

The Road to AmiGO 2

Seth Carbon

Berkeley BOP (<http://berkeleybop.org>),
Lawrence Berkeley National Lab

04 October 2012

Away from MySQL

We have been using an increasing number of tricks, extensions, and caches to keep the performance at an acceptable level. Things have to change. . .

- Complicated queries: enrichment, subsets, search, reports, etc.
- Data: ~1,500,000 -> ~13,000,000 -> ~80,000,000 -> ???
- Providing services

Into Solr

Introducing Solr

- A specialized HTTP server over the Lucene document store
- A common document-oriented approach (Lucene; similar to QuickGO)
- All of your data with just a URL

Why You Might Care

AmiGO 2 has greatly increased in flexibility, speed, and development turnaround time over SQL.

For example: a deep text search from ~30s. down to ~0.3s.

It has also made things that were previously not feasible easy to do.
For example:

- Very fast searches, even for previously impossible topics
- Widgets for search, graphs, etc.
- Easier direct calculations (counts, information content)
- Easy third-party use (data and clients)

AmiGO 2

Looking good!

Associations

Search: [reset]

Filter

- annotation_extension_class_closure_label cell [X]
- in vitro
- evidence_type: BMP [X]
- Source
- Evidence type

IDA	(12)	[+]	[-]
ISO	(3)	[+]	[-]
IGI	(2)	[+]	[-]
- taxon
- is-aPart-of closure (labels)
- Annotation extension class closure (labels)

Found entities

Total: 17 First: 11 Last: 17

Score	Annotation class	Evidence type	Gene Product	Source	Taxon	With	Annotation extension class
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	IDA	Zfp351	MGI	Mus musculus		permanent cell line cell
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	IDA	Zfp353	MGI	Mus musculus		permanent cell line cell
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	IDA	Zfp351	MGI	Mus musculus		permanent cell line cell
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	IDA	Zfp352	MGI	Mus musculus		permanent cell line cell
40.9%	protein ubiquitination involved in catabolic process	IDA	huh	MGI	Mus musculus		permanent cell line cell
32.7%	ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway	IDA	hcd56	MGI	Mus musculus		permanent cell line cell
32.7%	ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway	IGI	Hsfy1	MGI	Mus musculus	MGI:MGJ:97297	macrophage.primary cell culture cell macrophage.primary cell culture cell

AmiGO 2

Looking good!

The image shows two overlapping screenshots of the AmiGO 2 web interface. The background screenshot displays search results for the term 'catabolic'. The foreground screenshot shows a detailed view of the 'apoptosis' ontology term, including a list of related terms and their relationships.

AmiGO 2 Search Results (Background Screenshot):

Search: [reset]

Found entities: Total: 17 First: 11 Last: 17

Score	Annotation class	Evidence type	Gene Product	Source	Taxon	With	Annotation extension class
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	IDA	Zfp351	MGJ	Mus musculus		permanent cell line cell
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	IDA	Zfp353	MGJ	Mus musculus		permanent cell line cell
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	IDA	Zfp381	MGJ	Mus musculus		permanent cell line cell
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	IDA	Zfp352	MGJ	Mus musculus		permanent cell line cell
40.9%	protein ubiquitination involved in ubiquitin-dependent protein catabolic process	IDA	hsp	MGJ	Mus musculus		permanent cell line cell
32.7%	ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway	IDA	hsp58	MGJ	Mus musculus		
32.7%	ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway	IGI	hsp1	MGJ	Mus musculus		

AmiGO 2 Detailed View (Foreground Screenshot):

AmiGO - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://localhost:8080/ami-go/ontology/term.do?term=apoptosis

The Gene Ontology AmiGO Labs

You are using AmiGO Labs

Browse

Browser info

TOGO: simple explanation for why this looks different.

More information here.

Jump to term:

Transitive

- anti-apoptosis
- apoptosis fated cell
- apoptosis activator activity
- apoptosis regulator activity
- apoptosis inhibitor activity
- regulation of anti-apoptosis
- commitment to apoptosis
- induction of apoptosis
- execution phase of apoptosis
- negative regulation of anti-apoptosis

AmiGO 2

Looking good!

Associations

Search: [reset]

Filters

- annotation_extension_class_closure_label_cell [X]
- in vitro
- evidence_type: BIP [X]
- Source
- Evidence type

IDA	(12)	[+]	[-]
ISO	(3)	[+]	[-]
IGI	(2)	[+]	[-]

Found entities

Total: 17 First: 11 Last: 17

Score	Association class	Evidence type	Gene Product	Source	Taxon	With	Annotation extension class
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	EA	Zfp351	MG2	Mus musculus		permanent cell line cell
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	EA	Zfp353	MG2	Mus musculus		permanent cell line cell
40.9%	nucleus-transcribed mRNA catabolic process, deadenylation-dependent decay	EA	Zfp381	MG2	Mus musculus		permanent cell line cell

Transitive Term Neighborhood

- [is a] [\[GO:0008150\]](#) biological_process [i]
- [is a] [\[GO:0071840\]](#) cellular component organization or biogenesis [i]
- [is a] [\[GO:0009987\]](#) cellular process [i]
 - [is a] [\[GO:0016043\]](#) cellular component organization [i]
 - [is a] [\[GO:0071841\]](#) cellular component organization or biogenesis at cellular level [i]
 - [is a] [\[GO:0022411\]](#) cellular component disassembly [i]
 - [is a] [\[GO:0071842\]](#) cellular component organization at cellular level [i]
 - [is a] [\[GO:0043933\]](#) macromolecular complex subunit organization [i]
 - [is a] [\[GO:0071845\]](#) cellular component disassembly at cellular level [i]
 - [is a] [\[GO:0034621\]](#) cellular macromolecular complex subunit organization [i]
 - [is a] [\[GO:0032984\]](#) macromolecular complex disassembly [i]
 - [is a] [\[GO:0034623\]](#) cellular macromolecular complex disassembly [i]
 - [is a] [\[GO:0071824\]](#) protein-DNA complex subunit organization [i]
 - [[[]]] [\[GO:0032986\]](#) protein-DNA complex disassembly [i]
 - [is a] [\[GO:0006337\]](#) nucleosome disassembly [i]

Navigation icons: back, forward, search, etc.

AmiGO 2

Looking good!

Associations

Search: [reset]

Filters

- annotation_extension_class_obsolete_label_cell [X]
- in vitro
- evidence_type: IMP [X]
- Source
- Evidence type

Evidence type	IDA	ISO	IGI
	(12) [+]	(3) [-]	(2) [-]

Found entities

Total: 17 First: 11 Last: 17

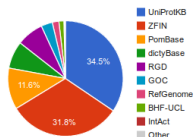
Score	Association class	Evidence type	Gen Prod
40.0%	nucleus-transcribed mRNA catalytic process, deadenylation-dependent decay	IDA	Zy3H
40.0%	nucleus-transcribed mRNA catalytic process, deadenylation-dependent decay	IDA	Zy3H
40.0%	nucleus-transcribed mRNA catalytic process, deadenylation-dependent	IDA	Zy3H

Transitive Term Neighborhood

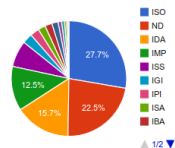
- [is a] [\[GO:0008150\]](#) biological_process [I]
- [is a] [\[GO:0071840\]](#) cellular component o
- [is a] [\[GO:0009987\]](#) cellular process [I]
- [is a] [\[GO:0016043\]](#) cellular componen
- [is a] [\[GO:0071841\]](#) cellular componen
- [is a] [\[GO:0022411\]](#) cellular compor
- [is a] [\[GO:0071842\]](#) cellular component organization at cellular level [I]
- [is a] [\[GO:0043933\]](#) macromolecular complex subunit organization [I]
- [is a] [\[GO:0071845\]](#) cellular component disassembly at cellular level [I]
- [is a] [\[GO:0034621\]](#) cellular macromolecular complex subunit organization [I]
- [is a] [\[GO:0032984\]](#) macromolecular complex disassembly [I]
- [is a] [\[GO:0034623\]](#) cellular macromolecular complex disassembly [I]
- [is a] [\[GO:0071824\]](#) protein-DNA complex subunit organization [I]
- [[[]]] [\[GO:0032986\]](#) protein-DNA complex disassembly [I]
- [is a] [\[GO:0006337\]](#) nucleosome disassembly [I]

Current Gene Ontology annotation information

Sources (381931)



Evidence (381931)



Secondary Tools

- Term Enrichment and Mapper/Slimmer
 - Being moved to Galaxy
- The Birds
 - GOOSE (SQL environment) not going anywhere for a while
 - Gannet (Solr environment) now included in AmiGO 2.
- New reporting tools
 - Scripting and remote use now possible

What We Have to Do

- Finishing up...
- Figuring out deployment
- Education and beta

Information and Demo

AmiGO 2 wiki and demo site

