DILS 2006

July 20, 2006

BioFacets: Solution Towards Leveraging the Wealth of Online Biological Databases

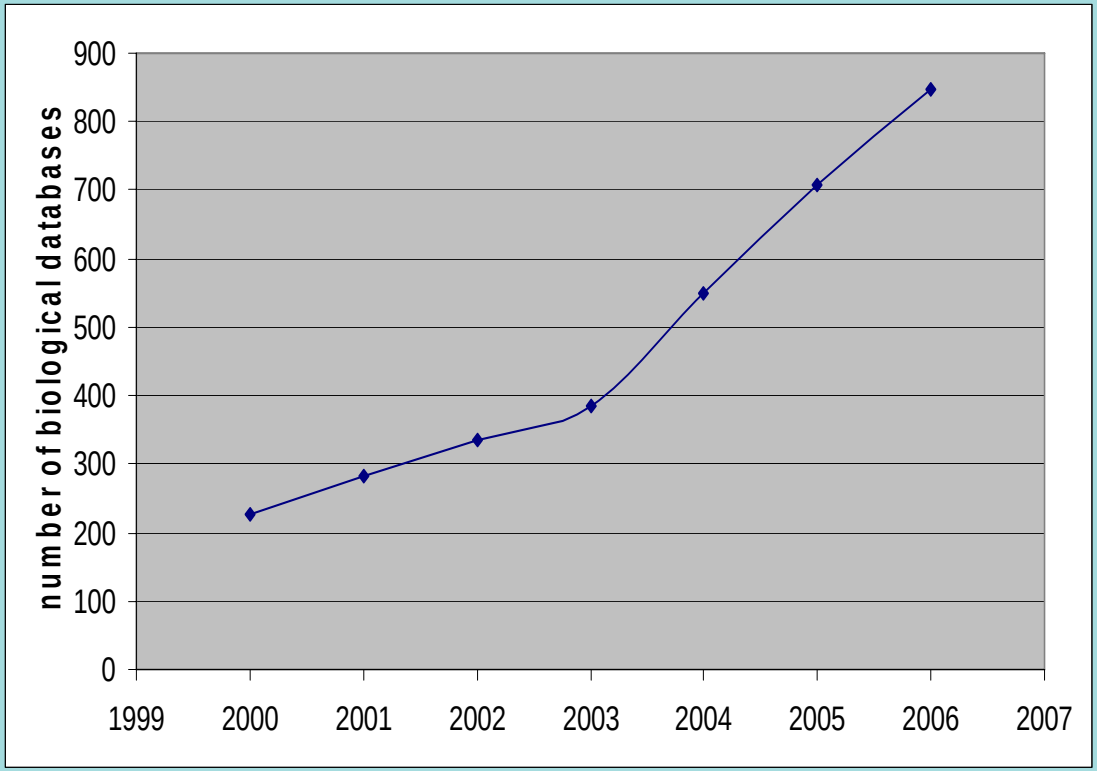
Malika Mahoui, Zina Ben Miled, Amey Godse, Harshad Kulkarni, Nianhua Li

Presented by: Malika Mahoui

Indiana University School of informatics @ IUPUI

Biological Domain

- Data intensive domain
- Gene
  - GenBank
  - EMBL
- Protein
  - SwissProt
  - PIR
  - PDB



Biological Research

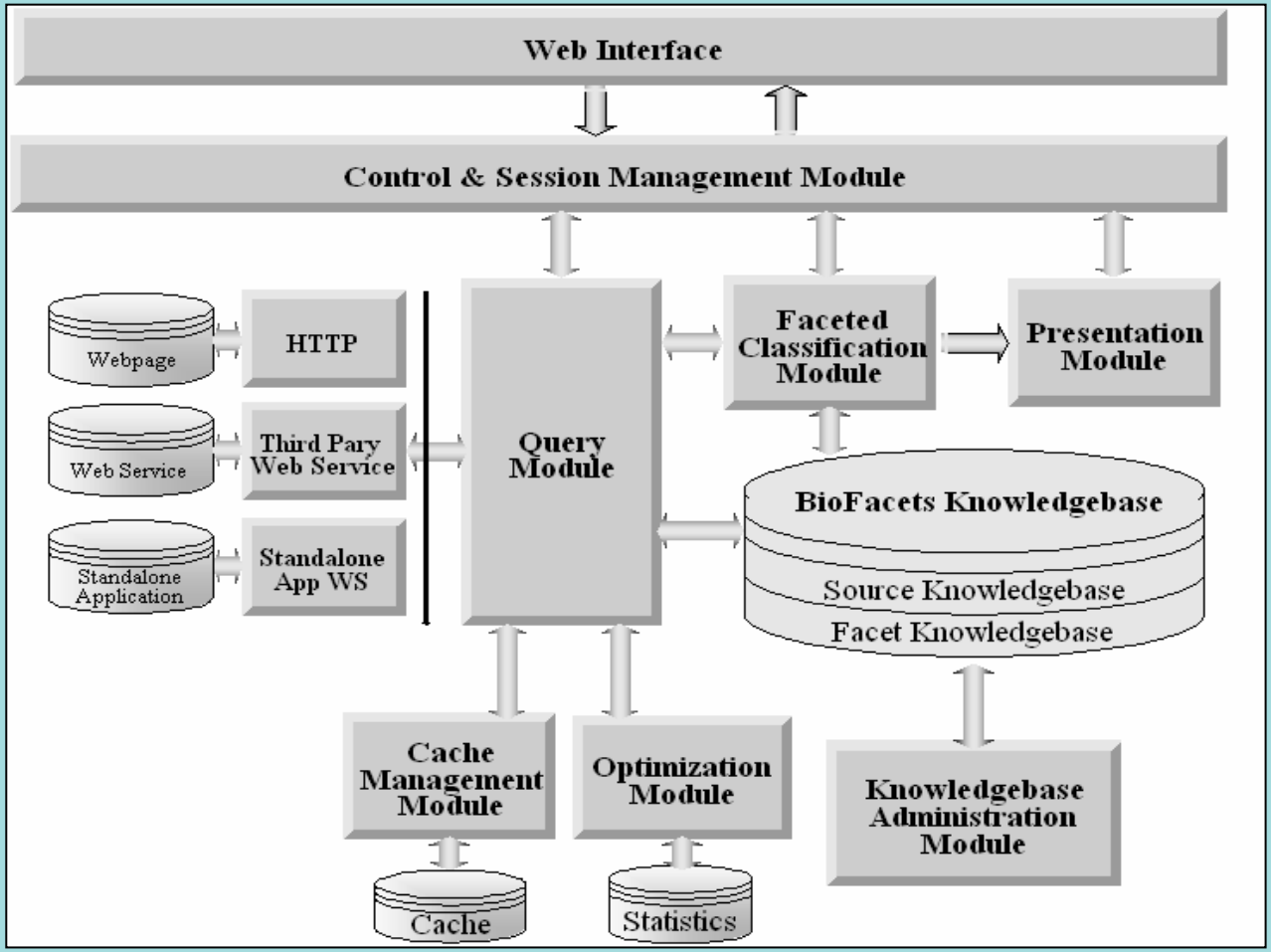
- **Characteristics: Biological Databases**
  - **Representational heterogeneity**
  - **Diversity of biological data**
  - **Large result sets**

1. **Querying remote databases**
2. **Integrating multiple databases**
3. **Representing result sets**

Biofacets Solution

- **features**
  - **Meta-search engine for biological databases**
  - **Wrapper-mediator approach for data integration**
  - **Dynamic Facetted approach for results classification**
  - **Results presentation and query refinement based on faceted classification**
  - **Cache management and query optimization to support system performance**

BioFacets Architecture



Faceted Classification

- Concept largely understood in digital libraries
- Assign multiple classifications to a result record
- Examples include Flamenco framework for image search

- Limitation: assume existence of data/metadata a priori

Facet & Facet Specification

- **A method of classification**

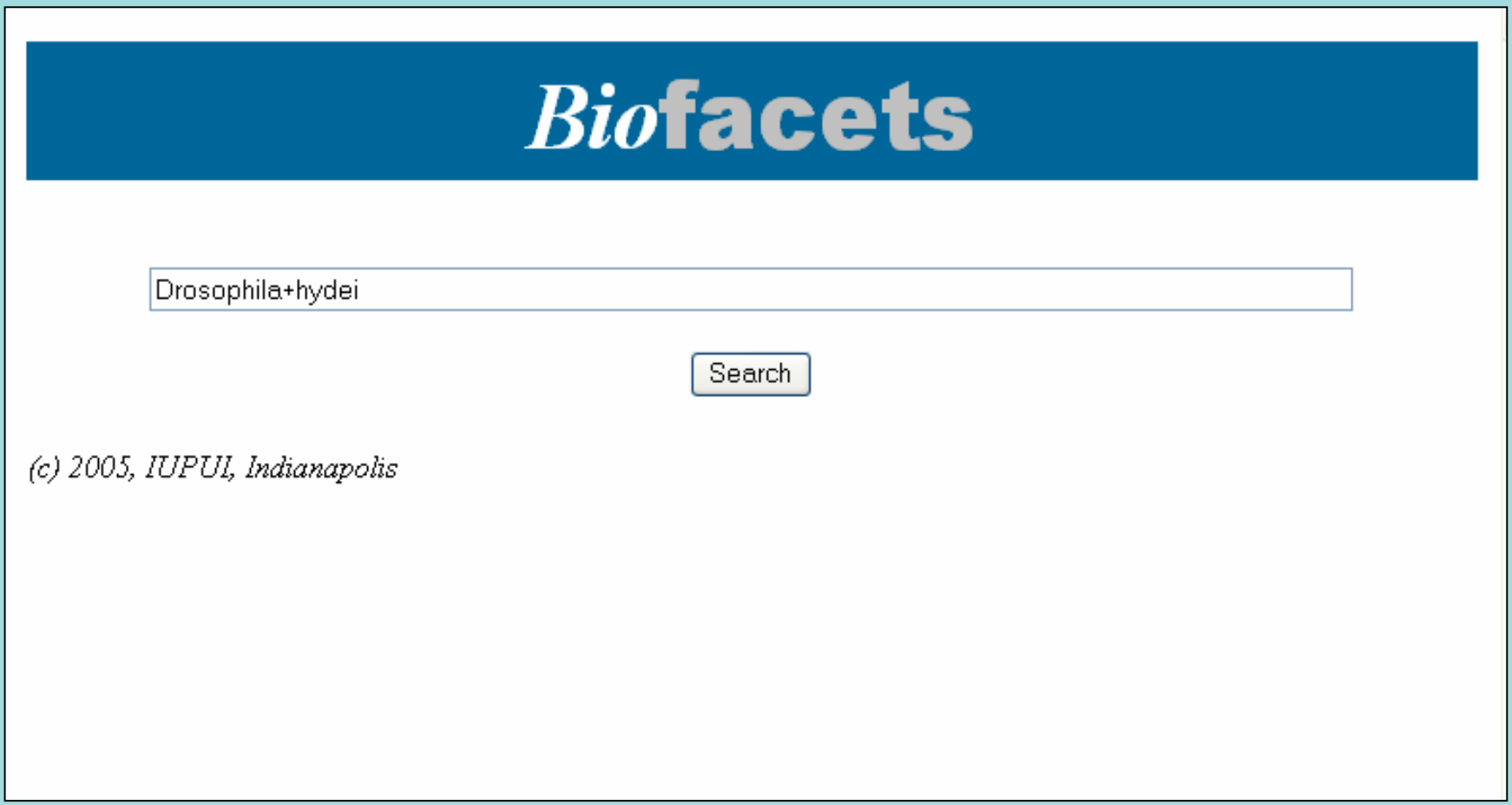
- **Facet name**
- **Assignment of value:**
  - **Static**
    - **Data Type**
  - **Dynamic**
    - **Protein Length**
- **Level:**
  - **Non-hierarchical**
    - **Gene function**
  - **Hierarchical**
    - **Organism Lineage**

```
<Facet
   fName=”data_type”
   type="static”
   isHierarchical="false”>
</Facet>
```

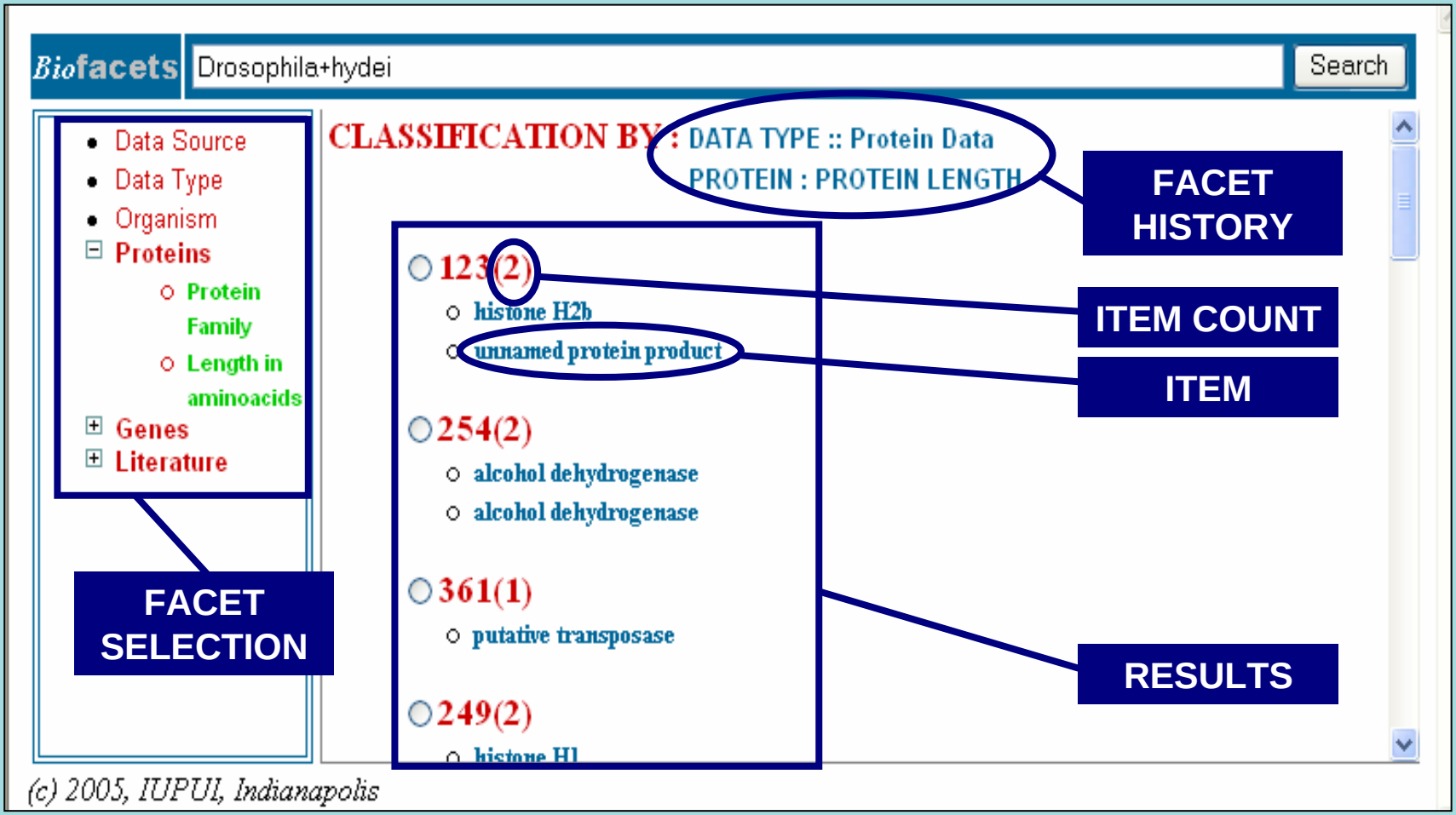
Classification Rules

Facet Type	Rule Type	Specification
Static	fixed value rule	<Rule> <ruleFacetName>data_type</ruleFacetName> <ruleMethod>fixed</ruleMethod> <Value>protein_data</Value> </Rule>
Dynamic	field value rule	<Rule> <ruleFacetName>organism</ruleFacetName> <ruleMethod>fieldvalue</ruleMethod> <Field>scientific_name</Field> </Rule>
Dynamic	lookup value rule	<Rule> <ruleFacetName>organism</ruleFacetName> <ruleMethod>lookup</ruleMethod> <DataSource>newt</DataSource> <LookupBaseURL> http://www.ebi.ac.uk/newt/display? from=au&match=taxonomy+identifier&search= </LookupBaseURL> <LookupField>tax_id</LookupField> <ValueField>scientific_name</ValueField> </Rule>

Start Page


Biofacets
Drosophila+hydei
Search
(c) 2005, IUPUI, Indianapolis

Faceted Browsing



Demonstration

- Demo

Conclusions

- **BioFacets has the potential to become the “Google” for biologists enhanced with a dynamic faceted classification approach for results presentation**

Acknowledgments

- NSF CAREER DBI-DBI-0133946
- NSF DBI-0110854