A Multi-Agent System Architecture for Knowledge Integration in Microarray Analysis

Miguel Vázquez
*Universidad Complutense de Madrid*
With: Pavón, Sanz, Boley, Lam, Fahrenkrug

Outline

- Scenario
  - What is Microarray analysis
  - How is it done
- Our proposal
  - How it can be improved
  - Why use agents
  - Architecture
Genomics

- Genes are DNA sequences that are in charge of the production of proteins, responsible biological functions of the cell
- Expression of genes at a given time indicate the production of proteins, and thus, the biological activity taking place

Microarray Analysis

- Used to measure the expression of thousands of genes in a given tissue
- Statistical analysis is used to determine the set of genes significantly expressed
Biological Analysis

- From the set of significantly expressed genes we draw biological conclusions
  - Relation between the genes
  - Functional activities they are involved in
  - ...
- Some genes are well documented, but many are not

Analysis cycle
Analysis cycle II

- Many work has been done in microarray design and statistical analysis
- Focus is shifting towards issues regarding biological analysis
- Let us take a closer look ...

Biological Analysis II

- Key word: Annotations
- Different type of information, about different organisms is stored in different databases and in different formats
- Knowledge can be shared
  - Relations between information types: genomics ~ proteomics
  - Different organisms share most of their genes
Where is the information?

- **Databases:**
  - GO: Gene Ontology
  - Swiss-Prot: Protein database
  - Kegg pathways
  - Organism specific databases

- **Literature:**
  - HUGE number of publications
  - Pubmed: Medical publication resource. Indexes and annotates publications

How this works?

- Curation, is the manual annotation of genes. This is done by reviewing the literature.
- Curation has a hard time caching up with the fast pace of publications
- Information grows in many different sources and integration is very loose
- Its hard and cumbersome to find information, many databases to check, literature to review
How we can help

- Improving the information flow
  - Ease access and integration of information
  - Allow results from each of the tasks in the analysis cycle to aid the others
- Providing a discovery driven workflow
  - Making conclusions in each step to suggest new hypothesis to test

Analysis workflow diagram
What we propose

- Design a framework that covers all the steps involved in the microarray analysis cycle.
- Provide tools to access and process different information sources
- Allow for information to flow between each of the steps

Agents

- Agents appear as a natural option for this implementation
  - Encapsulate logic for each information source and task
  - Design a common language to share information between them
What needs to be done

- Design the common language to represent the information in the system
- Implement agents for each particular task
- Develop new algorithms to process the information
- Provide a usable framework to the biologist

Questions?

- ...