KnowOS

Web Based
Listener
Object Persistence
Security v. Collaboration
Efficient Scripting
Process Management
Integrated Knowledge
Toolset Integration
Ability to read Upside Down

Goals of a Knowledge Operating System

Provide persistent object store (interconnected frames)

Provide storage for data as well as knowledge.

Provide integration with programs built by others.

Provide persistence of user environment across sessions.

Provide rich, efficient, extensible scripting.

Provide "the right amount" of user integration.

Provide universal access for both users and client code.

Provide access to remote services and databases.

Do all this in a convenient, integrated, user-friendly way.

KnowOS Approach

Turn Lisp into an Operating System

1. Start with ACL – fast compiler, multi-process model.

2. Run it on a server, accessed via a browser-based listener.

3. Integrate knowledge bases via a built-in frame system.

4. Run it on Linux – external tools, security model.

5. Rebuild user tools (editing, file manip., debugging, etc.)

6. Provide XML-RPC "Evalserver" for others to call in.

7. Try to avoid crashing it ("apparent persistence").

A Too-Brief History of Related Projects

Real running examples:

1970-...: APL – first PL as OS w/workspace concept

1975-?: LispMs – couldn't run external code

1979-present: Oracle/PLSQL – relational model

~1980-present: MatLab/Excel end-user PEs

~1995-?: FramerD – not really an OS (few services)

Research projects:

ErOS/CoyotOS – total persistence in a unix-style OS

TUNES (never implemented)

The infinitude of random persistent object gizmos

Plan of the Presentation

* Overview of KnowOS services

* Example 1: BioLingua biological knowledge environment

* Example 2: CACHE collaborative knowledge analysis

* Issues and approaches

* Near and long term goals

Plan of the Presentation

* Overview of KnowOS services

* Example 1: BioLingua biological knowledge environment

* Example 2: CACHE collaborative knowledge analysis

* Issues and approaches

* Near and long term goals

KnowOS Approach

Turn Lisp into an Operating System

1. Start with ACL – fast compiler, multi-process model.

2. Run it on a server, accessed via a browser-based listener.

3. Integrate knowledge bases via a built-in frame system.

4. Run it on Linux – external tools, security model.

5. Rebuild user tools (editing, file manip., debugging, etc.)

6. Provide XML-RPC "Evalserver" for others to call in.

7. Try to avoid crashing it ("apparent persistence").

KnowOS Approach

Turn Lisp into an Operating System

1. Start with ACL – fast compiler, multi-process model.

2. Run it on a server, accessed via a browser-based listener.

3. Integrate knowledge bases via a built-in frame system.

4. Run it on Linux – external tools, security model.

5. Rebuild user tools (editing, file manip., debugging, etc.)

6. Provide XML-RPC "Evalserver" for others to call in.

7. Try to avoid crashing it ("apparent persistence").

File   Edit   View   Favorites   Tools   Help

Address http://nostoc.stanford.edu:8002/redisplay.html?uid=8&pkg=JSHRAGER#TAG   Go   Links   Norton AntiVirus

Back   Forward   Stop   Refresh   Home   Search   Favorites   Media   History   Print   Edit   Discuss   Dell Home   Messenger

#$PEPCarboxykinaseC4Photosynthesis #$Hexokinase-IndependentSignaling
#$Hexokinase-DependentSignaling #$MannokinaseActivity
#$GluconokinaseActivity #$WaterPyruvateDikinaseActivity
#$ThiaminPyrophosphokinaseActi   #$Selenide,WaterDikinaseActivity
#$RibokinaseActivity #$Phosph   kinaseActivity
#$PhosphoenolpyruvateCarbox   Activity
#$PhosphoenolpyruvateCarboxyki   Diphosphate]Activity
#$PhosphoenolpyruvateCarboxyki   Atp]Activity
#$PhosphoenolpyruvateCarboxyki   ctivity
#$GTPPyrophosphokinaseActivit   XylulokinaseActivity
#$XylulokinaseActivity #$Ribos   sphatePyrophosphokinaseActivity
#$RhamnulokinaseActivity #$6-   ofructokinaseActivity
#$1-PhosphofructokinaseActivi   hosphofructokinaseActivity
#$L-RibulokinaseActivity #$L-   kinaseActivity
#$L-ArabinokinaseActivity #$Ke   okinaseActivity
#$HexokinaseActivity #$Glucok   tivity #$GalactokinaseActivity
#$FructokinaseActivity #$D-RibulokinaseActivity
#$2-Dehydro-3-DeoxygluconokinaseActivity
#$2-Dehydro-3-DeoxygalactonokinaseActivity
#$2-Amino-4-Hydroxy-6-HydroxymethyldihydropteridinePyrophosphokinaseActivity)

Results/History (green up arrow overlay)

```
*go-frames*
```

Simple Exprs (green left arrow overlay)

```
(loop for frame in *go-frames*
      when (search "kinase" (#^fname frame))
      collect frame)
```

Complex Exprs (green left arrow overlay)

Eval   Clear   Matching paren context: (loop for frame ...

BioLingua Listener v1.0

Lisp Spec   BioLingua Primitives   Upload facility   Help

KnowOS Approach

Turn Lisp into an Operating System

1. Start with ACL – fast compiler, multi-process model.

2. Run it on a server, accessed via a browser-based listener.

3. Integrate knowledge bases via a built-in frame system.

4. Run it on Linux – external tools, security model.

5. Rebuild user tools (editing, file manip., debugging, etc.)

6. Provide XML-RPC "Evalserver" for others to call in.

7. Try to avoid crashing it ("apparent persistence").

BioLingua: Frame #$GO.Acyl-CoaDehydrogenaseActivity - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://nostoc.stanford.edu:8002/frame?PKG=MT&NAME=GO.Acyl-CoaDehydrogen Go

Frame #$GO.Acyl-CoaDehydrogenaseActivity

Listener BioDocs FindFrames BioFiles LispDocs

Min Less More Max Format: Lisp Frame->Listener

-> | Slot Name | Slot Value

Slot Name	Slot Value
#^fName	GO.Acyl-CoaDehydrogenase
#^GO.DBXRefs	EC 1.3.99.3
#^GO.definition	Catalysis of the reaction: a
#^GO.ECRef	#$EC.1.3.99.3
#^GO.goid	3995
#^GO.Namespace	molecular_function
#^GO.prettyname	acyl-CoA dehydrogenase a
#^Go.products	#$MOL.2,3-Dehyd
#^Go.reactants	#$MOL.Acyl-C
#^Go.Related-Genes	#$Cwat.Cw?5697 wat
#^isA	#$Go.Reaction #$GO.Oxid
#^Source	GO
#^subClasses	#$GO.Very-Long-Chain-Ac #$GO.Butyryl-CoaDehydro

Parents
#$GO.Molecular_Function
..#$GO.CatalyticActivity
....#$GO.OxidoreductaseActivity
......#$GO.OxidoreductaseActivity,ActingOnTheCh-ChGrou
........#$GO.Acyl-CoaDehydrogenaseActivity
#$Go.Reaction
..#$GO.Acyl-CoaDehydrogenaseActivity

Children
#$GO.Acyl-CoaDehydrogenaseActivity
..#$GO.Very-Long-Chain-Acyl-CoaDehydrogenaseActivity
..#$GO.Short-Branched-Chain-Acyl-CoaDehydrogenaseAc
..#$GO.Isovaleryl-CoaDehydrogenaseActivity
..#$GO.Long-Chain-Acyl-CoaDehydrogenaseActivity

BioLingua Listener v4.1 Nostoc 8002 MT 5416 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://nostoc.stanford.edu:8002/redisplay.html?uid=25& =M Go

#$GO.ProtoporphyrinogenOxidaseActivity
#$GO.15,16-Dihydrobiliverdin:FerredoxinOxidoreductaseAct
<22>> (length *)
:: 1183
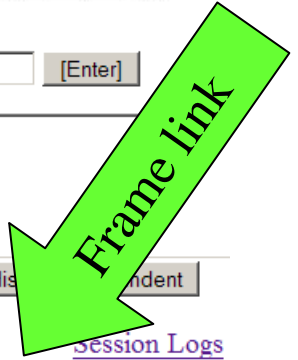<23>> (phylogeny-tree #$GO.Acyl-CoaDehydrogenaseActivity)

Frame links

Frame links

Frame link

TOP

0.05893 0.06066 0.17972

Tery_Te?6142   node1732

0.71184 0.18578

Cwat_Cw?5560   Cwat_Cw?1006

::
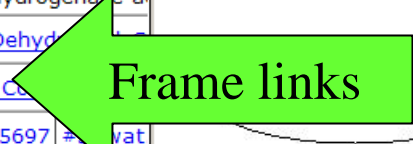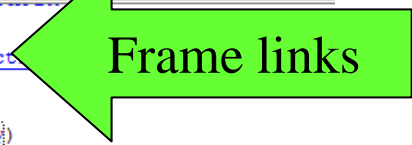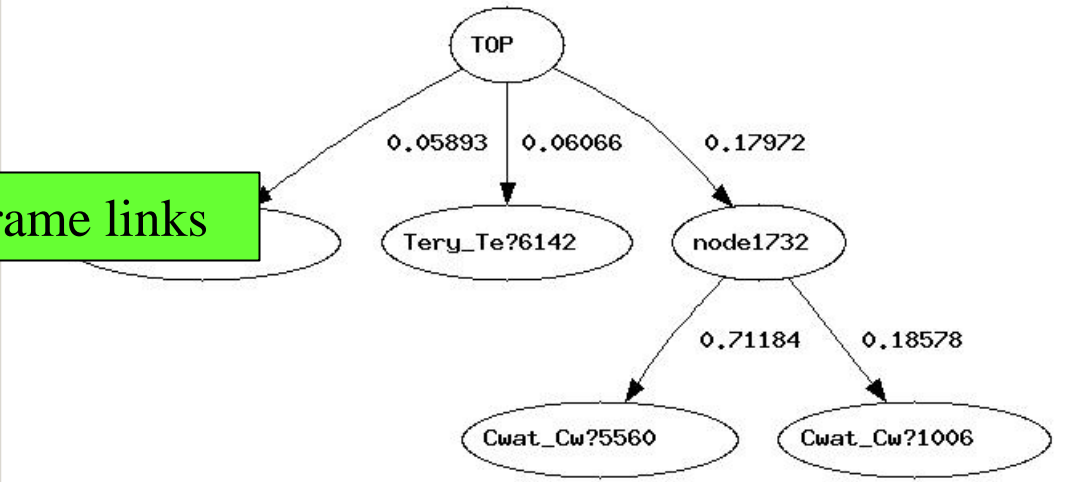
(phylogeny-tree #$GO.Acyl-CoaDehydrogenaseActivity) [Enter]

```
(defun phylogeny-tree (go-frame)
  (seegraph
   (run-phylip
    (#^alignments (align (#^Go.Related-Genes go-frame)))
    :labelfn #'first :seqfn #'second)))
```

Eval Clear Info: Package: MT   Clear His   ndent

Documentation   Bio Primitives   Browse BioFiles   Web Tools   Session Logs
Upload File   Files: Prev All   Lisp Hyperspec   Frame Browser   Feedback
BioLingua Listener v4.1

BioLingua Toplevel Frames

Listener BioDocs FindFrames BioFiles LispDocs

GO Frames

Go.Gene_Ontology   Go.Cellular_Component   Go.Biological_Process   Go.Molecular_Function   GO.Molecule   Go.Reaction
Go.Related-Genes   Go.reactantIn   Go.reactants   Go.productIn   Go.products   GO.directPrecursors   EC.Enzyme
EC.Nicezyme-link

KEGG Frames

Kegg.Thing   KEGG.Compound   Kegg.Reaction   Kegg.Glycan   Kegg.Enzyme   KEGG.name.s

OCELOT Frames

OC.Frames   OC.Publications   OC.Organisms   OC.Generalized-Reactions   OC.Enzymatic-Reactions   OC.Anatomical-
Structures   OC.People   OC.Chemicals   OC.Notes   OC.Databases

ORGANISMS Frames

thermosynechococcus_elongatus_bp1   crocosphaera_watsonii_wh8501   gloeobacter_violaceus_pcc7421
synechococcus_wh8102   synechococcus_elongatus_pcc7942   prochlorococcus_marinus_mit9313
prochlorococcus_marinus_ss120   prochlorococcus_marinus_med4   trichodesmium_erythraeum
anabaena_variabilis_atcc29413   anabaena_pcc7120   nostoc_punctiforme_atcc29133   synechocystis_pcc6803

Frame #$Kegg.Enzyme

Listener BioDocs FindFrames BioFiles LispDocs

Min Less More Max Format: Lisp Frame->Listener
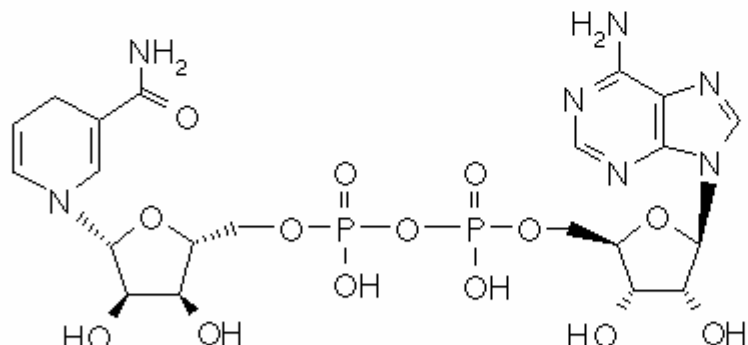
-> | *Slot Name* | *Slot Value*

#^allSubclasses	#$KEGG.Ec1.1.1.1	#$KEGG.Ec1.1.1.2	#$KEGG.Ec1.1.1.3	#$KEGG.Ec1.1.1.4	#$KEGG.Ec1.1.1.5
	#$KEGG.Ec1.1.1.6	#$KEGG.Ec1.1.1.7	#$KEGG.Ec1.1.1.8	#$KEGG.Ec1.1.1.9	#$KEGG.Ec1.1.1.10
	#$KEGG.Ec1.1.1.11	#$KEGG.Ec1.1.1.12	#$KEGG.Ec1.1.1.13	#$KEGG.Ec1.1.1.14	#$KEGG.Ec1.1.1.15
	#$KEGG.Ec1.1.1.16	#$KEGG.Ec1.1.1.17	#$KEGG.Ec1.1.1.18	#$KEGG.Ec1.1.1.19	#$KEGG.Ec1.1.1.20
	#$KEGG.Ec1.1.1.21	#$KEGG.Ec1.1.1.22	#$KEGG.Ec1.1.1.23	#$KEGG.Ec1.1.1.24	#$KEGG.Ec1.1.1.25
	#$KEGG.Ec1.1.1.26	#$KEGG.Ec1.1.1.27	#$KEGG.Ec1.1.1.28	#$KEGG.Ec1.1.1.29	#$KEGG.Ec1.1.1.30
	<<4280 more>>				
#^fName	Kegg.Enzyme				
#^isA	#$Kegg.Thing				
#^KEGG.product.s	NIL				
#^Kegg.Source.s	NIL				
#^KEGG.substrate.s	NIL				
#^Source	KEGG				
#^subClasses	#$KEGG.Ec6.6.1.2	#$KEGG.Ec6.6.1.1	#$KEGG.Ec6.5.1.4	#$KEGG.Ec6.5.1.3	#$KEGG.Ec6.5.1.2
	#$KEGG.Ec6.5.1.1	#$KEGG.Ec6.4.1.6	#$KEGG.Ec6.4.1.5	#$KEGG.Ec6.4.1.4	#$KEGG.Ec6.4.1.3
	#$KEGG.Ec6.4.1.2	#$KEGG.Ec6.4.1.1	#$KEGG.Ec6.3.5.10	#$KEGG.Ec6.3.5.9	#$KEGG.Ec6.3.5.8
	#$KEGG.Ec6.3.5.7	#$KEGG.Ec6.3.5.6	#$KEGG.Ec6.3.5.5	#$KEGG.Ec6.3.5.4	#$KEGG.Ec6.3.5.3
	#$KEGG.Ec6.3.5.2	#$KEGG.Ec6.3.5.1	#$KEGG.Ec6.3.4.17	#$KEGG.Ec6.3.4.16	#$KEGG.Ec6.3.4.15
	#$KEGG.Ec6.3.4.14	#$KEGG.Ec6.3.4.13	#$KEGG.Ec6.3.4.12	#$KEGG.Ec6.3.4.11	#$KEGG.Ec6.3.4.10

File Edit View Favorites Tools Help

Back · · · Search Favorites Media

Address http://nostoc.stanford.edu:8002/frame?pkg=JSHRAGER&printlimit=30&printispy=NIL&hideparents=T&hidech ▾ → Go

#^KEGG.genes.s	VVT: VVA3270	VPA: VPA0017	BPE: BP0373(ldh)
	HHE: HH1571(mdh)	CJE: Cj1167(ldh)	BSU: BG19003(ldh)
	BCE: BC1924 BC4870 BC4996	BCA: BCE2007(ldh) BCE5032 (ldh) BCE5135(ldh)	OIH: OB3279
	SAM: MW0217(lctE) MW2521	SEP: SE2145	LMO: lmo0210(ldh)
	<<34 more>>		
#^Kegg.link.s	http://www.genome.ad.jp/dbget-bin/www_bget?enzyme+1.1.1.27		
#^KEGG.motif.s	PS: PS00064 [LIVMA]-G-[EQ]-H-G-[DN]-[ST]		
#^KEGG.name.s	L-lactate dehydrogenase lactic acid dehydrogenase L(+)-nLDH L-(+)-lact L-lactic acid dehydrogenase lactate dehydrogenase lactate dehydrogenase NAD-dependent lactic deh		
#^KEGG.ortholog.s	KO: K00016 L-lactate dehydrogenase		
#^KEGG.pathway.s	PATH: MAP00010 Glycolysis / Gluconeogenesis PATH: MAP00272 Cysteine metabolism PATH: MAP00620		
#^KEGG.product.s	#$KEGG.C00080	#$KEGG.C00004 #$KEGG.C00022		
#^KEGG.reactstring.s	(S)-lactate + NAD+ = pyruvate + NADH + H+		
#^KEGG.reference.s	1 Dennis, D. and Kaplan, N.O. D and L-lactic acid dehydrogenase in Lactobacillus plantarum. J. Biol. Chem. :		
#^Kegg.Source.s	ENZYME		
#^KEGG.structures.s	PDB: 1EZ4 1CET 1A5Z 6LDH 1I10 8LDH 9LDB 9LDT 1CEQ 1PZG	1PZH 2LDB 2LDX 3LDH 5LDH :		
#^KEGG.substrate.s	#$KEGG.C00003	#$KEGG.C00186		
#^KEGG.sysname.s	(S)-lactate:NAD+ oxidoreductase		
#^Source	KEGG		

Parents

#$Kegg.Thing

Internet

File   Edit   View   Favorites   Tools   Help

Back   Search   Favorites   Media

Address   http://nostoc.stanford.edu:8002/frame?pkg=JSHRAGER&printlimit=30&printlispy=NIL&hideparents=T&hidech   Go

#^KEGG.accession.s	C00004				
#^KEGG.enzyme.s	#$KEGG.Ec1.1.1.1	#$KEGG.Ec1.1.1.2	#$KEGG.Ec1.1.1.3	#$KEGG.Ec1.1.1.4	#$KEGG.Ec1.1.1.5
	#$KEGG.Ec1.1.1.6	#$KEGG.Ec1.1.1.7	#$KEGG.Ec1.1.1.8	#$KEGG.Ec1.1.1.9	#$KEGG.Ec1.1.1.11
	#$KEGG.Ec1.1.1.12	#$KEGG.Ec1.1.1.13	#$KEGG.Ec1.1.1.14	#$KEGG.Ec1.1.1.15	#$KEGG.Ec1.1.1.16
	#$KEGG.Ec1.1.1.17	#$KEGG.Ec1.1.1.18	#$KEGG.Ec1.1.1.21	#$KEGG.Ec1.1.1.22	#$KEGG.Ec1.1.1.23
	#$KEGG.Ec1.1.1.24	#$KEGG.Ec1.1.1.26	#$KEGG.Ec1.1.1.27	#$KEGG.Ec1.1.1.28	#$KEGG.Ec1.1.1.29
	#$KEGG.Ec1.1.1.30	#$KEGG.Ec1.1.1.31	#$KEGG.Ec1.1.1.32	#$KEGG.Ec1.1.1.35	#$KEGG.Ec1.1.1.37
	<<365 more>>				
#^KEGG.formula.s	C21H29N7O14P2				
#^KEGG.GOframe.s	#$MOL.Nadh				
#^Kegg.image.s					
#^KEGG.name.s	NADH DPNH				
#^KEGG.pathway.s	PATH: MAP00190 Oxidative phosphorylation PATH: MAP00530 Aminosugars metabolism				
#^Kegg.Source.s	COMPOUND				
#^Source	KEGG				

Internet

```
#$S6803.sll1189 (ATTTTTTTA @ 25-34): glycolate oxidase subunit GlcE
#$S6803.slr1390 (ATTTTTTTA @ 78-87): cell division protein FtsH
#$S6803.slr1545 (ATTTTTTTA @ 121-130): RNA polymerase ECF-type (group 3) sigma-E factor
#$S6803.sll1710 (ATTTTTTTA @ 16-25): putative transposase [ISY523_b]
#$S6803.slr1738 (ATTTTTTTA @ 11-21): transcription regulator Fur family
#$S6803.sll1430 (ATTTTTTTA @ 11-20): adenine phosphoribosyltransferase
#$S6803.slr1334 (ATTTTTTTA @ 37-46): phosphoglucomutase/phosphomannomutase
#$S6803.trnN-GUU (ATTTTTTTA @ 133-142): tRNA-Asn (GUU); transfer RNA-Asn (GUU)
#$S6803.sll0337 (ATTTTTTTA @ 132-141): phosphate sensor, two-component sensor histidine kinase
#$S6803.ssl1533 (ATTTTTTTTA @ 86-96): unknown protein
#$S6803.slr0088 (ATTTTTTTA @ 158-167): beta-carotene ketolase
#$S6803.slr0039 (ATTTTTTTA @ 121-130): hypothetical protein
#$S6803.slr1655 (ATTTTTTTA @ 84-93): photosystem I subunit XI

> ((#$S6803.slr1655 84 93) (#$S6803.slr0039 121 130)
  (#$S6803.slr0088 158 167) (#$S6803.ssl1533 86 96)
  (#$S6803.sll0337 132 141) (#$S6803.trnN-GUU 133 142)
  (#$S6803.slr1334 37 46) (#$S6803.sll1430 11 20)
  (#$S6803.slr1738 11 21) (#$S6803.sll1710 16 25)
  (#$S6803.slr1545 121 130) (#$S6803.slr1390 78 87)
  (#$S6803.sll1189 25 34) (#$S6803.slr0242 188 197)
  (#$S6803.sll1061 10 19) (#$S6803.slr1488 11 21))
```

(find-promoted-genes2 :mints 7 :maxts 8)   [Enter]

Eval Clear Info: (find-promoted-

Clear History   Reindent

File Edit View Favorites Tools Help

Back · Search · Favorites · Media

Address http://nostoc.stanford.edu:8002/frame?PKG=JSHRAGER&NAME=S6803.slr1488

Frame #$S6803.slr1488

Listener BioDocs FindFrames BioFiles LispDocs

Min Less More Max Format: Lisp Frame->Listener

-> | *Slot Name* | *Slot Value*

Slot Name	Slot Value
#^ANNOTATION	multidrug resistance family ABC transporter
#^architecture	NIL
#^BEST-HIT-ACCESSION	ref|NP_486663.1|
#^BEST-HIT-ID-PCT	72
#^COG-DESCRIPTION	ABC transporter
#^COG-ID	COG1132
#^Contiguous-Sequence	#$S6803.chromosome
#^Direction	F
#^EC-DESCRIPTION	MDR (multidrug resistance) family ABC transporter
#^Encodes-protein	T
#^End-Unknown	NIL
#^fName	S6803.slr1488
#^From	36721
#^go-guesses	#$COG-DESCRIPTION \| 0.65686274 \| #$GO.WaterTransport | #$COG-DESCRIPTION \| 0.61111116 \| #$GO.SterolTransport | #
#^Go-Id	#$GO.Atp-BindingCassette[Abc]TransporterActivity \| #$GO.AtpBinding | #$GO.Membrane \| #$GO.Transport

Internet

File   Edit   View   Favorites   Tools   Help

Back   Search   Favorites   Media

Address   http://nostoc.stanford.edu:8002/frame?pkg=JSHRAGER&printlimit=30&printlispy=NIL&hideparents=T&hidech   Go

Frame #$GO.AtpBinding

Listener BioDocs FindFrames BioFiles LispDocs

Min Less More Max Format: Lisp Frame->Listen

-> | *Slot Name* | *Slot Value*

#^fName	GO.AtpBinding
#^GO.DBXRefs	ISBN 0198506732
#^GO.definition	Interacting selectively with triphosphate, a universally
#^GO.goid	5524
#^GO.Namespace	molecular_function
#^GO.prettyname	ATP binding

#^Go.Related-Genes	#$A7120.all8090	#$A7120.all8080	#$A7120.all8075	#$A7120.all8062	#$A7120.all8015
	#$A7120.alr9026	#$A7120.alr7635	#$A7120.alr7622	#$A7120.all7605	#$A7120.all7592
	#$A7120.all7583	#$A7120.alr7570	#$A7120.alr7552	#$A7120.alr7528	#$A7120.all7353
	#$A7120.alr7297	#$A7120.alr7295	#$A7120.all7229	#$A7120.all7227	#$A7120.all7218
	#$A7120.alr7200	#$A7120.all7196	#$A7120.all7172	#$A7120.alr7157	#$A7120.alr7129
	#$A7120.alr7046	#$A7120.all7026	#$A7120.alr7014	#$A7120.all7012	#$A7120.alr5362
	<<530 more>>				

#^isA	#$GO.AdenylNucleotideBinding
#^Source	GO

Parents
#$Go.Molecular_Function

(Inset popup window:)

#$A7120.all7646   #$A7120.all7620

<<530 more>>

#^isA	#$GO.AdenylNucleotideBinding
#^Source	GO

Parents
#$Go.Molecular_Function
..#$GO.Binding
....#$GO.NucleotideBinding
......#$GO.PurineNucleotideBinding
........#$GO.AdenylNucleotideBinding
..........**#$GO.AtpBinding**

Internet

KnowOS Approach

Turn Lisp into an Operating System

1. Start with ACL – fast compiler, multi-process model.

2. Run it on a server, accessed via a browser-based listener.

3. Integrate knowledge bases via a built-in frame system.

4. Run it on Linux – external tools, security model.

5. Rebuild user tools (editing, file manip., debugging, etc.)

6. Provide XML-RPC "Evalserver" for others to call in.

7. Try to avoid crashing it ("apparent persistence").

Frame #$GO.Acyl-CoaDehydrogenaseActivity

Listener BioDocs FindFrames BioFiles LispDocs

Min Less More Max Format: Lisp Frame->Listener

-> | **Slot Name** | **Slot Value**

#^fName	GO.Acyl-CoaDehydrogenase
#^GO.DBXRefs	EC│1.3.99.3
#^GO.definition	Catalysis of the reaction: a
#^GO.ECRef	#$EC.1.3.99.3
#^GO.goid	3995
#^GO.Namespace	molecular_function
#^GO.prettyname	acyl-CoA dehydrogenase a
#^Go.products	#$MOL.2,3-Dehydroacyl-C
#^Go.reactants	#$MOL.Acyl-Coa │ #$MOL.A
#^Go.Related-Genes	#$Cwat.Cw?5697 │ #$Cwat
#^isA	#$Go.Reaction │ #$GO.Oxid
#^Source	GO
#^subClasses	#$GO.Very-Long-Chain-Ac #$GO.Butyryl-CoaDehydro

Parents
#$GO.Molecular_Function
..#$GO.CatalyticActivity
....#$GO.OxidoreductaseActivity
......#$GO.OxidoreductaseActivity,ActingOnTheCh-ChGrou
........#$GO.Acyl-CoaDehydrogenaseActivity
#$Go.Reaction
..#$GO.Acyl-CoaDehydrogenaseActivity

Children
#$GO.Acyl-CoaDehydrogenaseActivity
..#$GO.Very-Long-Chain-Acyl-CoaDehydrogenaseActivity
..#$GO.Short-Branched-Chain-Acyl-CoaDehydrogenaseA(
..#$GO.Isovaleryl-CoaDehydrogenaseActivity
..#$GO.Long-Chain-Acyl-CoaDehydrogenaseActivity

#$GO.ProtoporphyrinogenOxidaseActivity
#$GO.15,16-Dihydrobiliverdin:FerredoxinOxidoreductaseActivity ..
<22>> (length *)
:: 1183
<23>> (phylogeny-tree #$GO.Acyl-CoaDehydrogenaseActivity)

Dotty output

Call Dotty

Call Phylip

Call Clustal


```
(phylogen                    eActivity)        [Enter]

(defun                gen
  (seegraph
   (run-phylip
    (#^alignments (align (#^Go.Related-Genes go-frame)))
     :labelfn #'first :seqfn #'second)))
```

Eval   Clear   Info: Package: MT   Clear History   Reindent

Documentation   Bio Primitives   Browse BioFiles   Web Tools   Session Logs
Upload File   Files: Prev All   Lisp Hyperspec   Frame Browser Feedback

BioLingua Listener v4.1

KnowOS Approach

Turn Lisp into an Operating System

1. Start with ACL – fast compiler, multi-process model.

2. Run it on a server, accessed via a browser-based listener.

3. Integrate knowledge bases via a built-in frame system.

4. Run it on Linux – external tools, security model.

5. Rebuild user tools (editing, file manip., debugging, etc.)

6. Provide XML-RPC "Evalserver" for others to call in.

7. Try to avoid crashing it ("apparent persistence").

BioLingua Directory Listing

Listing for /home/visitors/jshrager/

New file or subdirectory   Delete directory files

---

up ..
dir bkup
dir session-logs
dir stem

edit	+cih+.lisp	06/08/05 15:55	edit	jnk.cons	09/10/04 13:21
edit	+cih-cache+.lisp	06/10/05 09:06	edit	jnk3.cons	09/10/04 13:21
edit	200409271144clean.txt	09/29/04 11:04	edit	kid.xml	04/25/05 16:44
edit	arglist-data.js	06/10/05 09:01	edit	lysis.cons	04/25/05 16:00
edit	biolisp.ext	12/10/04 05:51	edit	lysis.lisp	12/10/04 08:11
edit	biolisp.ini	12/09/04 09:07	edit	make-new-temp-file-path.src	05/02/05 11:51
edit	cjtu.lisp	06/10/05 09:05	edit	merge.cons	09/10/04 13:21
edit	count-base.src	05/02/05 12:14	edit	misc.lisp	03/03/05 11:34
edit	filepath-utils.lisp	03/06/05 16:26	edit	neuro.cons	09/10/04 13:21
edit	ql.src	11/30/04 11:56	edit	nhs.lisp	05/02/05 15:51
edit	qoexamples.lisp	01/20/05 15:01	edit	ocean.lisp	04/25/05 16:42
edit	hihara.tbl	08/07/03 19:40	edit	photo.cons	09/10/04 13:21
edit	hybrow.lisp	01/24/05 23:36	edit	scraper.lisp	04/28/05 07:46
edit	je.lisp	04/03/04 11:02	edit	seegraph.src	05/03/05 14:10
edit	jkf-setup.src	05/02/05 11:24	edit	snpscoretest.lisp	10/04/04 12:24
edit	jkf.cons	10/04/04 10:30	edit	uberprims.lisp	03/06/05 16:34
edit	jkf.lisp	05/03/05 12:54	edit	uberprims.lisp~	03/06/05 16:29

BioLingua Lisp File Listing
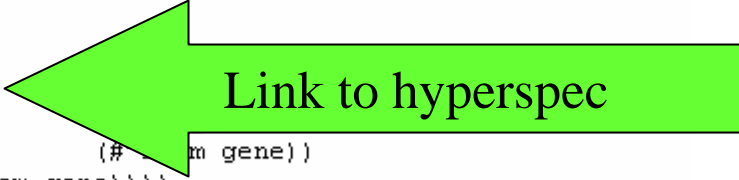
/home/visitors/jshrager/cjtu.lisp   Download source   Edit source

```
;;; Another approach, run through all the genes, grab the +-edgelimit
;;; and see if there are one or more occurrances of the relevant
;;; pattern in that segment.  The complexity here is to ensure that we
;;; don't grab inter-region that overlaps into the next gene.
(defun find-promoted-genes2 (&key (mints 6) (maxts 7) (edgelimit 200))
  (let* (
            (promex (format nil "A(T(~a
            (cs (extract-sequence #$
            (genevec (#^Genes-Sorted-by
            )
    (loop with last = nil
          with hits = nil
          for gene across genevec
          do (cond ((null last) nil)
                   (t (let ((interseq (subseq
                                             (#      m gene))
                             (#^from gene))))
                      (ppcre::do-matches (ms me promex interseq)
                        (format t "~a (~a @ ~a-~a): ~a~%"
                                gene (subseq interseq ms me) ms me
                                (#^annotation gene))
                        (push (list gene ms me) hits)))))
            (setq last gene)
            finally (return hits)
            )))
```

Link to internal code

Link to hyperspec

/home/visitors/jshrager/ cjtu.lisp   [Return to Listener]

Documentation   Bio Primitives   Browse BioFiles   Web Tools   Session Logs
Upload File   Files: Prev All   Lisp Hyperspec   Frame Browser   Feedback

Save   Save-Compile-Load   Reindent   New   Info:

```
;;; Another approach, run through all the genes, grab the +-edgelimit
;;; and see if there are one or more occurrances of the relevant
;;; pattern in that segment.  The complexity here is to ensure that we
;;; don't grab inter-region that overlaps into the next gene.

(defun find-promoted-genes2 (&key (mints 6) (maxts 7) (edgelimit 200))
  (let* (
          (promex (format nil "A(T{~a,~a})A" mints maxts))
          (cs (extract-sequence #$S6803.chromosome))
          (genevec (#^Genes-Sorted-by-Position #$S6803.chromosome))
          )
     (loop with last = nil
           with hits = nil
           for gene across genevec
           do (cond ((null last) nil)
                    (t (let ((interseq (subseq cs (min (max (#^to last)
                                                            (- (#^from gene)
edgelimit))
                                                       (#^from gene))
```

SESSION LOGS FOR USER JSHRAGER

Listener BioDocs FindFrames BioFiles LispDocs

Link	Last entry	Log file name
Log 1	Today, at 09:06 AM	ryoy-20050610-090106.log
Log 2	Wednesday (2 days ago), at 05:17 PM	ruwj-20050608-154838.log
Log 3	Wednesday (2 days ago), at 02:36 PM	rurv-20050608-135653.log
Log 4	Monday (4 days ago), at 02:34 PM	rqne-20050606-143203.log
Log 5	Friday (7 days ago), at 07:51 AM	rjna-20050603-075106.log
Log 6	Wed, May 25, 2005 at 08:49 AM	qqkv-20050525-084934.log
Log 7	Tue, May 24, 2005 at 10:19 PM	qpnn-20050524-221832.log
Log 8	Fri, May 20, 2005 at 10:59 PM	qhag-20050520-225417.log
Log 9	Tue, May 17, 2005 at 10:44 PM	qaqa-20050517-224417.log
Log 10	Mon, May 16, 2005 at 10:20 PM	pyml-20050516-221954.log
Log 11	Fri, May 13, 2005 at 08:10 AM	pqvy-20050513-081007.log
Log 12	Fri, May 13, 2005 at 01:10 AM	pqbn-20050512-235500.log
Log 13	Tue, May 03, 2005 at 05:56 PM	owje-20050503-175112.log
Log 14	Tue, May 03, 2005 at 04:40 PM	owgq-20050503-162238.log
Log 15	Tue, May 03, 2005 at 04:22 PM	owgq-20050503-162228.log
Log 16	Tue, May 03, 2005 at 04:02 PM	owgm-20050503-160202.log
Log 17	Tue, May 03, 2005 at 02:17 PM	ovxq-20050503-125152.log
Log 18	Mon, May 02, 2005 at 03:51 PM	otwp-20050502-135945.log
Log 19	Mon, May 02, 2005 at 12:15 PM	otrv-20050502-113816.log
Log 20	Mon, May 02, 2005 at 11:33 AM	otrs-20050502-112211.log
Log 21	Sat, Apr 30, 2005 at 09:21 PM	ooeo-20050430-211305.log
Log 22	Sat, Apr 30, 2005 at 09:05 PM	ooem-20050430-210321.log
Log 23	Sat, Apr 30, 2005 at 05:33 PM	ontd-20050430-162743.log
Log 24	Thu, Apr 28, 2005 at 10:34 AM	oiyn-20050428-102603.log
Log 25	Thu, Apr 28, 2005 at 07:46 AM	oipn-20050428-065702.log
Log 26	Mon, Apr 25, 2005 at 05:37 PM	odci-20050425-164437.log
Log 27	Mon, Apr 25, 2005 at 04:42 PM	odad-20050425-155601.log
Log 28	Mon, Apr 25, 2005 at 01:10 PM	ocqq-20050425-112028.log

```
;;; Compiling file /home/visitors/jshrager/cjtu.lisp
;;; Writing fasl file /home/visitors/jshrager/cjtu.fasl
;;; Fasl write complete
; Fast loading /home/visitors/jshrager/cjtu.fasl

06/08/05 16:53:43 : Out: T
06/08/05 16:53:51 : In : (find-promoted-genes2 :edgelimit 50)
06/08/05 16:53:51 : <<< Evaluation failed. Actual error: >>>
<<< In SUBSEQ, `start' (16743) is greater than `end' (16740).>>>
<<< Use (EXPLAIN) to see a stack trace. >>>


06/08/05 16:53:58 : In : (find-promoted-genes2 :edgelimit 50)
06/08/05 16:53:58 : <<< Evaluation failed. Actual error: >>>
<<< In SUBSEQ, `start' (16743) is greater than `end' (16740).>>>
<<< Use (EXPLAIN) to see a stack trace. >>>


06/08/05 16:55:01 : In : T
06/08/05 16:55:01 : Out: T
06/08/05 16:55:05 : In : (explain)
06/08/05 16:55:05 : Eval. printout:


(EXCL::.ERROR "~@[In ~a~], `start'~@[ (~s)~] is greater than `end'~@[ (~s)~]."
              SUBSEQ 16743 16740)
(EXCL::.SEQUENCE-BOUNDS-ERROR SUBSEQ 16743 16740)
(SUBSEQ "GGCGCGCCATCGCCGGCTGGCGGAAATTACCGAAATGATCCACACCGCTAGTTTGGTCCACGATGACGTGGTGGATGAGGC(
...
*** OUTPUT OF 3573920 CHARACTERS TRUNCATED.
*** CHANGE VALUE OF *OUTPUT-LIMIT* (CURRENTLY 10000) TO SEE MORE
*** THE FIRST 400 CHARACTERS OF OUTPUT ARE SHOWN ABOVE, AND THE LAST 200 ARE SHOWN BELOW.
...
LAMBDA 253))
((METHOD WEBLISTENER::REPL-EXECUTE-FORM (T T T T))
 #<COMMON-LISP-USER:BIOLINGUA @ #x717c742a>
 #<Function (:ANONYMOUS-LAMBDA 253) @ #x95eb8cf2>
 (FIND-PROMOTED-GENES2 :EDGELIMIT 50) T)
06/08/05 16:55:05 : Out: NIL
06/08/05 16:55:49 : In : (load (compile-file "/home/visitors/jshrager/cjtu.lisp"))
06/08/05 16:55:49 : Eval. printout:
```

Done

Documentation

LOAD-HIHARA-TABLE   (&KEY FORCE-RELOAD? FORCE-RECOMPILE?)

PRELOAD-ORGANISMS   (&OPTIONAL ORGANISM-NAMES)

Loads auxiliary information for all the organisms available to this instanc...

DUMP-HASH-TO-FILE   (HT FILE &KEY HASH-TABLE-NAME HASH-TABLE...)

Writes out a hash table to a file or stream, in a way that the hash table ...

SELECT-MICROARRAY-TABLE-DATA (TABLE TEST)

Given a table, loaded by read-microarray-table, and a test with two args: ...

RESTORE-WORKSPACE   (&KEY PACKAGE VERSION USER DIRECTORY-PAT...)

Load a workspace file designated by PACKAGE and VERSION and located in USER...

LOAD-MODULE   (NAME &KEY USE-MODULE? SEARCH-DIRECTORIE...)

Loads a module, possibly created by CREATE-LOADABLE-MODULE, or represented ...

```
;; There are 3 other related symbols not shown.  Use
;; (help "LOAD A TABLE" :all t) to see them all.
```

(help "load a table")   [Enter]

Eval   Clear   Info: Package: JSHRAGER   Clear History   Reindent

Documentation   **Bio Primitives**   **Browse BioFiles**   **Web Tools**   **Session Logs**

Upload File   **Files: Prev All**   **Lisp Hyperspec**   **Frame Browser**   **Feedback**

BioLingua Listener v5.0

BioLingua Doc

File Edit View

Back

Address http://

Do

All contents (specified)

About BioLingua

Tutorial examples

News!

- 20050607 our colla

Done   Internet

KnowOS Approach

Turn Lisp into an Operating System

1. Start with ACL – fast compiler, multi-process model.

2. Run it on a server, accessed via a browser-based listener.

3. Integrate knowledge bases via a built-in frame system.

4. Run it on Linux – external tools, security model.

5. Rebuild user tools (editing, file manip., debugging, etc.)

6. Provide XML-RPC "Evalserver" for others to call in.

7. Try to avoid crashing it ("apparent persistence").

KnowOS Approach

Turn Lisp into an Operating System

1. Start with ACL – fast compiler, multi-process model.

2. Run it on a server, accessed via a browser-based listener.

3. Integrate knowledge bases via a built-in frame system.

4. Run it on Linux – external tools, security model.

5. Rebuild user tools (editing, file manip., debugging, etc.)

6. Provide XML-RPC "Evalserver" for others to call in.

7. Try to avoid crashing it ("apparent persistence").

Plan of the Presentation

* Overview of KnowOS services

* Example 1: BioLingua biological knowledge environment

* Example 2: CACHE collaborative knowledge analysis

* Issues and approaches

* Near and long term goals

The BioLingua Vision:
Biologist as Programmer

Give biologists a program and they'll make you program more and more.

But give them an ***integrated knowledge and programming environment***, and teach them to use it, and you'll change their lives!

(Not to mention saving yourself a lot of boring programming!)

Current Best Practice:
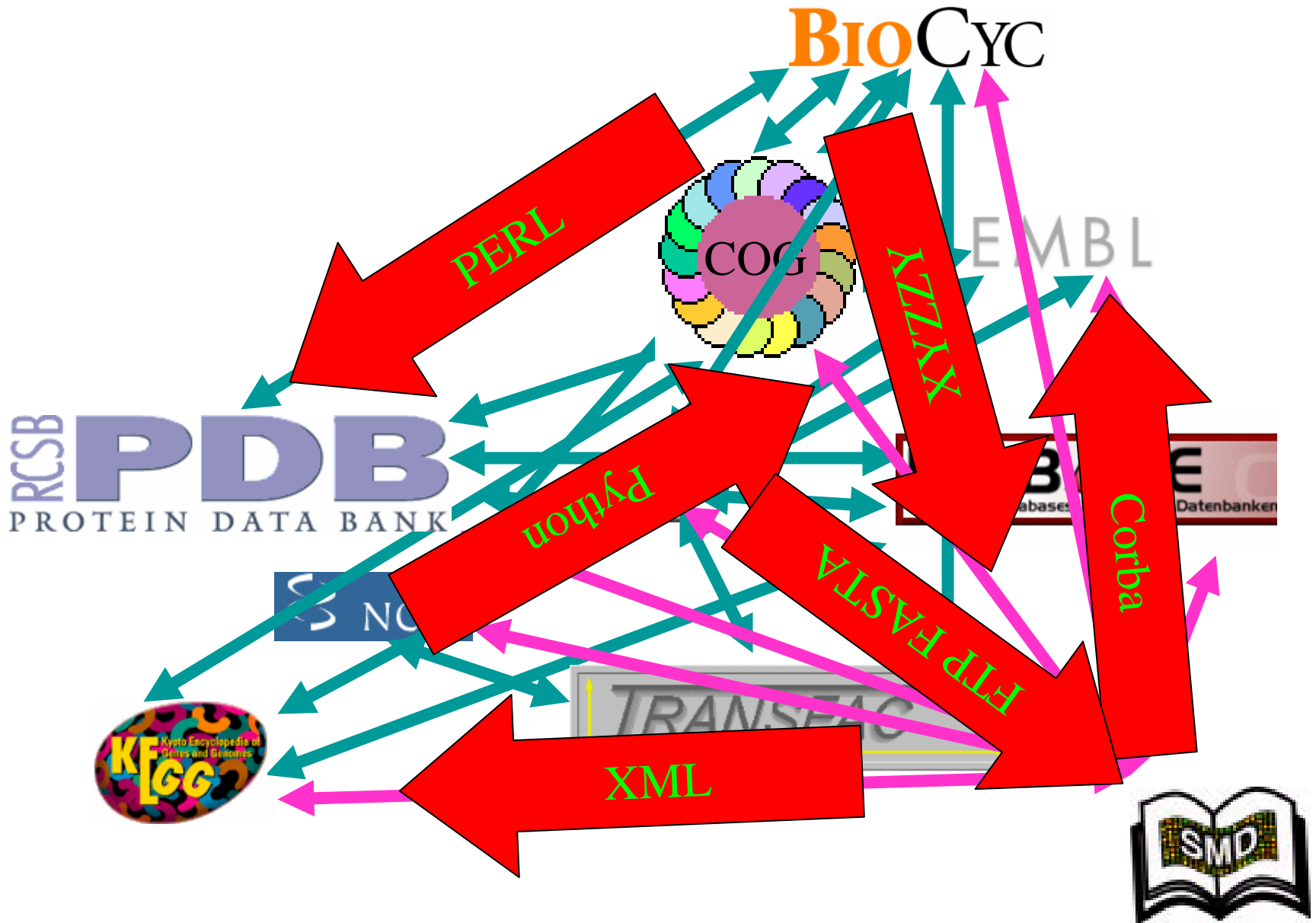
KnowOS Approach:

Integrated DBs
On central server

BIOCYC

EMBL

COG

GO

K**E**GG

RCSB **PDB**
PROTEIN DATA BANK

NCBI

Microarray DB

Organism
Models

#$trichodesmium_erythraeum #$anabaena_variabilis_atcc29413
#$synechocystis_pcc6803 #$prochlorococcus_marinus_ccmp1375
#$anabaena_pcc7120 #$nostoc_punctiforme_atcc29133 ...

SMD
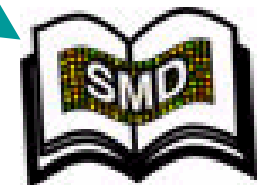
BioLingua Prime Directive:

All data and knowledge can be
manipulated by user-written program
that approximate user's natural protocols.

```
(loop for pm4gene in (#^Genes ProcMed4)
    as all-orthologous =  (all-blast-orthologs pm4gene)
    as 6803ortholog = (intersect (#^Genes Syny6803) all-orthologous)
    when (and (not-any #'member-geneid
                        (#^Genes slotv Proc9313) all-orthologous))
                (any #'member-geneID 6803ortholog)
                (>= ma-ratio (ma-select 6803ortholog Hihara1) 2)))
    collect light-specific-genes 6803ortholog)
```

Count the genes of an organism.

```
<3>> *load-organisms*
:: Attempt to take the value of the unbound variable `*LOAD-ORGANISMS*'.
<4>> *loaded-organisms*
:: (#$anabaena_pcc7120 #$musmusculus #$nostoc_punctiforme_atcc29133
   #$promed4 #$anabaena7120 #$synechocystis6803)
<5>> (setq s #$synechocystis6803)
:: #$synechocystis6803
<6>> (#^genes s)
:: (#$sll0400 #$sll0401 #$sll0402 #$sll0403 #$sll0404 #$sll0405 #$sll0406
   #$sll0408 #$sll0409 #$sll0410 #$sll0412 #$sll0413 #$sll0414 #$sll0415
   #$sll1500 #$sll0416 #$sll1501 #$sll1502 #$sll0418 #$sll1503 #$sll0419
   #$ssl0312 #$sll1504 #$sll1505 #$sll1507 #$sll1508 #$sll1509 #$ssl0318
   #$sll0420 #$sll0421 #$sll0422 #$sll0423 #$sll0424 #$sll1510 #$sll0426
   #$sll1511 #$sll0427 #$sll1512 #$sll0428 #$sll1513 #$sll1514 #$ssl0323
   #$sll1515 #$sll1516 #$ssl1417 #$ssl2501 #$ssl2502 #$ssl2507 #$sll0430
   #$sll0431 #$sll1520 #$sll0436 #$sll1521 #$sll1522 #$ssl0331 #$sll1524
   #$sll1525 #$sll1526 #$sll1527 #$sll1528 #$ssl1426 #$sll0441 #$sll0442
   #$sll0443 #$sll0444 #$sll0445 #$sll1530 #$sll0446 #$sll1531 #$sll0447
   #$sll1532 #$sll0448 #$sll1533 #$sll0449 #$sll1534 #$sll1535 #$sll1536
   #$sll1537 #$sll1538 #$slr0400 #$slr0401 #$slr0402 #$slr0404 #$slr0406
   #$slr0407 #$slr0408 #$ssl3615 #$sll0450 #$sll0451 #$sll0454 #$sll0455
   #$sll1540 #$sll0456 #$sll1541 #$ssl0350 #$sll1542 #$sll1543 #$sll0459
   #$ssl0352 #$sll1544 ...)
<7>> (length *)
:: 3264
```

How many of those are transporters?

```
<10>> (remove-if-not #'(lambda (g) (member #$transport (#^go-frames g))) (#^genes s)))
:: (#$sll0415 #$sll1521 #$sll1533 #$sll0477 #$sll0478 #$sll0479 #$sll1599
   #$slr1596 #$slr2045 #$slr2077 #$slr2097 #$sll1600 #$sll0537 #$sll0550
   #$sll0594 #$sll1685 #$sll1699 #$sll1017 #$sll1082 #$sll1087 #$sll0616
   #$sll0640 #$sll1762 #$sll0679 #$slr1728 #$slr1730 #$slr1740 #$slr0677
   #$slr0678 #$sll1104 #$slr0019 #$sll0064 #$sll1164 #$slr0063 #$slr0075
   #$slr0079 #$sll0771 #$slr0753 #$sll0108 #$sll1206 #$sll0174 #$sll0182
   #$sll1270 #$slr0161 #$slr1257 #$slr1270 #$slr1908 #$slr0875 #$slr1962
   #$sll0217 #$sll0219 #$slr0896 #$sll0224 #$sll1314 #$sll1319 #$slr1316
   #$slr1317 #$sll1374 #$slr1392 #$sll1404 #$sll1405 #$sll1406 #$sll1409
   #$sll0355 #$sll1450 #$sll1451 #$sll1452 #$slr0341 #$slr1452 #$slr1453
   #$slr1454 #$slr1455 #$ssr1386 #$slr1490)
<11>> (length *)
:: 74
```

Slots

DBLINKS	(#$CY #$sll0415 () () () () ())
fName	sll0415
go-frames	(#$Atp-BindingCassette[Abc]Transporter #$ATPBinding #$ATPBinding #$Transport)
Left-End-Position	2540821
organism	#$synechocystis6803
PRODUCT	#$Sll0415-Monomer
Right-End-Position	2540821
SequenceInfo	(2540821 2541885 354 c)
source	ocelot

Frame #$Glycolysis

slots

Assume-Unique-Enzymes	no
Common-Name	glycolysis
definition	The breakdown of a monosaccharide (generally glucose) into simpler components, including pyruvate. definition_
Deltag0	-8.5d0
fName	Glycolysis
Goid	6096
In-Pathway	(#$Glucfermen-Pwy #$Glycolysis+Citric-Acid-Pwy)
IsaLevel	5
Net-Reaction-Equation	Glucose + 2 Pi + 2 ADP + 2 NAD = 2 pyruvate + 2 ATP + 2 NADH + 2 H + 2 H(2)O
Pathway-Links	((#$Pyruvate #$Fermentation-Pwy #$Amino-Acid-Biosynthesis) (#$Glc-6-P #$Pentose-P-Pwy))
PREDECESSORS	((#$Pglucisom-Rxn) (#$6pfructphos-Rxn #$Pglucisom-Rxn) (#$F16aldolase-Rxn #$6pfructphos-Rxn) (#$Triosepisomerization-Rxn #$F16aldolase-Rxn) (#$Gapoxnphosphn-Rxn #$Triosepisomerization-Rxn #$F16aldolase-Rxn) (#$Phosglyphos-Rxn #$Gapoxnphosphn-Rxn) (#$3pgarearr-Rxn #$Phosglyphos-F #$2pgadehydrat-Rxn #$3pgarearr-Rxn) (#$Pepdephos-Rxn #$2pgadehydrat-Rxn))
Prettyname	glycolysis
Reaction-List	(#$Pglucisom-Rxn #$6pfructphos-Rxn #$F16aldolase-Rxn #$Triosepisomerization-Rxn #$Gapoxnphos Rxn #$Phosglyphos-Rxn #$3pgarearr-Rxn #$2pgadehydrat-Rxn #$Pepdephos-Rxn)
source	ocelot
Super-Pathways	(#$Glucfermen #$Glycolysis+Citric-Acid-Pwy)
SYNONYMS	Embden-Meyerhof hway

Internet

Frame #$Pglucisom-Rxn

Slots	
Common-Name	Glucose-6-phosphate isomerase
DBLINKS	(([#$PIR](#) T51720 () pkarp 3199817462) ([#$PIR](#) T46970 () pkarp 3199817462) ([#$PIR](#) T44843 () pkarp 3199817462) ([#$PIR](#) T43196 () pkarp 3199817462) ([#$PIR](#) T14631 () pkarp 3199817462) ([#$PIR](#) T0915 pkarp 3199817462) ([#$PIR](#) T09153 () pkarp 3199817462) ([#$PIR](#) T05572 () pkarp 3199817462) ([#$PIR](#) () pkarp 3199817462) ([#$PIR](#) T03948 () pkarp 3199817462) ([#$PIR](#) T02094 () pkarp 3199817462) ([#$PIR](#) S75607 () pkarp 3199817462) ([#$PIR](#) S73908 () pkarp 3199817462) ([#$PIR](#) S58164 () pkarp 31998174 ([#$PIR](#) S57831 () pkarp 3199817462) ([#$PIR](#) S57830 () pkarp 3199817462) ([#$PIR](#) S41808 () pkarp 3199817462) ([#$PIR](#) S41807 () pkarp 3199817462) ([#$PIR](#) S41806 () pkarp 3199817462) ([#$PIR](#) S235 pkarp 3199817462) ([#$PIR](#) NUZQF () pkarp 3199817462) ([#$PIR](#) NUVKL () pkarp 3199817462) ([#$PIR](#) pkarp 3199817462) ([#$PIR](#) NUPG () pkarp 3199817462) ([#$PIR](#) NUMS () pkarp 3199817462) ([#$PIR](#) NU pkarp 3199817462) ([#$PIR](#) NUBY () pkarp 3199817462) ([#$PIR](#) NUBSSA () pkarp 3199817462) ([#$PIR](#) pkarp 3199817462) ([#$PIR](#) I48073 () pkarp 3199817462) [...15 more...](#))
Deltag0	0.4d0
ec-number	5.3.1.9
Enzymatic-Reaction	[#$Enzrxn-256](#)
fName	Pglucisom-Rxn
In-Pathway	([#$Sucsyn-Pwy](#) [#$P341-Pwy](#) [#$Glycolysis](#) [#$Gluconeo-Pwy](#) [#$P124-Pwy](#) [#$P125-Pwy](#))
LEFT	[#$Glc-6-P](#)
RIGHT	[#$Fructose-6p](#)
source	ocelot
SYNONYMS	(Hexose monophosphate isomerase Phosphohexoisomerase Phosphoglucoisomerase Phosphosaccharom Hexosephosphate isomerase Oxoisomerase Phosphohexomutase Phosphohexose isomerase Phosphoglu isomerase)

Find the genes involved in glycolysis, and their reactions.

```
<38>> (remove-if-not #'(lambda (g) (member #$glycolysis (#^go-frames g))) (#^gene
:: (#$sll0587 #$sll0593 #$slr1096 #$sll0018 #$sll1196 #$sll0745 #$slr0752
   #$slr0783 #$sll1275 #$slr1945 #$slr0884 #$sll1342 #$slr1349 #$slr0943
   #$slr0394)
<39>> (first *)
:: #$sll0587
<40>> (#^product *)
:: #$Sll0587-Monomer
<41>> (#^catalyzes *)
:: #$Enzrxn-491
<42>> (#^reaction *)
:: #$Pepdephos-Rxn
<43>> (#^left *)
:: (#$Pyruvate #$Atp)
<44>> (defun extract-reaction (gene)
   (ignore-errors (let ((reaction (#^reaction (#^catalyzes (#^product gene)))))
      (list gene (#^left reaction) (#^right reaction)))))
:: EXTRACT-REACTION
<45>> (mapcar #'extract-reaction (hop 38))
:: ((#$sll0587 (#$Pyruvate #$Atp)
   (#$Phospho-Enol-Pyruvate #$Adp #$PROTON))
 NIL
 (#$slr1096 (#$Red-Dihydrolipoamide #$NAD)
   (#$Ox-Lipoamide #$Nadh #$PROTON))
 (#$sll0018 #$Fructose-16-Diphosphate
   (#$Dihydroxy-Acetone-Phosphate #$GAP))
 NIL NIL (#$slr0752 #$2-Pg (#$Phospho-Enol-Pyruvate #$WATER))
 (#$slr0783 #$GAP #$Dihydroxy-Acetone-Phosphate)
 (#$sll1275 (#$Pyruvate #$Atp)
   (#$Phospho-Enol-Pyruvate #$Adp #$PROTON))
 (#$slr1945 #$G3p #$2-Pg)
```

BioLingua: Frame #$GO.Acyl-CoaDehydrogenaseActivity - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://nostoc.stanford.edu:8002/frame?PKG=MT&NAME=GO.Acyl-CoaDehydrogen ▾ ▶ Go G

Frame #$GO.Acyl-CoaDehydrogenaseActivity

Listener BioDocs FindFrames BioFiles LispDocs

Min Less More Max Format: Lisp Frame->Listener

-> | *Slot Name* | *Slot Value*

#^fName	GO.Acyl-CoaDehydrogenase
#^GO.DBXRefs	EC|1.3.99.3
#^GO.definition	Catalysis of the reaction: a
#^GO.ECRef	#$EC.1.3.99.3
#^GO.goid	3995
#^GO.Namespace	molecular_function
#^GO.prettyname	acyl-CoA dehydrogenase a
#^Go.products	#$MOL.2,3-Dehydroacyl-C
#^Go.reactants	#$MOL.Acyl-Coa \| #$MOL.A
#^Go.Related-Genes	#$Cwat.Cw?5697 \| #$Cwat
#^isA	#$Go.Reaction \| #$GO.Oxide
#^Source	GO
#^subClasses	#$GO.Very-Long-Chain-Ac / #$GO.Butyryl-CoaDehydro

Parents
#$GO.Molecular_Function
..#$GO.CatalyticActivity
....#$GO.OxidoreductaseActivity
......#$GO.OxidoreductaseActivity,ActingOnTheCh-ChGrou
........#$GO.Acyl-CoaDehydrogenaseActivity
#$Go.Reaction
..#$GO.Acyl-CoaDehydrogenaseActivity

Children
#$GO.Acyl-CoaDehydrogenaseActivity
..#$GO.Very-Long-Chain-Acyl-CoaDehydrogenaseActivity
..#$GO.Short-Branched-Chain-Acyl-CoaDehydrogenaseAd
..#$GO.Isovaleryl-CoaDehydrogenaseActivity
..#$GO.Long-Chain-Acyl-CoaDehydrogenaseActivity

BioLingua Listener v4.1 Nostoc 8002 MT 5416 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://nostoc.stanford.edu:8002/redisplay.html?uid=25&pkg=M ▾

#$GO.ProtoporphyrinogenOxidaseActivity
#$GO.15,16-Dihydrobiliverdin:FerredoxinOxidoreductaseActivity ..
<22>> (length *)
:: 1183
<23>> (phylogeny-tree #$GO.Acyl-CoaDehydrogenaseActivity)

Dotty output

TOP

0.05893   0.06066   0.17972

Cwat_Cw?5697   Tery_Te?6142   node1732

0.71184   0.18578

Cw?5560   Cwat_Cw?1006

::

(phyloge... eb... eActivity)   [Enter]

Call Dotty
Call Phylip
Call Clustal

```
(defun phylogen
  (seegraph
   (run-phylip
     (#^alignments (align (#^Go.Related-Genes go-frame)))
     :labelfn #'first :seqfn #'second)))
```

Eval   Clear   Info: Package: MT   Clear History   Reindent

Documentation   Bio Primitives   Browse BioFiles   Web Tools   Session Logs
Upload File   Files: Prev All   Lisp Hyperspec   Frame Browser   Feedback

BioLingua Listener v4.1

BioLingua-Lite (Jeff Elhai, James Mastros, and others @ VCU)

Challenge problem:

Find 100 bp of sequence upstream from a set of orthologs for all genes in an organism and align them.

SEED Version:

```
for i in `pegs $1`
do
    (echo "$i"; echo "$i" | similar_to 1.0e-50 | is_prokaryotic | head -n
40 )
|  upstream upstream=100 plus=10 | tr -d A-Z  >
"Output-intergenic.$1/$i.fasta"
        cd Output-intergenic.$1; clustalw -infile=$i.fasta -align > /dev/null
        cd ..
    echo $i
 done
```
(by Rick Stevens, co-developer of The Seed)

BioLingua-Lite version:

```
(FOR-EACH gene IN (GENES-OF Npun)
    AS orthologs = (ORTHOLOGS-OF gene)
    AS upstream-seqs = (SEQUENCES-UPSTREAM-OF orthologs LENGTH 100)
    COLLECT (ALIGNMENT-OF upstream-seqs))
```
(by Jeff Elhai, developer of BioLite)

BioLingua: A Computational Biology Workbench
                              Based on the KnowOS platform

- Integrates Genomic and Data Analysis Tools

- Integrates Organism-specific as well as General Knowledge

- Unifies Important Knowledge Bases

- Offers a Flexible "Open Programming" Methodology

- Provides Convenient Universal Access (fully web-enabled)

Free demo server: www.biolingua.org

Open Source software on SourceForge

Plan of the Presentation

* Overview of KnowOS services

* Example 1: BioLingua biological knowledge environment

* Example 2: CACHE collaborative knowledge analysis

* Issues and approaches

* Near and long term goals

ACH$_0$

		Code	Type	Weight	H: 1 Disgruntled Michael's employee or customer	H: 2 Foreign terrorist(s)	H: 3 Two black males in blue car with unknown, possibly criminal, motives	H: 4 Lone serial killer (almost certainly male, 80% prob white	H: 5 Domestic terrorist(s), white militiamen
	Inconsistency Score				-89	-55	-5	-47	-43
	Create Evidence								
E43	22 Oct Johnson shot on steps of Ride-on-Bus		Police report	HIGH	I	C	C	C	C
E42	Accent sounded Caribbean, Jamaican?		From sniper(s)	LOW	N	I	I	I	I
E41	Sniper calls from Ponderosa; five red stars note		From sniper(s)	HIGH	I	I	C	II	I
E40	Ala police report Armalite catalog dropped by suspect		Police report	HIGH	NA	NA	C	NA	NA
E39	Accent of phone caller Hispanic or Jamaican		Analysis	MEDIUM	N	I	I	I	I
E38	Sniper calls Pastor Sullivan (2 men, accent, Ala)		From sniper(s)	MEDIUM	I	I	C	C	I
E37	Credit card used in Tacoma WA linked with Alabama		Police report	MEDIUM	I	I	C	I	I
E36	Cinnaraison snack bag		Analysis	LOW	NA	NA	NA	NA	NA
E35	Handwriting matches Tarot card		Analysis	HIGH	I	I	C	C	I
E34	Ziploc bag with letter (use of "we") and demands		From sniper(s)	HIGH	II	II	C C	II	II
E33	.233 casing found consistent in most cases		Analysis	LOW	C	C	C	C	C
E32	19 Oct Hopper shot at Ashlawn Ponderosa		Police report	HIGH	I	C	C	C	C
E31	Sniper calls Baliles; provides Alabama info (2 suspects)		From sniper(s)	HIGH	II	I	C C	II	I
E30	Sniper calls dispatcher Lefkoff; young man using "we"		From sniper(s)	HIGH	II	I	C C	II	C
E29	14 Oct Franklin killed at Home Depot		Police report	HIGH	I	C	C	C	C

# Collaborative ACH

ACH_0

Incoming Intelligence

Incoming intelligence is distributed to the analysts in relevance-sorted order according to the hypotheses they are working with based upon an underlying knowledge model

Client/server architecture permits collaboration among analysts through "publication" of hypothesis and linking in as evidence

Linked matrices project a Bayesian influence network
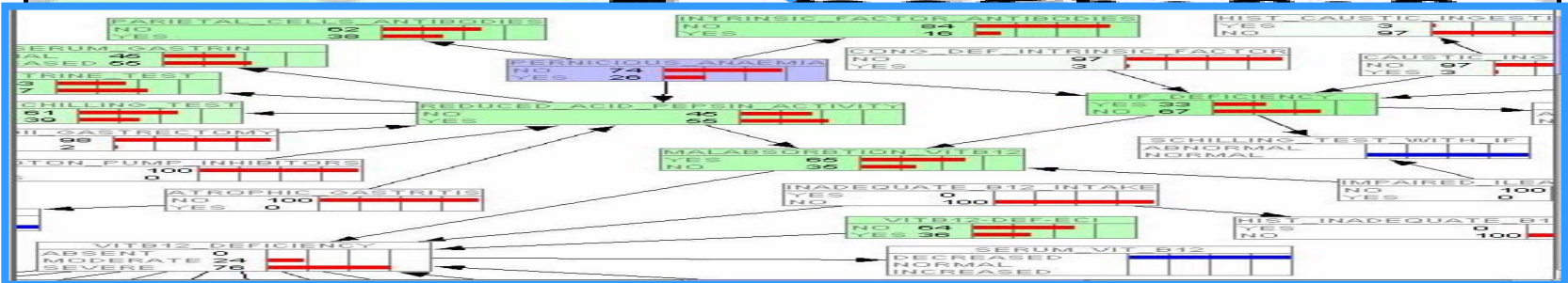
Top Matrix	Weight:	3301930362: Lone Wolf Neo Nazi Sniper	Diagnosticity:
Operations:		[▾]	
3301917788: TV profile: white, male, military background	MEDIUM ▾	C ▾	Non-diagnostic (1)!
Support:		0.5	

Select new evidence:

13.06: Sniper calls Baliles; [...]les Alabama info (E2 suspects)
12.[...] Sniper calls Pastor Su[...] (E2 men, accent, Ala)
9.5[...] [...]te Chevy Astro w lad[...]
9.0[...] [...] calls dispatcher Lefk[...]; y[...]

Enter a n[...] [...]othesis: [_____]

[Submit]

Both intelligence and hypotheses are linked to underlying knowledge layer

Incoming intelligence directed to the analysts working on relevant problems

Underlying knowledge layer in a frame system:

Frame #$SHOOT

Listener BioDocs FindFrames BioFiles BioCliki LispDocs

Min Less More Max Lispier **Frame->Listener**

Frames representing concepts

Slots

#^associates	#$WINDOW	#$STORE	#$FIRED	#$WAREHOUSE	#$SHOPPER
	#$MARTIN	#$GRAS	#$MOWING	#$LANDSCAPER	#$BUCHANAN
	#$BLOCK	#$HALF	#$MOBILE	#$WALEKAR	#$LOT
	#$WORLD	#$BENCH-LEISURE	#$RAMO	#$STATION	#$SHELL
	#$RIVERA	#$MARKSMAN			
	#$DC	#$AVE			
	...27 more...				
#^fName	SHOOT				
#^sentences	#$3301917333	#$3301917779	#$3301917253	#$3301917186	#$3301917629
	#$3301917525	#$3301917860	#$3301917476	#$3301917397	#$3301917413
	#$3301917577	#$3301917500	#$3301916950	#$3301917714	

Frames representing pieces of evidence

Parents [Hide]

Children [Hide]

Examine Evidence

Ranking Expression: `A terrorist was arrested in Madrid.` `Semantic distance ▾`

Matching basis: ☑Input string | ☐Matrix title | ☐Matrix hypotheses | ☐Matrix evidence | ☐Matrix explanations |

`Search / Refresh`

[Enter new external evidence]

Intelligence is ranked by "semantic similarity" (distance in knowledge space)

Evidence sorted by: **Semantic distance**
Ranking expression: **A terrorist was arrested in Madrid.**
Matching bases: **Matrix title = off | Matrix hypotheses = off | Matrix evidence = off | Matrix explanations = off |**

`Add` **1.00 : 20050412 Carl Louis arrested in Detroit [object]**

`Add` 0.47 : 20050412 Omar Bakri Qatada is arrested in London [object]

`Add` 0.47 : 20050412 C4 was stolen from manufacturer in Keswick, Virginia [object]

`Add` 0.46 : 20050412 Tawfiq al Adel and Saeed Khallad arrested in Netherlands, radioactive truck [object]

`Add` 0.42 : 20050412 Wallace Wilson arrested at Camp George West [object]

`Add` 0.39 : 20050412 Package bombs are found in Seattle [object]

`Add` 0.34 : 20050412 Muhammad Shamzai is arrested in Paris [object]

`Add` 0.28 : 20050412 Holland Queen arrives in Boston from Amsterdam April 29 [object]

`Add` 0.27 : 20050412 Omar Clark and Joseph Nizar are arrested at O'Hare [object]

`Add` 0.21 : 20050412 Car bomb goes off in Buenos Aires on April 19 [object]

`Add` 0.21 : 20050412 50 pounds of C4 are found in a storage space in Keswick, Virginia [object]

Interconnectivity of Individual Analyses:

-- Inference sharing and peer group critical analysis

-- Ability to track the chain of inference

Sharing of Hypotheses (or of evidence) by multiple matrices.

Linking of Hypotheses in one matrix as evidence in another matrix.
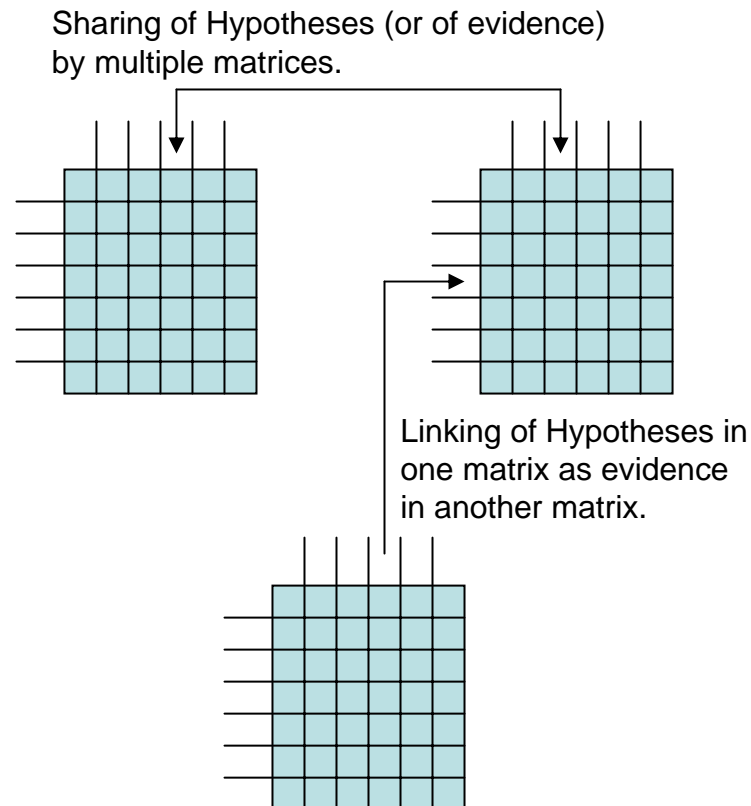
Analysts can "promote" hypotheses as if they were intelligence. The system guides these to other analysts working on related problems, those other analysts can link these into their ongoing analytical process.

Top Matrix	Weight:	3301930422: Foreign Terrorist	33 Wa Sn
Operations:			
3301930131: Terrorists sniping	MEDIUM	CC	
3301917159: Accent sounded Caribbean, Jamaican?	MEDIUM	C	
3301917788: TV profile: white, male, military background	MEDIUM	I	
Support:		0.5	-1

user: Heuer:

Top Matrix	Weight:	3301930556: Terrorists sniping	3301929911: Kids out joyriding with guns	Diagnosticity:
Operations:				
3301917397: Forensics shows all shot at long range with .223 bullets	MEDIUM	Delete / Promote	II	1 confirm, and 1 disconfirm (0)
3301917333: 2 Oct shot fired thru Michael's store window	MEDIUM	I	CC	1 confirm, and 1 disconfirm (0)
Support:		0.5	0.0	

CACHE

Incoming Intelligence

Client/server architecture permits collaboration among analysts through "publication" of hypothesis and linking in as evidence

Incoming intelligence is distributed to the analysts in relevance-sorted order according to the hypotheses they are working with based upon an underlying knowledge model

Analysis of Competing Hypotheses

Analysis of Competing Hypotheses

Analysis of Competing Hypotheses

Analysis of Competing Hypotheses
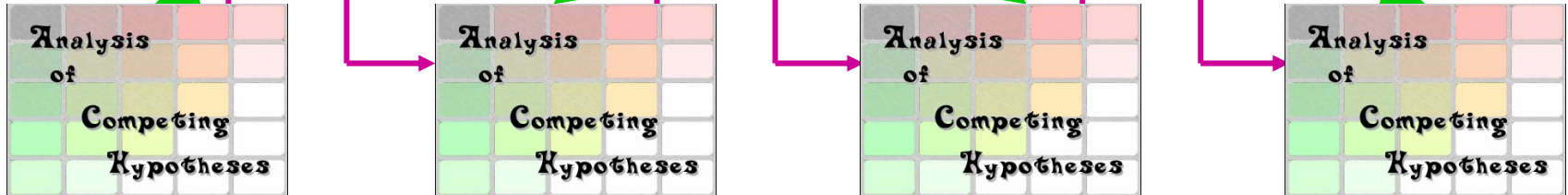
Linked matrices project a Bayesian influence network
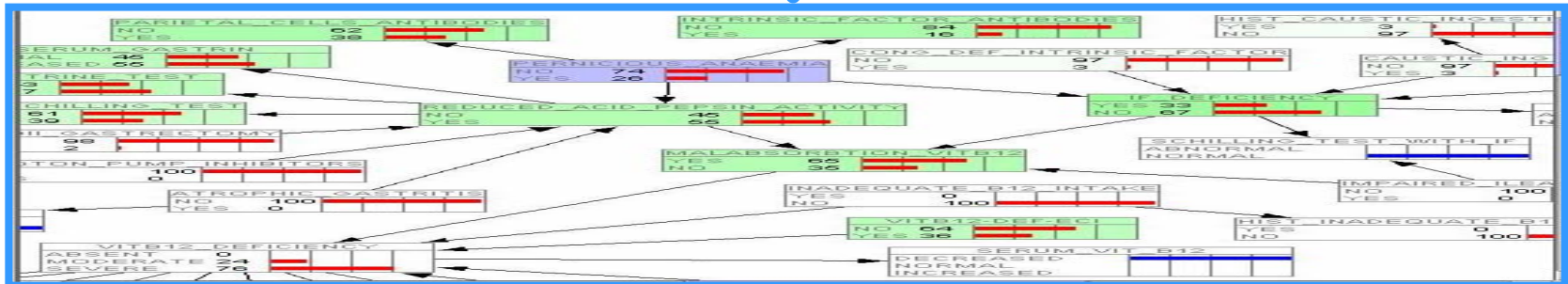
CACHE: A Collaborative Analysis Methodology

Based on the KnowOS platform

- Integrates Analyses across a Community of Analysts

- Enables Semantics-based Sharing of Evidence and Hypotheses

- Unifies Important Knowledge Bases

- Offers a Flexible "Open Programming" Methodology

- Provides Convenient Universal Access (fully web-enabled)

Sorry, no demo server yet ☹

Plan of the Presentation

* Overview of KnowOS services

* Example 1: BioLingua biological knowledge environment

* Example 2: CACHE collaborative knowledge analysis

* Issues and approaches

* Near and long term goals

Issues and Approaches

* Pure HTML limits interactivity (e.g., debugging)

* Various poor core algorithms have been discovered

* Users share the Lisp image (pros and cons)

    Name management issues (conficting exports)

    Thread management issues (GC can hang everyone)

    Incompatible with high security

Plan of the Presentation

* Overview of KnowOS services

* Example 1: BioLingua biological knowledge environment

* Example 2: CACHE collaborative knowledge analysis

* Issues and approaches

* Near and long term goals

Community Resources...

BioCL^{iki}

the Bio(Common)Lisp wiki

Home | Recent Changes | About CLiki | Text Formatting | Create New Page

sylvesterlab;init.lisp

;;;

This is a Lisp code file; Please don't remove or edit this line!

```
;;; Take the name of a constriants file (something like:
;;; "test.cons"), solves the constriants and displays a graphical
;;; average solution.  Normally it runs 100 iterations.
;;; To use it, do this: (consolve "test.cons")  The file (test.cons in
;;; this example) must appear in the CLiki world as username;test.cons,
;;; where username is your BioLingua login name.

(defun consolve (name &key (result-type :plot) (trials 100))
  (import-cliki-file name)
  (solve-constraints name :trials trials :result-type result-type))

;;;
```

This is a Lisp code file; Please don't remove or edit this line!

Edit page | View source | Last edit: Mon, 29 Dec 2003 00:35:51 GMT | [   ] search

CLiki pages can be edited by anyone at any time. Imagine a fearsomely comprehensive disclaimer of liability. Now fear, comprehensively

BioCL^{iki}

Home | Recent Ch

index

This is the main page of

BioLingua users should

If you have a page called
see mine, in case you w

Users' Private Pages:

JP's page
Mike's page
Jeff S's Page
James Evans' page
Jeremy Zucker's page

User Group Pages:

User Groups Control Pa
Sylvester Lab Main Pag
Bhaya Lab Main Page
Developer's Group Init

Developer's To Do L

Known Bugs

Toward More Real Persistence...

Everyone wants true persistence...

...until they actually get it!

Envisioned approach:

AllegroCache and the concept of
a "knowledge CVS"

Alternatives to having to type code...

```
(loop for pm4gene in (#^Genes ProcMed4)
    as all-orthologous =  (all-blast-orthologs pm4gene)
    as 6803ortholog = (intersect (#^Genes Syny6803) all-orthologous)
    when (and (not-any #'member-geneid
                   (#^Genes slotv Proc9313) all-orthologous))
        (any #'member-geneID 6803ortholog)
        (>= ma-ratio (ma-select 6803ortholog Hihara1) 2)))
    collect light-specific-genes 6803ortholog)
```
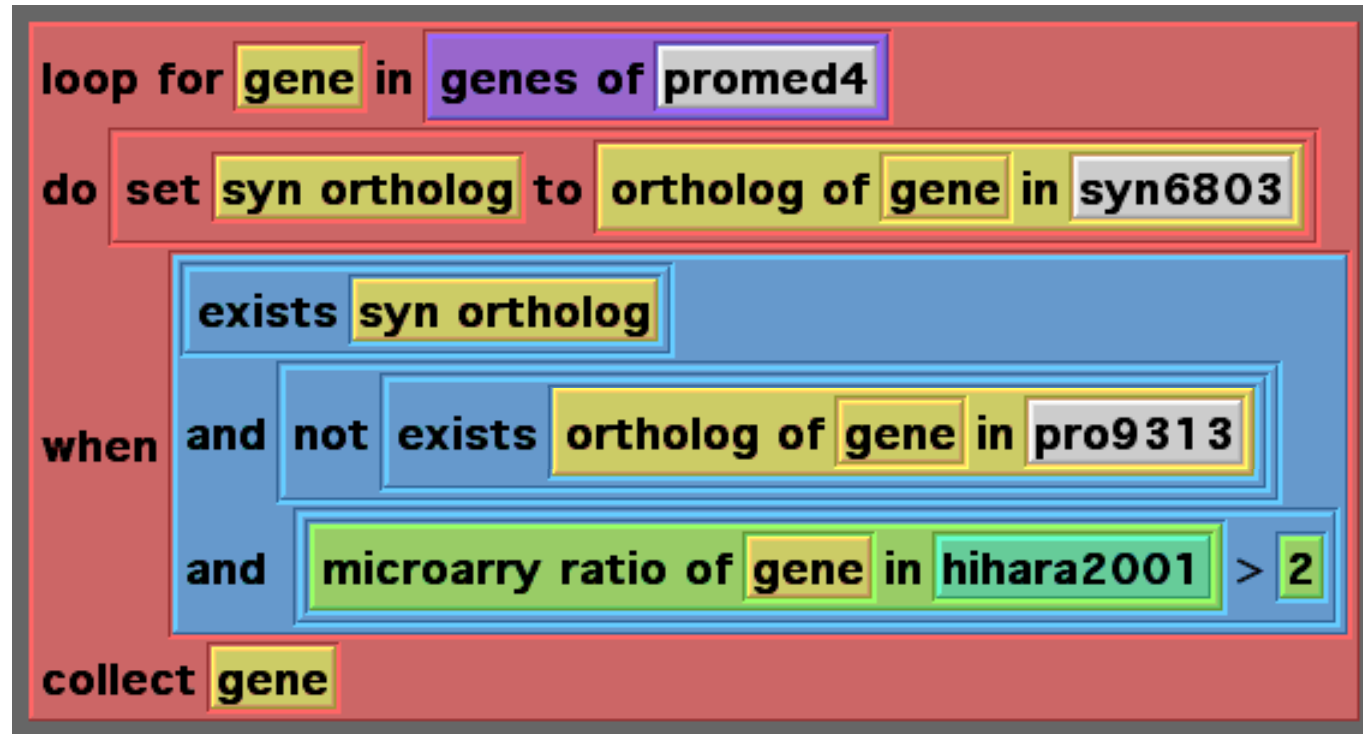
Advanced Reasoning Tools...

Richard Waldinger and Mark Stickel
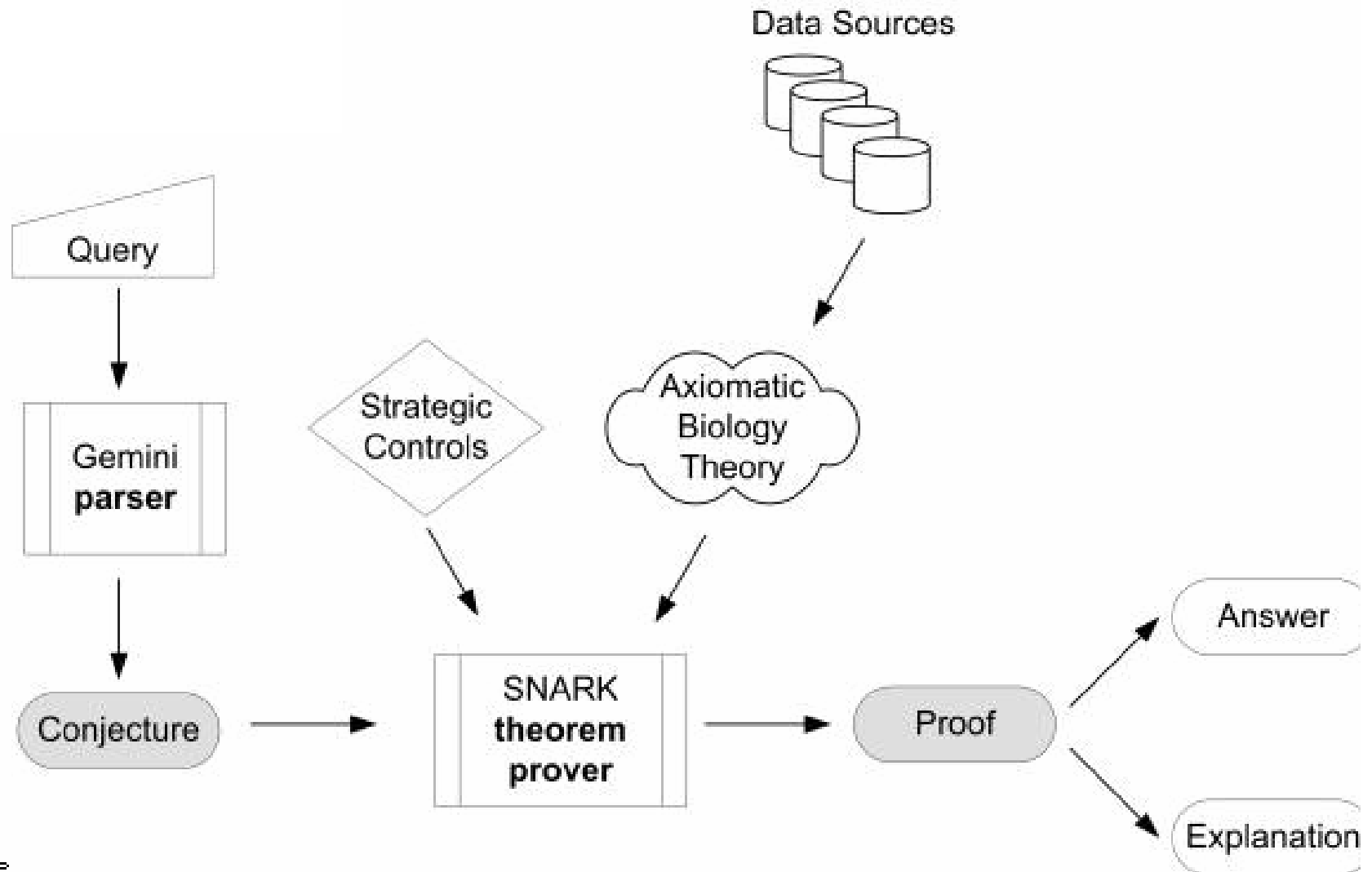
Advanced Reasoning Tools...

Richard Waldinger and Mark Stickel

(loop for pm4gene **in (#^Genes ProcMed4)**
    as all-orthologous = (all-blast-orthologs pm4gene**)**
    as 6803ortholog = (intersect (#^Genes Syny6803) all-orthologous)
    when (and (not-any #'member-geneid
                     (#^Genes slotv Proc9313) all-orthologous))
                (any #'member-geneID 6803ortholog)
                (>= ma-ratio (ma-select 6803ortholog Hihara1) 2)))
    collect light-specific-genes 6803ortholog)

Advanced Reasoning Tools...

Richard Waldinger and Mark Stickel

For each gene in ProMed4,
    Find all the gene's Blast orthologs,
    Find those from Syny6803,
    When there are *not* any Pro9313 genes in the Blast orthologs,
      and there are any the 6803 orthologs
      and the expression ratio for the 6803 orthologs
        in the Hihara microarray data is >= 2,
collect the 6803 orthologs in a list, called light-specific-genes.

Advanced Reasoning Tools...

Richard Waldinger and Mark Stickel

English Query: *List the genes that pertain to med4 and that have an ortholog in s6803 that has a hihara ratio greater than 2 and that do not have orthologs in mit9313.*

Advanced Reasoning Tools...

Richard Waldinger and Mark Stickel

```
(imp
  (exists
    (A)
    (and
      (and
        (holds gene A)
        (and
          (exists
            (B)
            (and
              (holds pertain B)
              (actor B A) true
              (none B med4)))
          (exists
            (C)
            (and
              (and
                (holds ortholog C)
                (in C s6803)
                (exists
                  (D)
                  (and
                    (and
                      (holds ratio D)
                      (
                        (lambda
                          (E)
                          (and
                            (and
                              (holds hihara E))))
                        (D))
                      (and
                        (exists
                          (F)
                          (exists
                            (G)
                            (exists
                              (H)
                              (and
                                (great F D G)
                                (exceeds_degree H G
                                  (number_to_x 2))))))
```

Advanced Reasoning Tools...

Richard Waldinger and Mark Stickel

```
(find-all
  '(and
    (gene-pertains-to-organism ?gene4 med4)
    (forall ((gene9313))
      (not (gene-has-ortholog-in-organism ?gene4 gene9313 mit9313)))
    (gene-has-ortholog-in-organism ?gene4 ?gene44 s6803)
    (= ?number (hihara-mean-regulation-ratio ?gene44))
    (> ?number 2))
  :answer '(ans ?gene4 ?gene44 ?number))
```

Advanced Reasoning Tools...

Richard Waldinger and Mark Stickel

```
(Refutation
(Row hihara-problem
   (or
    (not
     (gene-pertains-to-organism ?gene
      |hashdollar-prochlorococcus_marinus_med4|))
    (not
     (gene-has-ortholog-in-organism ?gene ?gene1
      |hashdollar-synechocystis_pcc6803|))
    (not (= ?number (hihara-mean-regulation-ratio ?gene1)))
    (not (> ?number 2))
    (gene-has-ortholog-in-organism ?gene
     (snark-user::gene-skolemkibs1 ?gene)
     |hashdollar-prochlorococcus_marinus_mit9313|))
  negated_conjecture
  Answer (answer-- (ans ?gene ?gene1 ?number)))
(Row 230
   (or
    (not
     (gene-has-ortholog-in-organism |hashdollar-PMED4.PMM0226| ?gene
      |hashdollar-synechocystis_pcc6803|))
```

Advanced Reasoning Tools...

Richard Waldinger and Mark Stickel

(ANSWER-- (ANS #$PMED4.PMM0817 #$S6803.ssr2595 2.2025))
(ANSWER-- (ANS #$PMED4.PMM0226 #$S6803.slr1604 2.17)))

KnowOS Applications

Real running servers:

Multi-Cyano BioLingua (CIW / VCU / others) [+ teaching]

Parasite BioLingua (VCU)

Arabidopsis BioLingua (CIW / NTT / U.Chicago)

CACHE (PARC / NIMD)

Proposed:

Human BioLingua (Stanford Genome Tech. Ctr.)

BioCACHE for Multi-Cyano Annotation (CIW / MIT)

Space Sciences Discovery Platform (NASA)

Community Hypothesis Browser (Penn State)

KnowOS Core Tech

JP Massar
Mike Travers
Mark Slupesky

Additional Code

Edi Wietz
Dan Barlow

Sever Support

Bob Haxo
Daniela Puiu
Mike Chapman

BioLingua

Jeff Elhai
Andrew Pohorille
Stephen Bay
Pat Langley

Students

Monica Jain
Ashvin Kumar
Sumudu Watagala
Marc Santoro

CACHE

Doritt Billman
Pete Pirolli
Stu Card

Sources of Support: NASA, NSF, CIW, NTT, VCU, Franz, LispWorks, Stanford