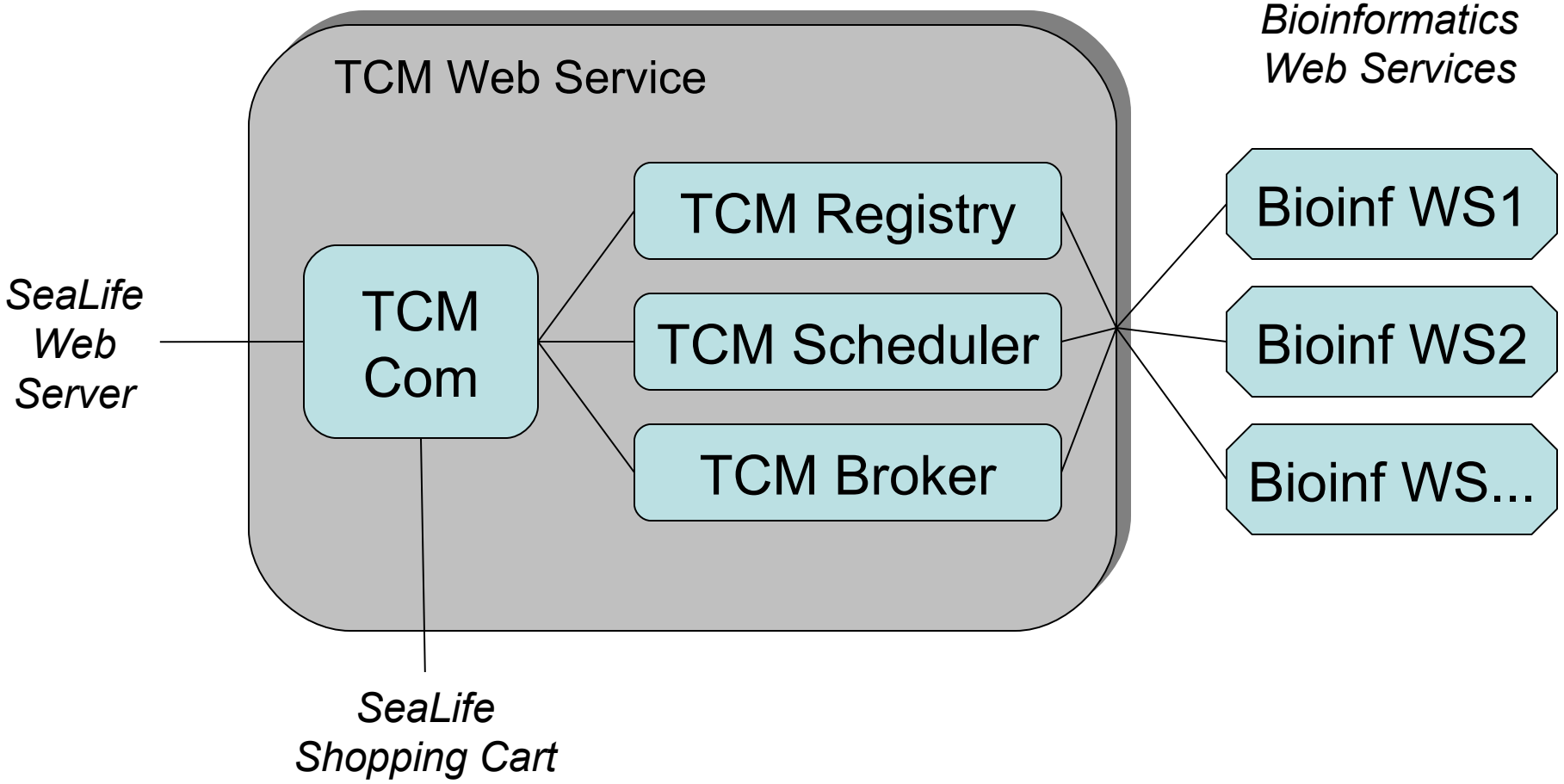
myGrid

- UK e-Science project for distributed middleware services in bioinformatics
- SeaLife plans to use:
  - for semantic services: FETA, Pedro;
  - for workflow enactment: FreeFluo;
  - for brokering: services ontology;

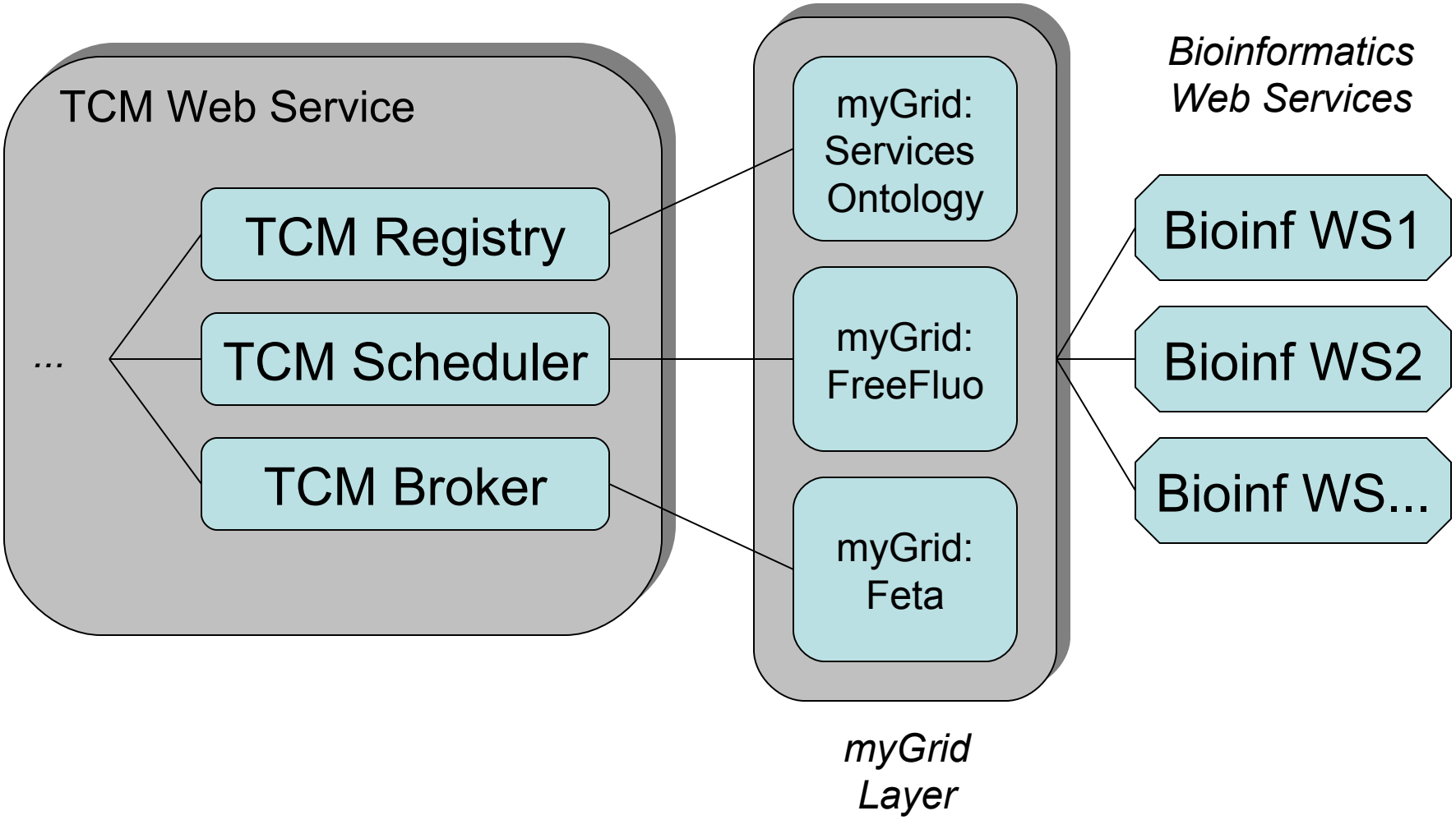
SeaLife Task Composition Manager

- responsible for linking to the bioinformatics web/grid services infrastructure;
- interacts with:
  - SeaLife Web Server;
  - Shopping Cart;
  - Web/Grid Services;
- responsible for identifying available Web/Grid services based on shopping cart content;

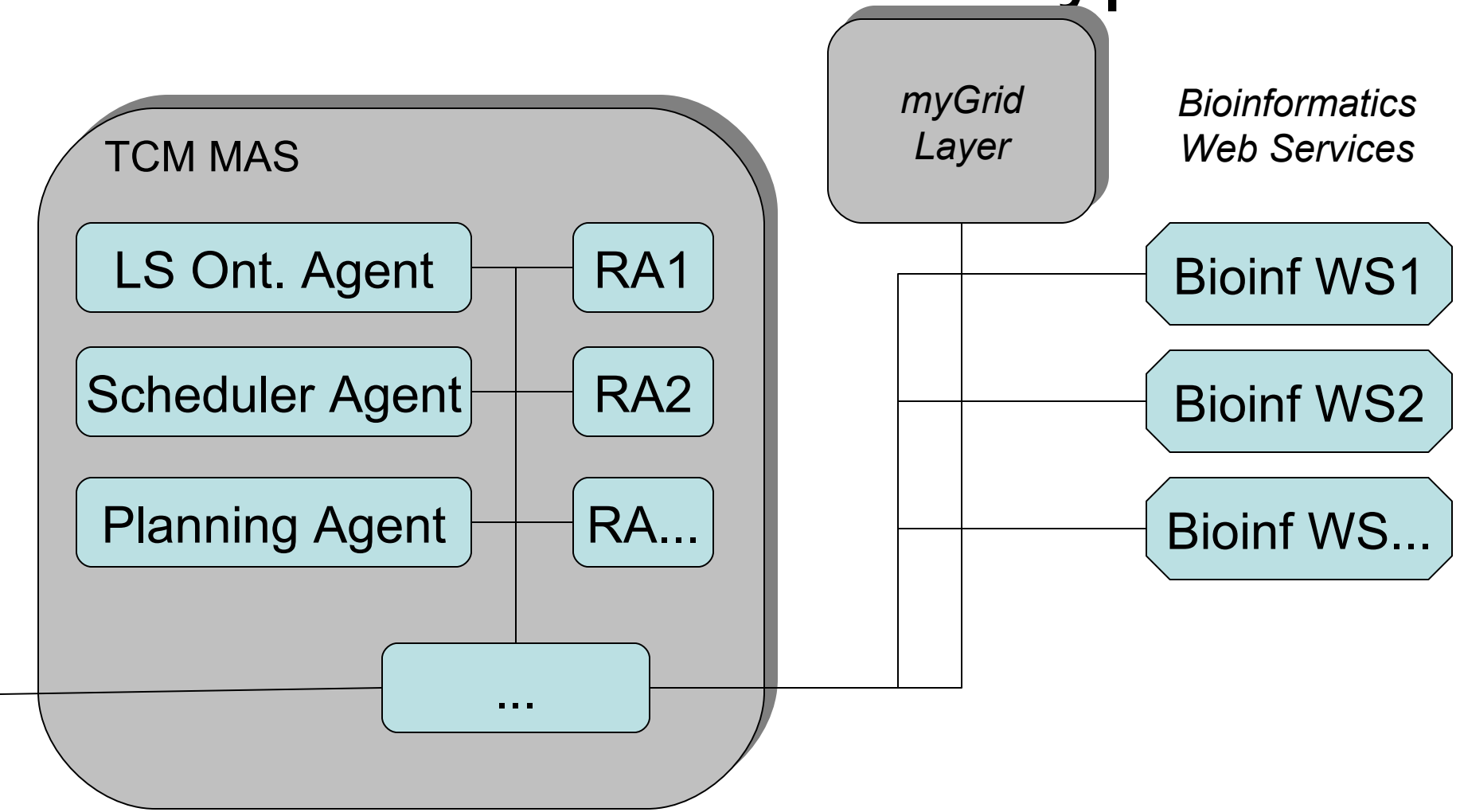
TCM: Initial Prototype



TCM: myGrid Prototype



TCM: MAS Prototype



Conclusion

- SeaLife: A Semantic Web/Grid Browser;
- Linking the Web and Web/Grid Services for the Life Sciences;
- Main areas:
  - Ontologies;
  - Text mining for concept mapping;
  - Service Composition;
- Building on existing technology;
- Towards Agent Systems;

Acknowledgements

- Slides have been contributed by various members of the SeaLife project consortium.
- SeaLife Web Site:

http://www.biotec.tu-dresden.de/sealife/