The Semantic Web …

Joel Sachs

jsachs@cs.umbc.edu

Joint work with: Cyndy Parr, Andriy Parafiynyk, Tim Finin and the whole SPIRE gang

The Semantic Web … Is it Good for Anything?

Joel Sachs

jsachs@cs.umbc.edu

Joint work with: Cyndy Parr, Andriy Parafiynyk, Tim Finin and the whole SPIRE gang

UMBC
an Honors University in Maryland

The Semantic Web … What is it good for?

Joel Sachs

jsachs@cs.umbc.edu

Joint work with: Cyndy Parr, Andriy Parafiynyk, Tim Finin and the whole SPIRE gang

The Semantic Web … It Just Might Work.

Joel Sachs

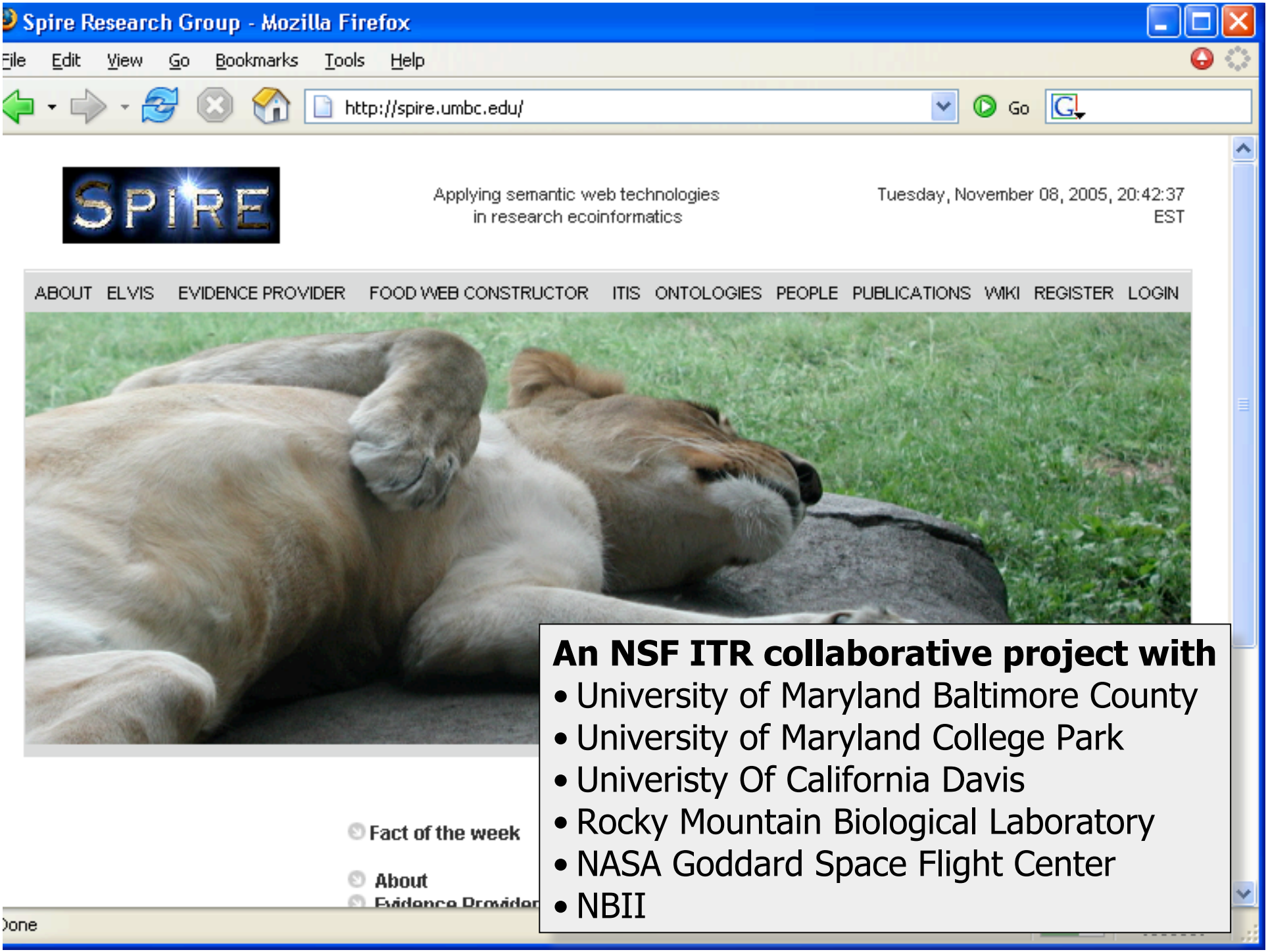
jsachs@cs.umbc.edu

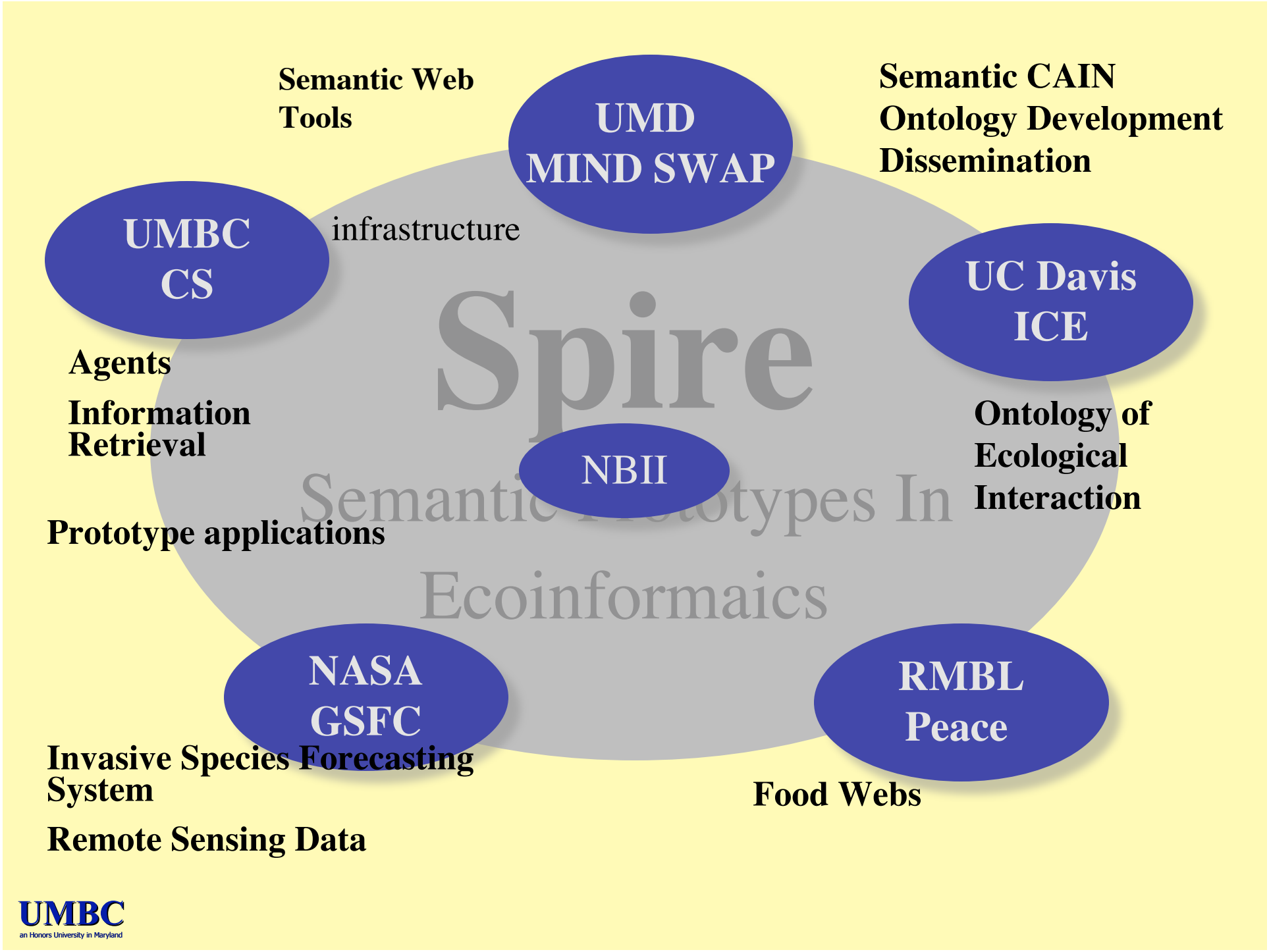
Joint work with: Cyndy Parr, Andriy Parafiynyk, Tim Finin and the whole SPIRE gang

UMBC
an Honors University in Maryland

Overview of Talk

- Introduction and background (very brief)
- ELVIS
  - The Food Web Constructor and the Evidence Provider
- ETHAN
  - An ontology for evolutionary trees and natural history
- Swoogle and Tripleshop
- Back to the drawing board?
- Deep Thoughts
  - Microformats, ontology engineering, etc.
- Questions/Objections/Better ideas


Spire Research Group - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://spire.umbc.edu/
Go
SPIRE
Applying semantic web technologies
in research ecoinformatics
Tuesday, November 08, 2005, 20:42:37
EST
ABOUT ELVIS EVIDENCE PROVIDER FOOD WEB CONSTRUCTOR ITIS ONTOLOGIES PEOPLE PUBLICATIONS WIKI REGISTER LOGIN
Fact of the week
About
An NSF ITR collaborative project with
• University of Maryland Baltimore County
• University of Maryland College Park
• Univeristy Of California Davis
• Rocky Mountain Biological Laboratory
• NASA Goddard Space Flight Center
• NBII



UMBC
an Honors University in Maryland

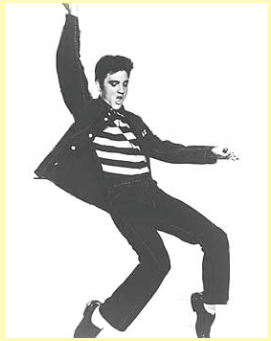
Invasive Species

- Invasive species cost the U.S. economy over $138 billion per year [1].
- By various estimates, these species contribute to the decline of 35 to 46 percent of U.S. endangered and threatened species
- The invasive species problem is growing, as the number of pathways of invasion increases.

[1] Pimental et al. 2000 Environmental and economic costs associated with non-indigenous species in the United States. Bioscience 50:53-65.
[2] Charles Groat, Director U.S. Geological Survey, http://www.usgs.gov/invasive_species/plw/usgsdirector01.html

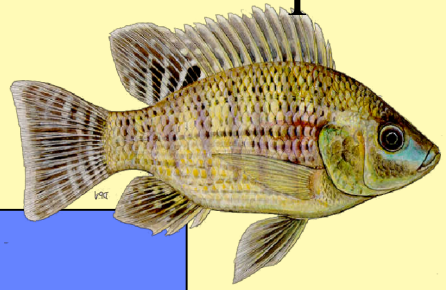
An invasive species scenario

- Nile Tilapia fish have been found in a California lake.
- Can this invasive species thrive in this environment?
- If so, what will be the likely consequences for the ecology?
- So…we need to understand the effects of introducing this fish into the food web of a typical California lake



ELVIS: Ecosystem Localization, Visualization, and Information System

Oreochromis niloticus
Nile tilapia



Species list constructor

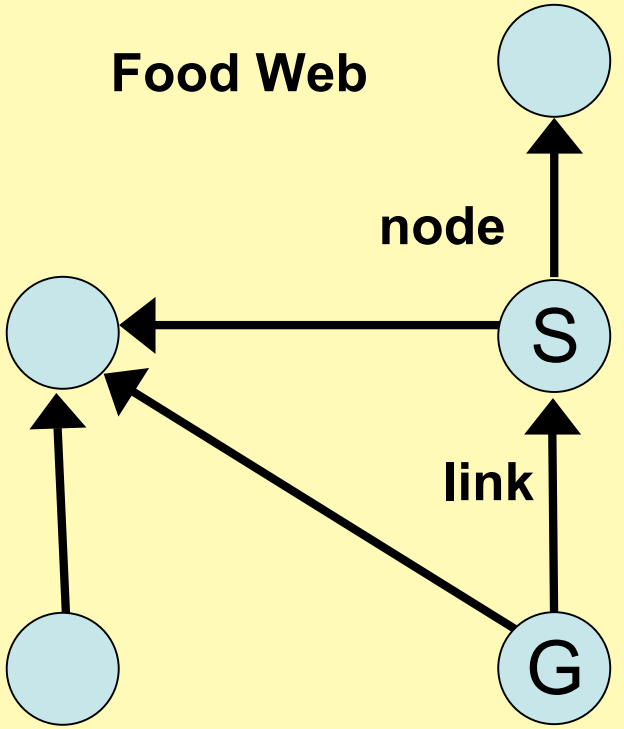
Bacteria
Microprotozoa
Amphithoe longimana
Caprella penantis
Cymadusa compta
Lembos rectangularis
Batea catharinensis
Ostracoda
Melanitta
Tadorna tadorna . . .

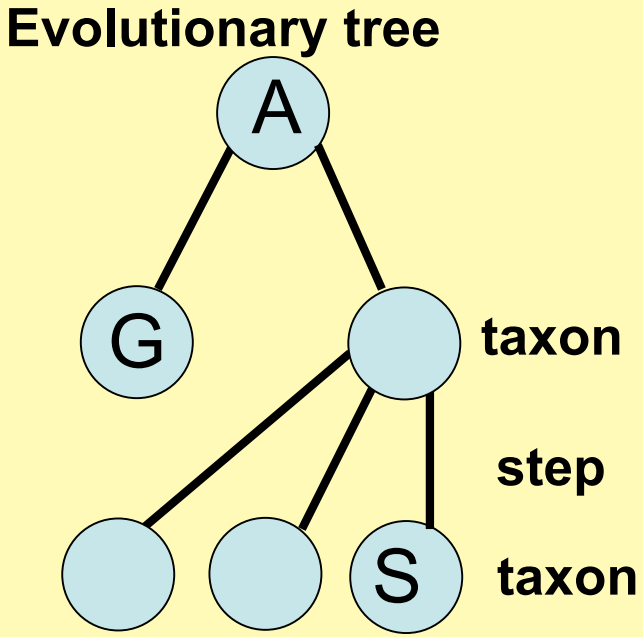
Food web constructor

?

The problem

- We have data on what species are known to be in the location and can further restrict and fill in with other ecological models
- But we don't know which of these the Nile Tilapia eats of who might eat it.
- We can reason from taxonomic data (similar species) and known natural history data (size, mass, habitat, etc.) to fill in the gaps.


Food Web
node
S
link
G


Evolutionary tree
A
G
taxon
step
S
taxon

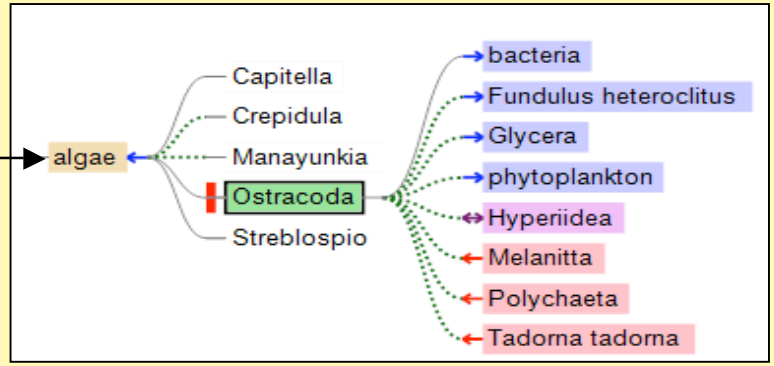
Show the ELVIS Demo, Joel.

Food Web Constructor

Predict food web links using database and taxonomic reasoning.

Habitat	Study ID	Study Details	Publication Year
Agricultural	45	L. J. Tilly, The structure and dynamics of Cone Spring. Ecol. Monogr. 38(2):169-197, from p. 183 (1968).	1968
	58	N. N. Smirnov, Food cycles in sphagnous bogs, Hydrobiologia 17:175-182, from p. 179 (1961).	1961
	62	V. S. Summerhayes and C. S. Elton, Further contributions to the ecology of Spitzbergen, J. Ecol. 16:193-268, from p. 217 (1928).	1928
	90	D. J. Shure, Radionuclide tracer analysis of trophic relationships in an old-field ecosystem, Ecol. Monogr. 43(1):1-19, from p. 15 (1973).	1973
	153	M. A. Mayse and P. W. Price, 1978. Seasonal development of soybean arthropod communities in east central Illinois. Agro-Ecosys. 4:387-405, from p. 401.	1978
	154	M. A. Mayse and P. W. Price, 1978. Seasonal development of soybean arthropod communities in east central Illinois. Agro-Ecosys. 4:387-405, from p. 402.	1978
Brackish water	1	S. Z. Qazim, Some problems related to the food chain in a tropical estuary. In: Marine Food Chains, J. H. Steele, Ed. (Oliver and Boyd, Edinburgh, 1970), pp. 45-51, from p. 50.	1970
	57	A. Yanez-Arancibia, Taxonomia, ecologia y estructura de las comunidades de peces en lagunas costeras con bocas efimeras del Pacifico de Mexico. Cent. Cienc. del Mar y Limnol. Univ. Nal. Auton. Mex. Publ. Espec. 2:1-306 (1978).	1978
		K. Hogetsu, Biological productivity of some coastal regions of Japan. In: Marine Production Mechanisms	

Link#	Certainty idx	Predator	Prey
1	0.9825	Gammarus pulex	Chironomidae
2	0.2933	Gammarus pulex	algae
3	5.4717	Gammarus pulex	detritus
4	2.5	Perla carlukiana	Baetis
5	0.4933	Perla carlukiana	Hydropsyche
6	0.9817	Perla carlukiana	Chironomidae
7	0.995	Perla carlukiana	Simulium
8	2.65	Baetis	encrusting algae
9	7.9625	Baetis	detritus
10	1.5	Ecdyonurus venosus	detritus



In a new estuary, Nile Tilapia could compete with ostracods (green) to eat algae. Predators (red) and prey (blue) of ostracods may be affected

UMBC
an Honors University in Maryland

Food Web Constructor generates possible links

POSITIVE EVIDENCE

Link#	Certainty idx	Predator	Prey
1	15.87	bacteria	detritus
2	14.83	Amphipoda	detritus
3	9.9417	Ostracoda	detritus
2380	0.72	detritus	bacteria
2381	0.7	Pagurus	Maldanidae
2382	0.6667	Oreochromis niloticus	detritus
2383	0.6667	Oreochromis niloticus	algae
2384	0.6667	Sterna forsteri	Fundulus similis
2385	0.6667	Sterna forsteri	Fundulus confluentus
2386	0.6667	Larus atricilla	Bivalvia

UMBC
an Honors University in Maryland

SPIRE

Evidence provider gives details

Oreochromis niloticus (rank Species) is a likely predator for algae (rank Unknown)

Similar links observed between:

1

Predator: juvenile Cichlidae (taxon Cichlidae, rank Family)

Prey: bottom algae and detritus (taxon algae, rank Unknown)

In habitat:
River

The link is discovered in the study: G. Fryer, The trophic interrelationships and ecology of some littoral communities of Lake Nyasa, Proc. London Zool. Soc. 132:153-229, from p. 219 (1959).

Published in: 1959 year

The study was conducted in: Country: Malawi

Link Proximity: **33.33%**

UMBC
an Honors University in Maryland

SPIRE

Testing the algorithm

- Take each web out of the database
- Attempt to predict its links
- Compare prediction with actual data

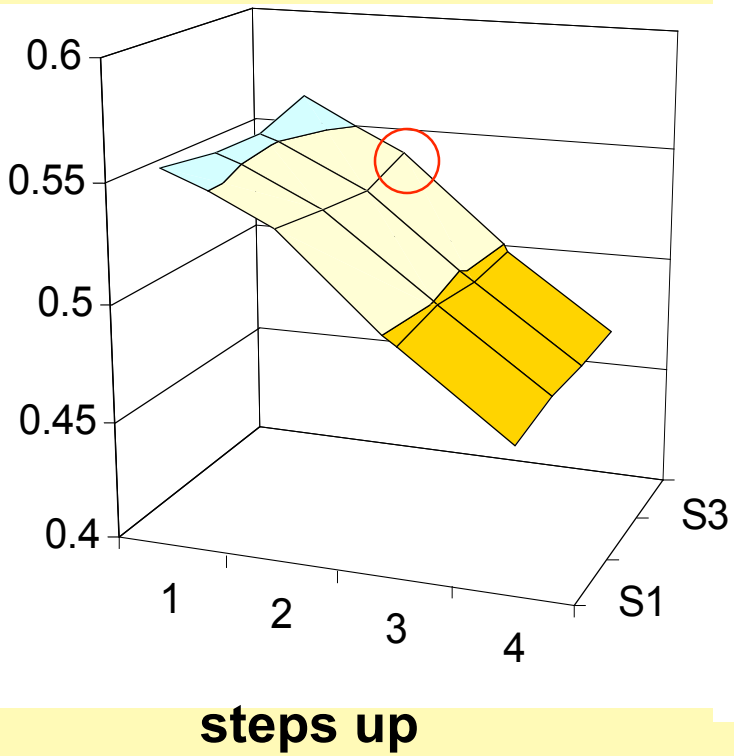
Accuracy *percentage of all predictions that are correct* **89%**

Precision *percentage of predicted links that are correct* **55%**

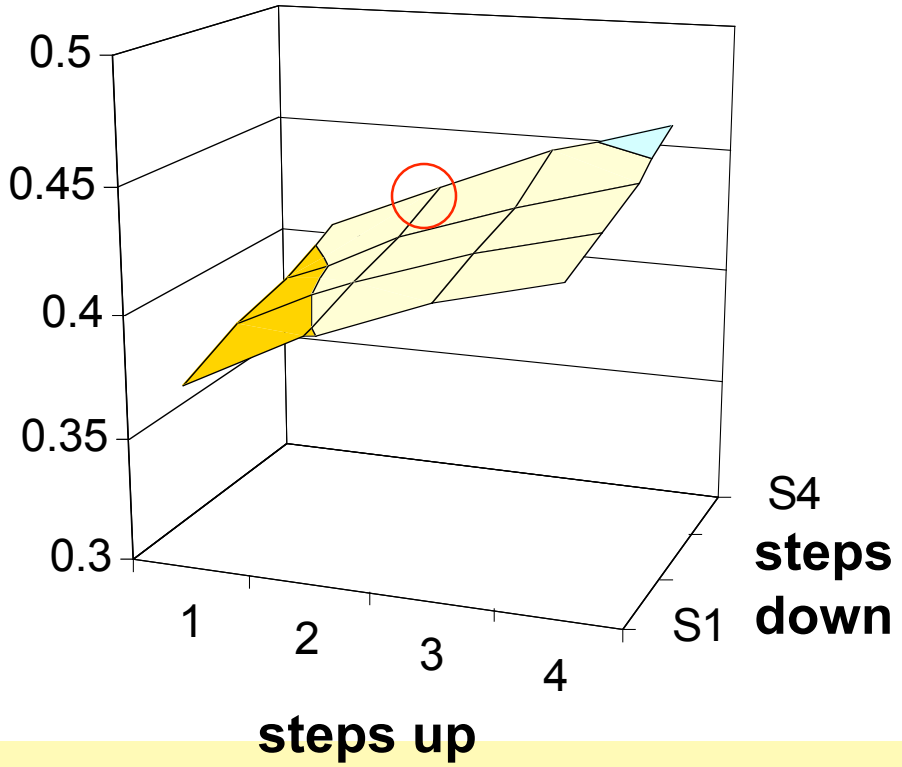
Recall *percentage of actual links that are predicted* **47%**

Evolutionary distance threshold
2 steps up and 4 steps down

precision



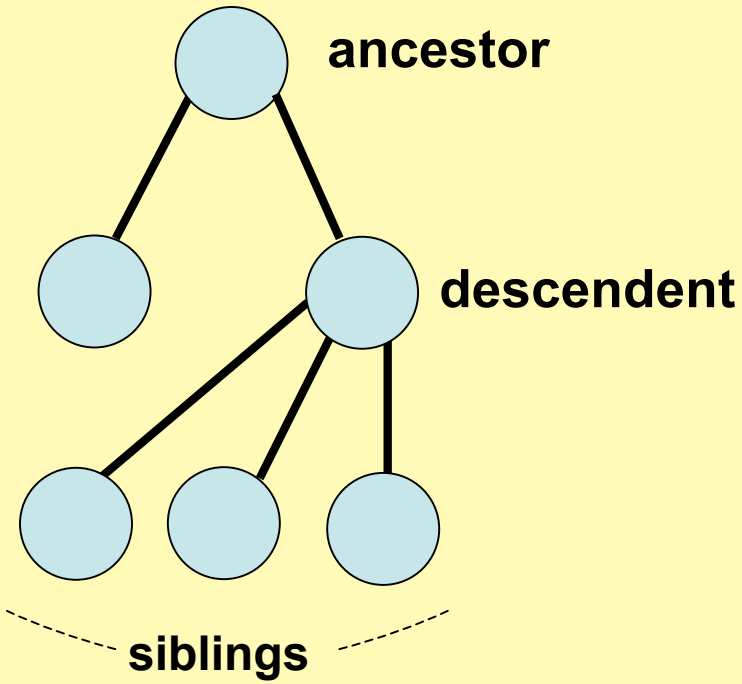
recall



UMBC
an Honors University in Maryland

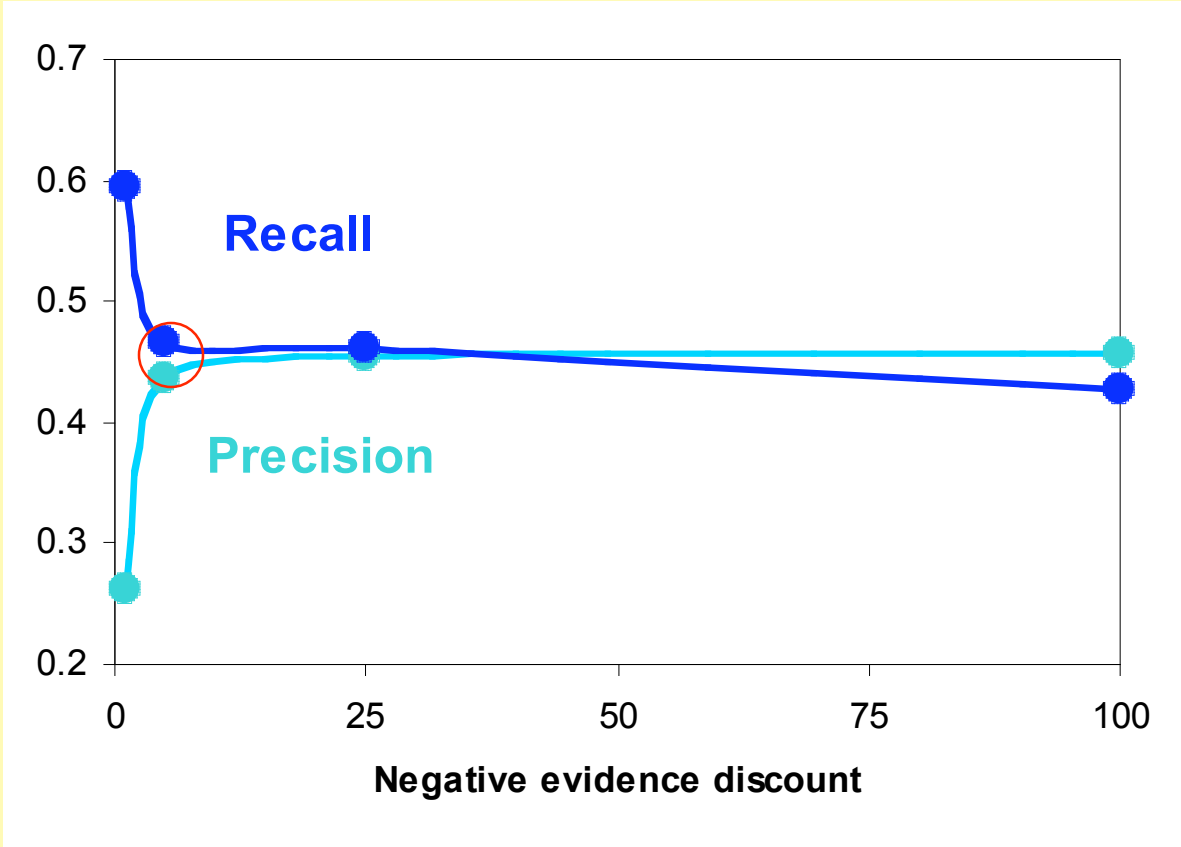
Evolutionary direction penalty not very sensitive

$$Weight_{AB} = \frac{1}{1 + (Distance_{XA} \times Penalty_{XA}) + (Distance_{YB} \times Penalty_{YB})}$$

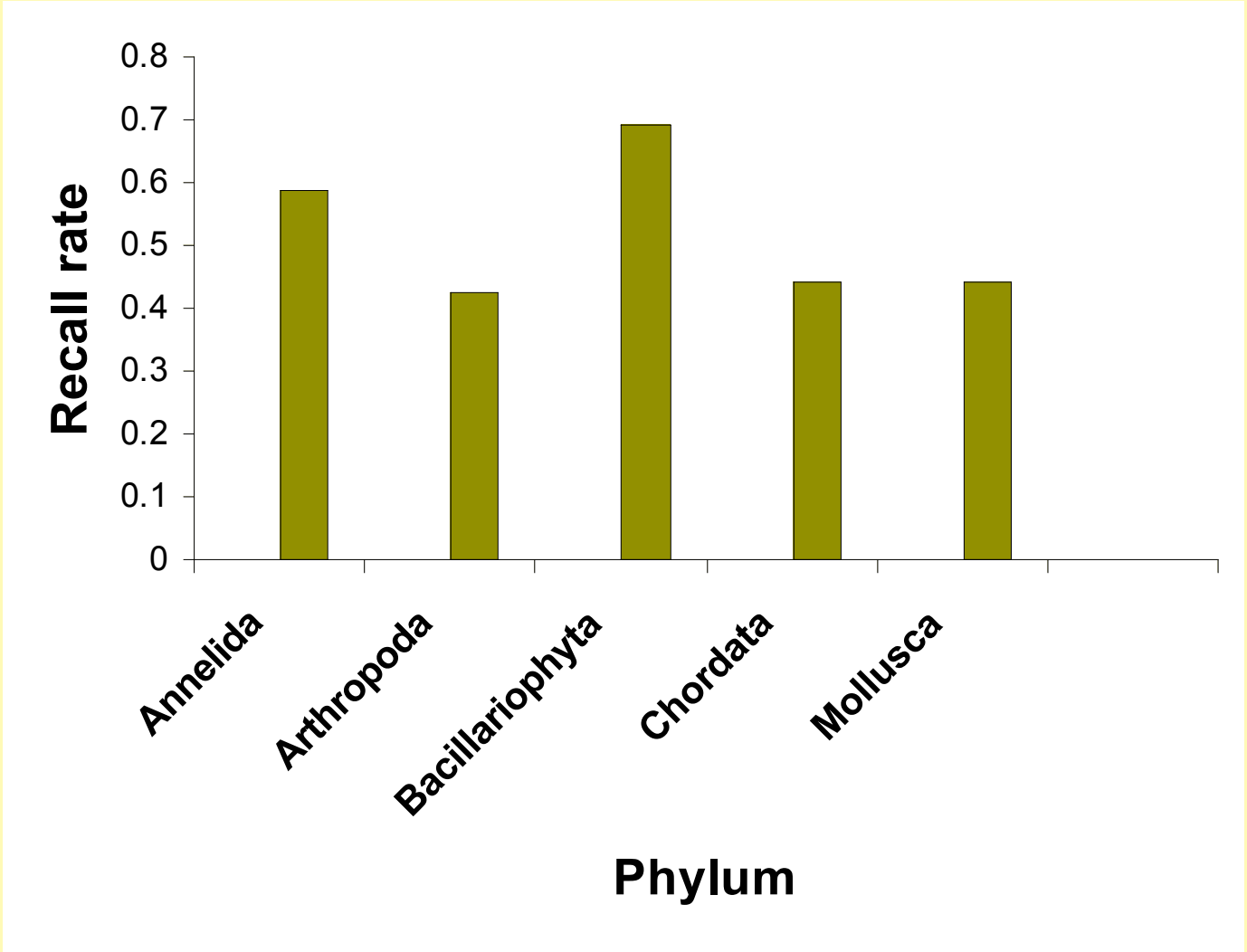


Negative evidence discount is sensitive

$$CertaintyIdx_{XY} = \sum_{i=1}^{N} \frac{weight_i}{discount}(LinkValue_i)$$



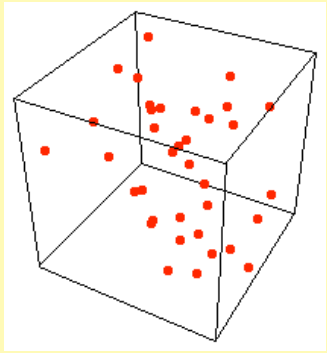
Some phyla are easier to predict than others



How can we do better predicting links?

Trait space distance weighting

Euclidean distance in natural history N-space



Parameterize functions from the literature that might predict links using characteristics of taxa. For example, size or stoichiometry.

$$\text{LinkStatus}_{AB} = f(\alpha, \text{size}_A, \text{size}_B), f(\beta, \text{stoich}_A, \text{stoich}_B) \ldots$$

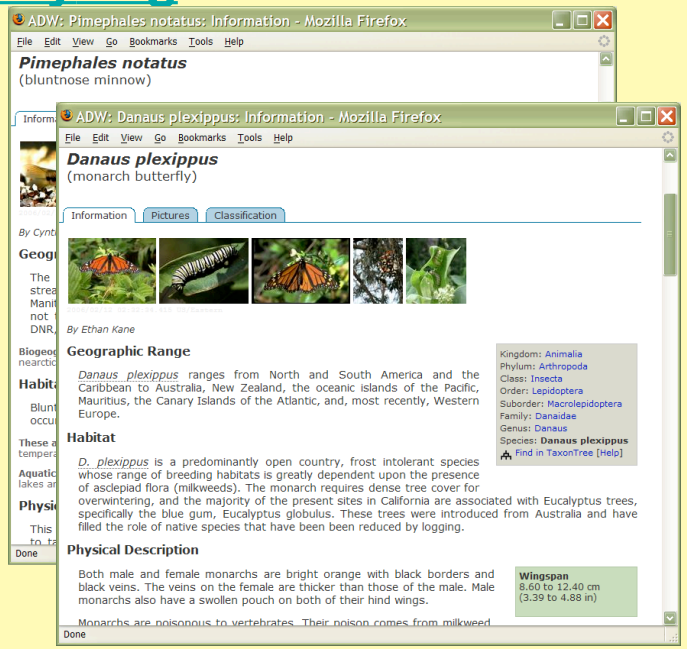
…need more data

ETHAN
Evolutionary Trees and Natural History ontology

Animal Diversity Web

http://www.animaldiversity.org

- geographic range
- habitats
- physical description
- reproduction
- lifespan
- behavior and trophic info
- conservation status



Triples

“Esox lucius” hasMaxMass “1.4 kg”

“Esox lucius” isSubclassOf “Esox”

“Esox” eats “Actinopterygii”

UMBC
an Honors University in Maryland

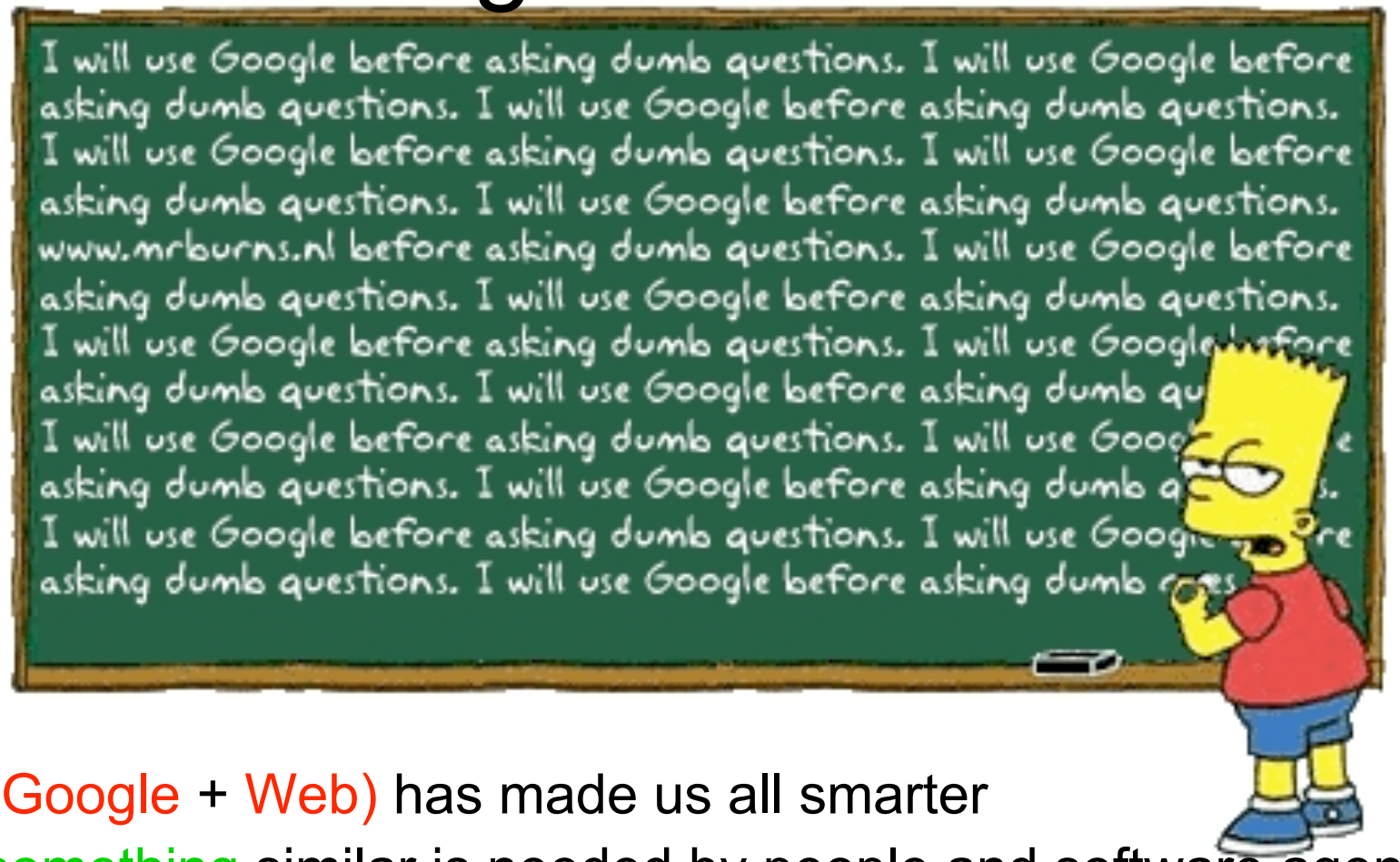
Ethan Requirements

- Be able to clean data, even if it comes from multiple sources.
  - Are there any organisms described both as a flier and glider?
  - Are there cases where the reported values for average mass of a species do not fall within the max and the min of its family.
- Be able to query to find (distributed) data.
  - Find the geographic ranges of birds (Class Aves) that chorus or duet.
- Be able to take advantage of inheritance and aggregation.
  - If I add a triple that family Corvidae is precocial (a reproductive keyword) can all birds in this family inherit that characteristic?
  - If I know that five species of butterflies have various maximum lifespans, I want to be able to get an answer to the question, "What isthe maximum lifespan in the family that contains those butterflies."
- Be able to take advantage of property hierarchies.
- Be able to extend easily.

Joel, show the people the following

- http://spire.umbc.edu/ontologies/ethan_keywords.owl
- http://spire.umbc.edu/ontologies/taxa/Gorilla_gorilla_gorilla.owl

Swoogle: Motivation



- (Google + Web) has made us all smarter
- something similar is needed by people and software agents for finding information on the semantic web

UMBC
an Honors University in Maryland

Okay, now show Swoogle and Tripleshop

Swoogle Triple Shop

Account Query Dataset Run

User: default

Query Name: new query

What are body masses of fishes that eat fishes?

Creator:	default	**Created Time:**	
Visible to Others:	true	**Updated Time:**	
Description:	click here	**Tags:**	

SPARQL Query:

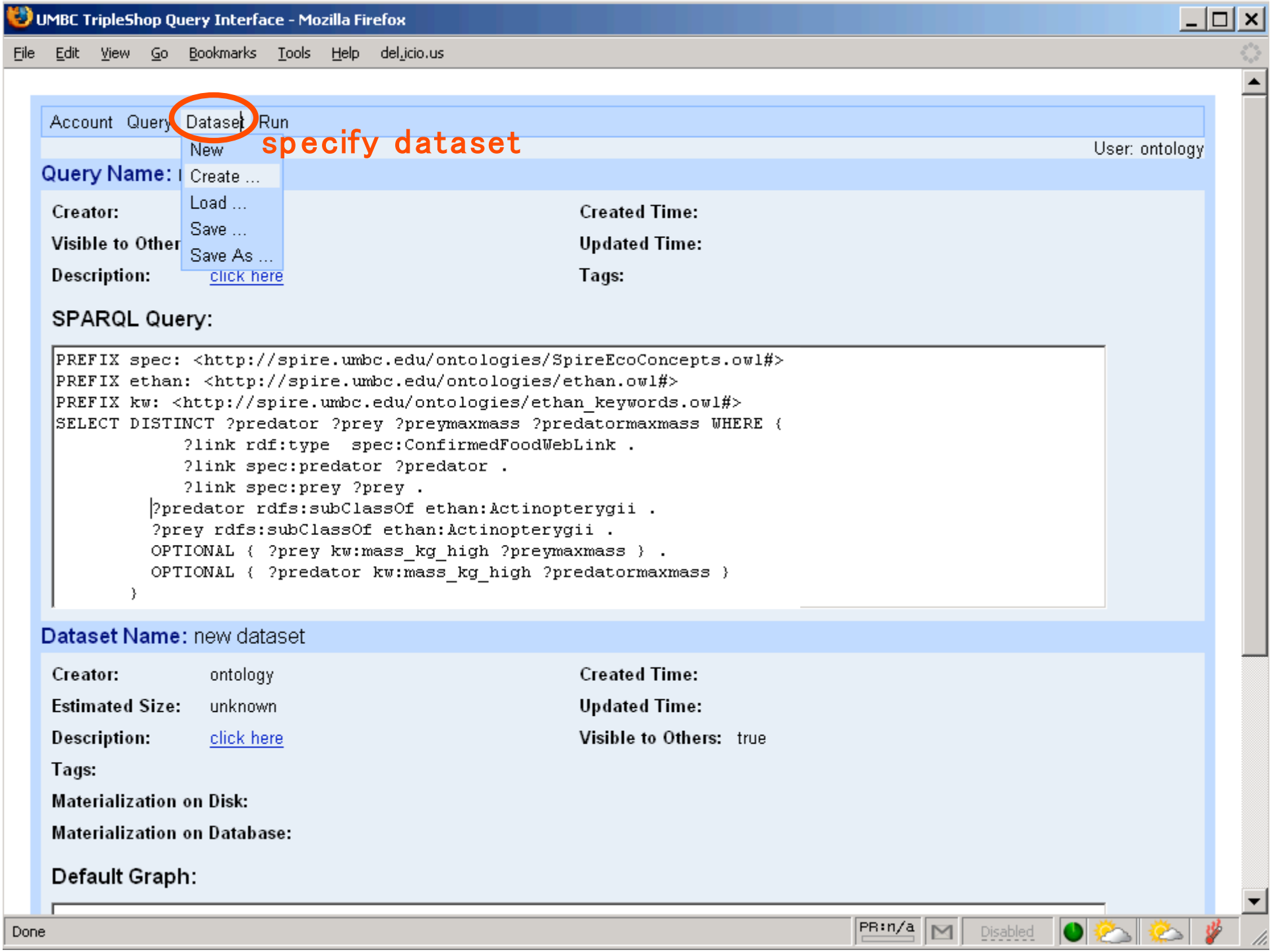
```
SELECT DISTINCT ?predator ?prey ?preymaxmass ?predatormaxmass
WHERE {
            ?link rdf:type  spec:ConfirmedFoodWebLink .
            ?link spec:predator ?predator .
            ?link spec:prey ?prey .
            ?predator rdfs:subClassOf ethan:Actinopterygii .
            ?prey rdfs:subClassOf ethan:Actinopterygii .
            OPTIONAL { ?predator kw:mass_kg_high ?predatormaxmass }  .
            OPTIONAL { ?prey kw:mass_kg_high ?preymaxmass }
      }
```

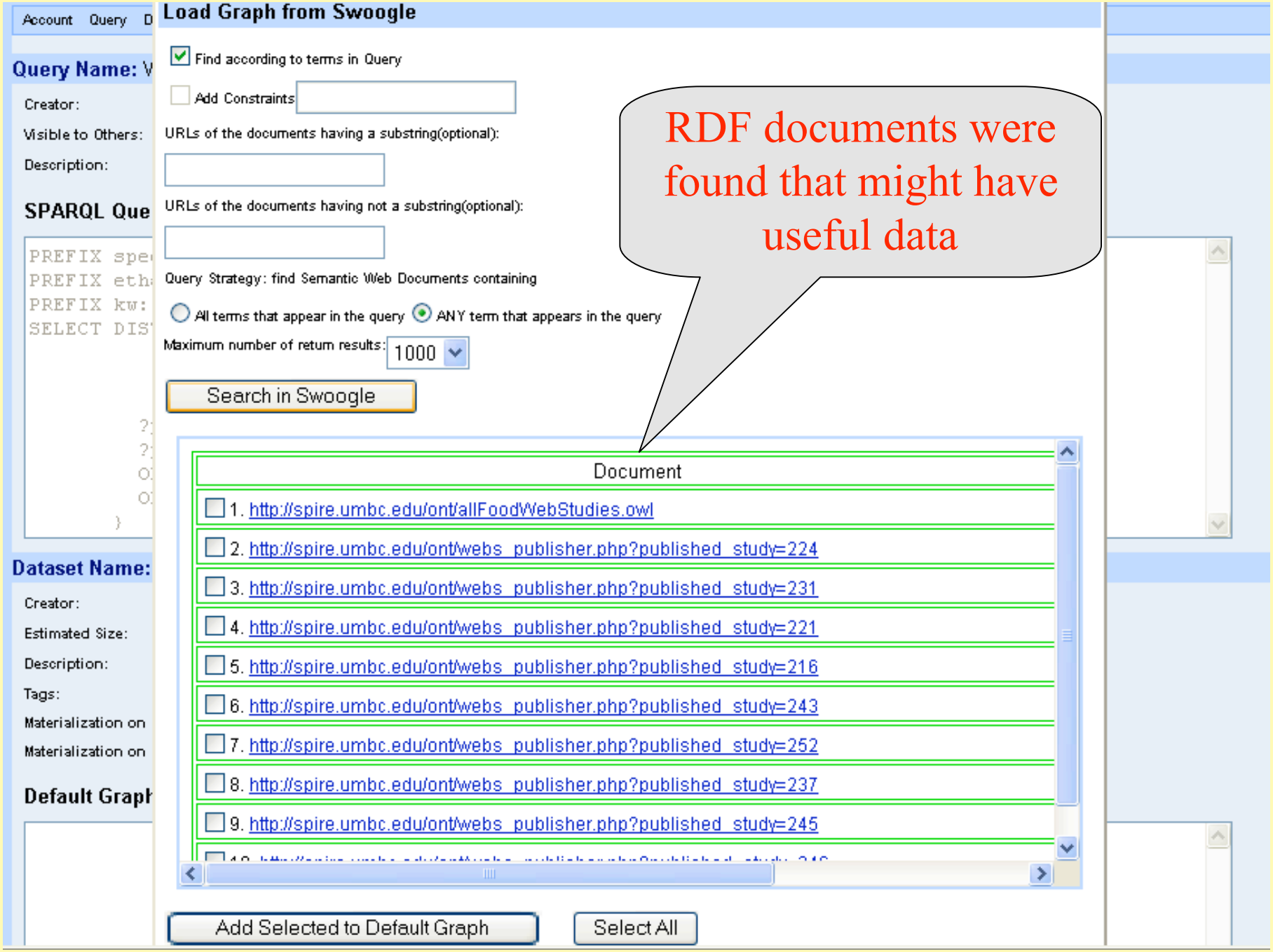
. . . leaving out the FROM clause

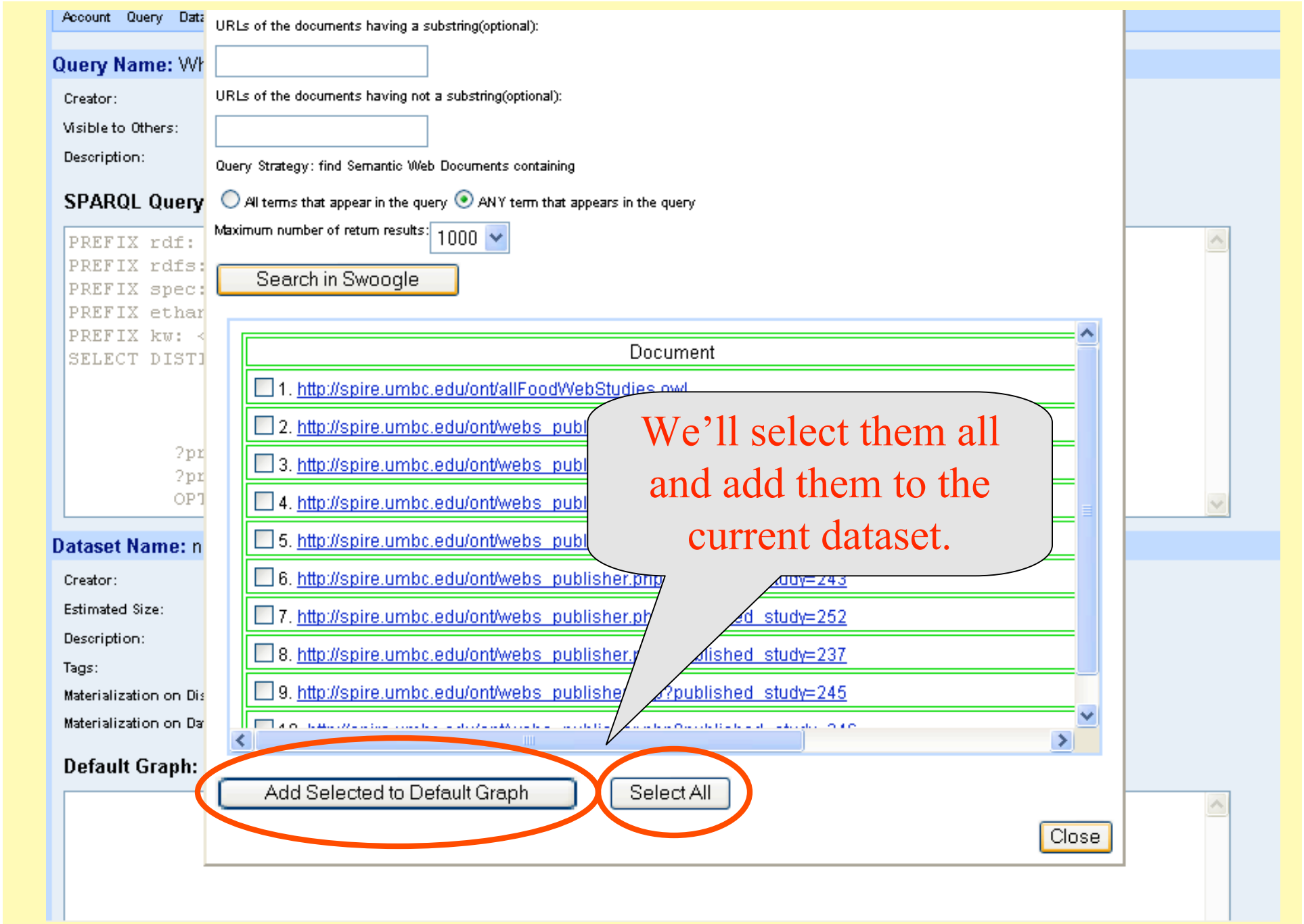
Dataset Name: new dataset

Creator:	default	**Created Time:**	
Estimated Size:	unknown	**Updated Time:**	
Description:	click here	**Visible to Others:**	true
Tags:			

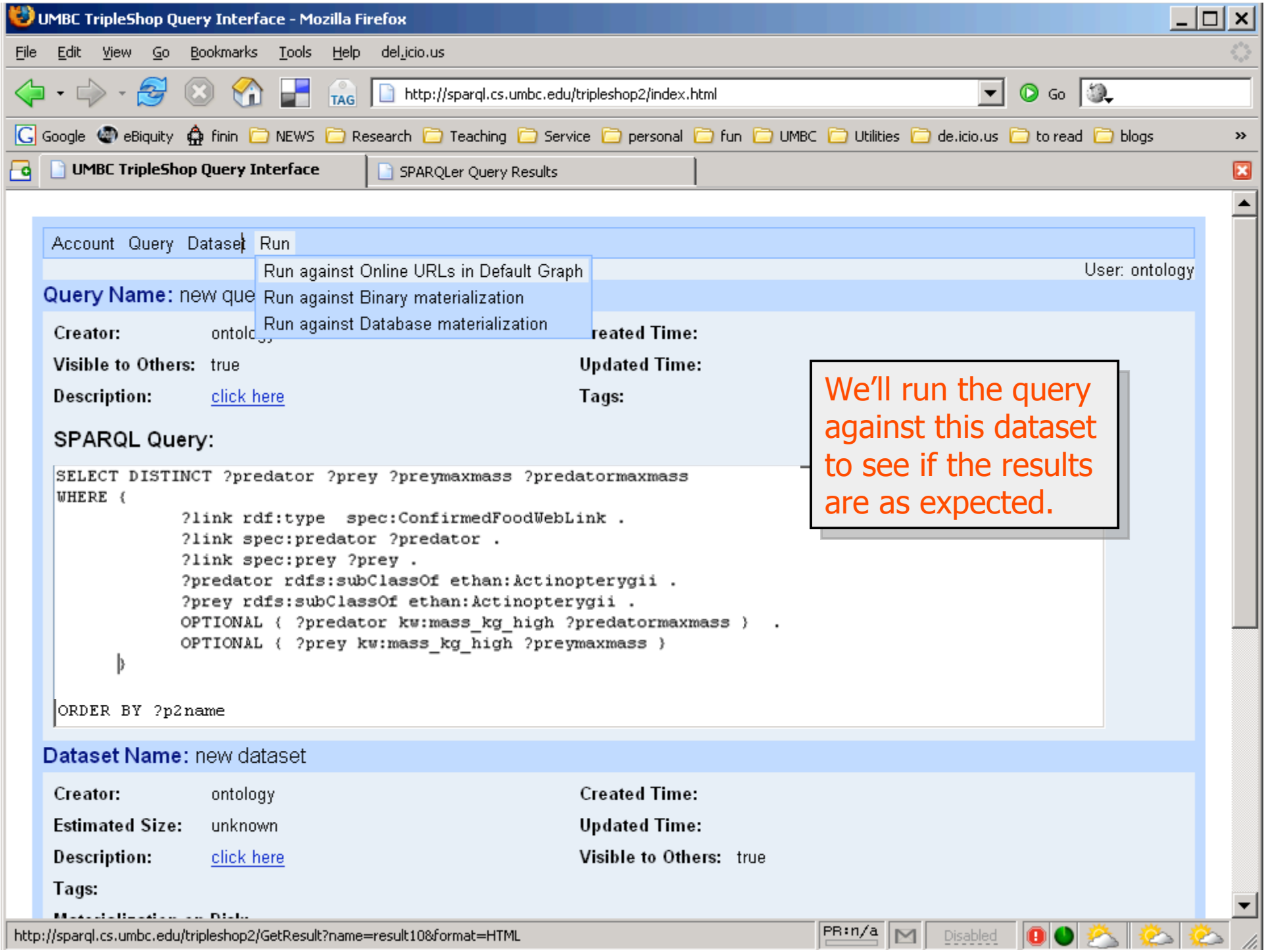
UMBC
an Honors University in Maryland

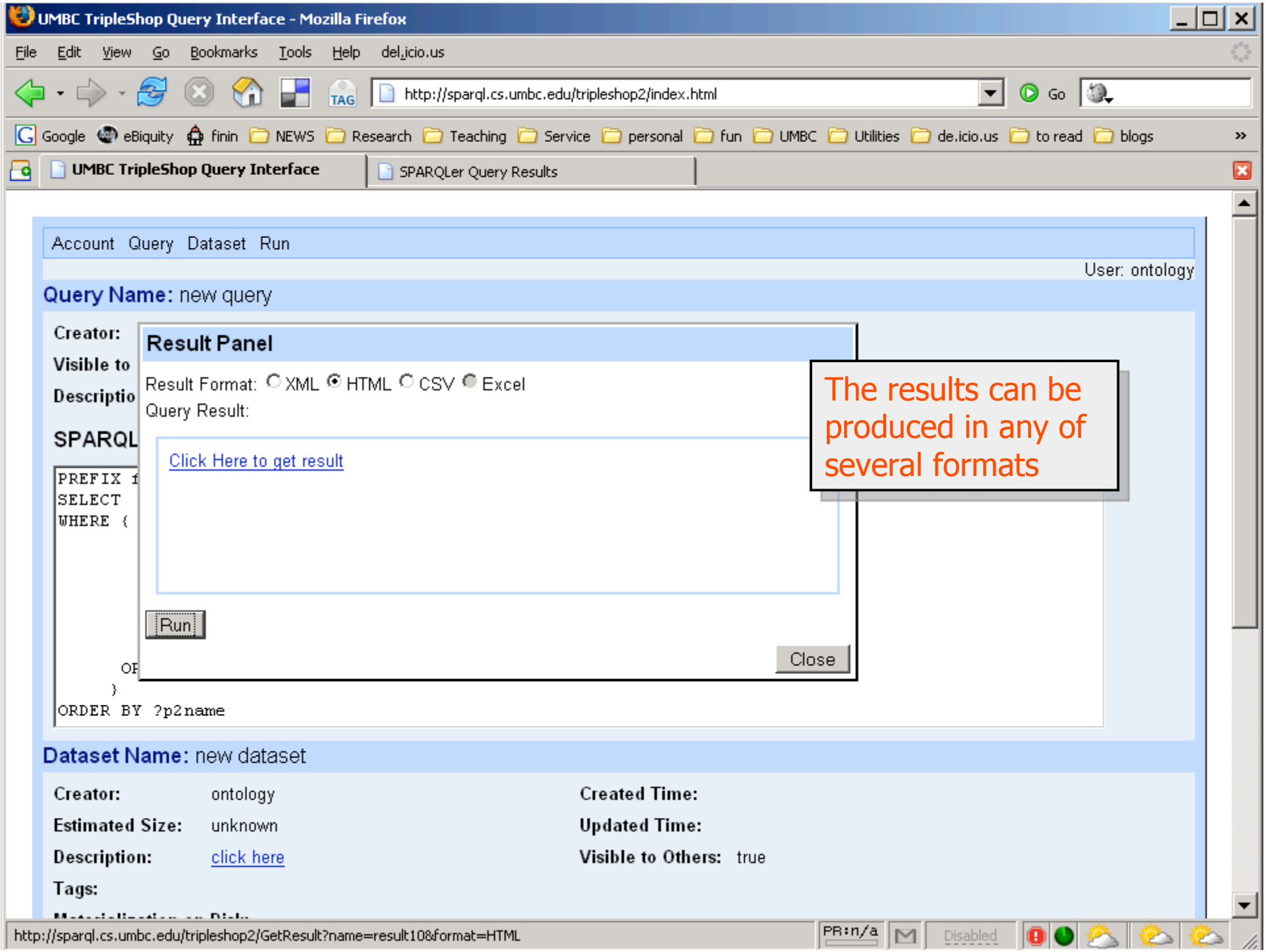

UMBC TripleShop Query Interface - Mozilla Firefox
File Edit View Go Bookmarks Tools Help del.icio.us
Account Query Dataset Run
specify dataset
New
Create ...
Load ...
Save ...
Save As ...
User: ontology
Query Name:
Creator:
Visible to Other
Description: click here
Created Time:
Updated Time:
Tags:
SPARQL Query:
PREFIX spec: <http://spire.umbc.edu/ontologies/SpireEcoConcepts.owl#>
PREFIX ethan: <http://spire.umbc.edu/ontologies/ethan.owl#>
PREFIX kw: <http://spire.umbc.edu/ontologies/ethan_keywords.owl#>
SELECT DISTINCT ?predator ?prey ?preymaxmass ?predatormaxmass WHERE {
?link rdf:type spec:ConfirmedFoodWebLink .
?link spec:predator ?predator .
?link spec:prey ?prey .
?predator rdfs:subClassOf ethan:Actinopterygii .
?prey rdfs:subClassOf ethan:Actinopterygii .
OPTIONAL { ?prey kw:mass_kg_high ?preymaxmass } .
OPTIONAL { ?predator kw:mass_kg_high ?predatormaxmass }
}
Dataset Name: new dataset
Creator: ontology
Estimated Size: unknown
Description: click here
Tags:
Materialization on Disk:
Materialization on Database:
Created Time:
Updated Time:
Visible to Others: true
Default Graph:
Done
PR:n/a
Disabled


Account Query D
Load Graph from Swoogle
Query Name: V
Creator:
Visible to Others:
Description:
SPARQL Que
PREFIX spe
PREFIX eth
PREFIX kw:
SELECT DIS
Dataset Name:
Creator:
Estimated Size:
Description:
Tags:
Materialization on
Materialization on
Default Graph
Find according to terms in Query
Add Constraints
URLs of the documents having a substring(optional):
URLs of the documents having not a substring(optional):
Query Strategy: find Semantic Web Documents containing
All terms that appear in the query
ANY term that appears in the query
Maximum number of return results:
1000
Search in Swoogle
RDF documents were found that might have useful data
Document
1. http://spire.umbc.edu/ont/allFoodWebStudies.owl
2. http://spire.umbc.edu/ont/webs_publisher.php?published_study=224
3. http://spire.umbc.edu/ont/webs_publisher.php?published_study=231
4. http://spire.umbc.edu/ont/webs_publisher.php?published_study=221
5. http://spire.umbc.edu/ont/webs_publisher.php?published_study=216
6. http://spire.umbc.edu/ont/webs_publisher.php?published_study=243
7. http://spire.umbc.edu/ont/webs_publisher.php?published_study=252
8. http://spire.umbc.edu/ont/webs_publisher.php?published_study=237
9. http://spire.umbc.edu/ont/webs_publisher.php?published_study=245
Add Selected to Default Graph
Select All



UMBC
an Honors University in Maryland


UMBC TripleShop Query Interface - Mozilla Firefox
File Edit View Go Bookmarks Tools Help del.icio.us
http://sparql.cs.umbc.edu/tripleshop2/index.html
Go
Google eBiquity finin NEWS Research Teaching Service personal fun UMBC Utilities de.icio.us to read blogs
UMBC TripleShop Query Interface
SPARQLer Query Results
Account Query Dataset Run
Run against Online URLs in Default Graph
Run against Binary materialization
Run against Database materialization
User: ontology
Query Name: new que
Creator:
reated Time:
Visible to Others: true
Updated Time:
Description: click here
Tags:
SPARQL Query:
SELECT DISTINCT ?predator ?prey ?preymaxmass ?predatormaxmass
WHERE {
?link rdf:type spec:ConfirmedFoodWebLink .
?link spec:predator ?predator .
?link spec:prey ?prey .
?predator rdfs:subClassOf ethan:Actinopterygii .
?prey rdfs:subClassOf ethan:Actinopterygii .
OPTIONAL { ?predator kw:mass_kg_high ?predatormaxmass } .
OPTIONAL { ?prey kw:mass_kg_high ?preymaxmass }
}
ORDER BY ?p2name
We'll run the query against this dataset to see if the results are as expected.
Dataset Name: new dataset
Creator: ontology
Created Time:
Estimated Size: unknown
Updated Time:
Description: click here
Visible to Others: true
Tags:
http://sparql.cs.umbc.edu/tripleshop2/GetResult?name=result10&format=HTML
PR:n/a
Disabled


UMBC TripleShop Query Interface - Mozilla Firefox
File Edit View Go Bookmarks Tools Help del.icio.us
http://sparql.cs.umbc.edu/tripleshop2/index.html
Go
Google eBiquity finin NEWS Research Teaching Service personal fun UMBC Utilities de.icio.us to read blogs
UMBC TripleShop Query Interface
SPARQLer Query Results
Account Query Dataset Run
User: ontology
Query Name: new query
Creator:
Visible to
Descriptio
SPARQL
PREFIX f
SELECT
WHERE {
ORDER BY ?p2name
Result Panel
Result Format: XML HTML CSV Excel
Query Result:
Click Here to get result
Run
Close
The results can be produced in any of several formats
Dataset Name: new dataset
Creator: ontology
Estimated Size: unknown
Description: click here
Tags:
Created Time:
Updated Time:
Visible to Others: true
http://sparql.cs.umbc.edu/tripleshop2/GetResult?name=result10&format=HTML
PR:n/a
Disabled

Results

predator	prey	predatormass	preymass
Salmo_trutta	Lepomis_macrochirus		2.2
Micropterus_salmoides	Lepomis_macrochirus	10	2.2
Stizostedion_vitreum	Carassius_auratus	10	3
Esox_masquinongy	Stizostedion_vitreum	29.48	10
Esox_lucius	Stizostedion_vitreum	1.4	10
Ictalurus_punctatus	Stizostedion_vitreum		10
Micropterus_salmoides	Stizostedion_vitreum	10	10
Micropterus_salmoides	Micropterus_salmoides	10	10

http://sparql.cs.umbc.edu/tripleshop2/

And now, the punchline.

- http://spire.umbc.edu/ontologies/InvasivesOntology.owl
- http://spire.umbc.edu/ontologies/descurainia_pinnata.owl
- http://spire.umbc.edu/ontologies/CaliforniaWeeds.owl

Queries

- List unique predators and prey and their body masses
  - Combine SpireEcoConcepts with the invasives lists and Ethan (for the sizes)
- Q: What taxa on a survey list are outside their known invasive range.
  - Combine a survey list with invasive lists and a geographic
- Q: Who are predators of the birds (class aves) on the lists of concern?
  - List unique predators and prey.
  - Will combine SpireEcoConcepts with EthanAnimals and the invasives lists
- Q: What are reproductive characteristics (e.g. number of offspring) of animals onthe invasives lists?
  - List taxon, and characteristics
  - Will combine Ethan taxon documents and invasive lists
- Q: What taxa are on List X but not on List Y?
  - List taxaCombines different invasives lists.

Deep Thoughts

- Some cool things require less formal semantics than we thought.
  - Mashups, etc.
- Other things require more shared, formal semantics than we seem to be capable of.
  - Is the ozone whole getting bigger or smaller?

- A little tour of flickr, fieldmarking, species microformats, etc.

Tripleshop is_a Work in Progress

- There are a host of performance issues
- We plan on supporting some special datasets, e.g.,
  - FOAF and SPIRE data collected from Swoogle
  - Definitions of RDF and OWL classes and properties from all ontologies that Swoogle has discovered
- Expanding constraints to select candidate SWDs to include arbitrary metadata and embedded queries
  - FROM “documents trusted by a member of the SPIRE project”
- “Qurantine” needed to handle conflicts.

Review

- All Elvis functionality is encapsulated as web services, and all input and output is OWL based.
  - So Elvis integrates easily with other semantic web applications, like the TripleShop.
- ELVIS as a platform for experimenting with different approached to food web prediction.
- TripleShop as an integrating platform
- TripleShop allows researchers to semi-automatically construct datasets in response to ad-hoc queries.
- Contact jsachs@cs.umbc.edu to participate.