

WebScripter

Tools for Building a Semantic Web from the Ground Up

May 26, 2004

Dr. Robert MacGregor
USC/Information Sciences Institute

macgregor@isi.edu

End-user Knowledge Engineering

- Semantic experts are scarce, knowledge acquisition is expensive and inexact, semantic expert unavailable after initial session(s).
 - End-users are the real experts – need to insure that their expertise is fully-exploited
- ➔ Need tools that help end-users to evolve ontologies while working with data

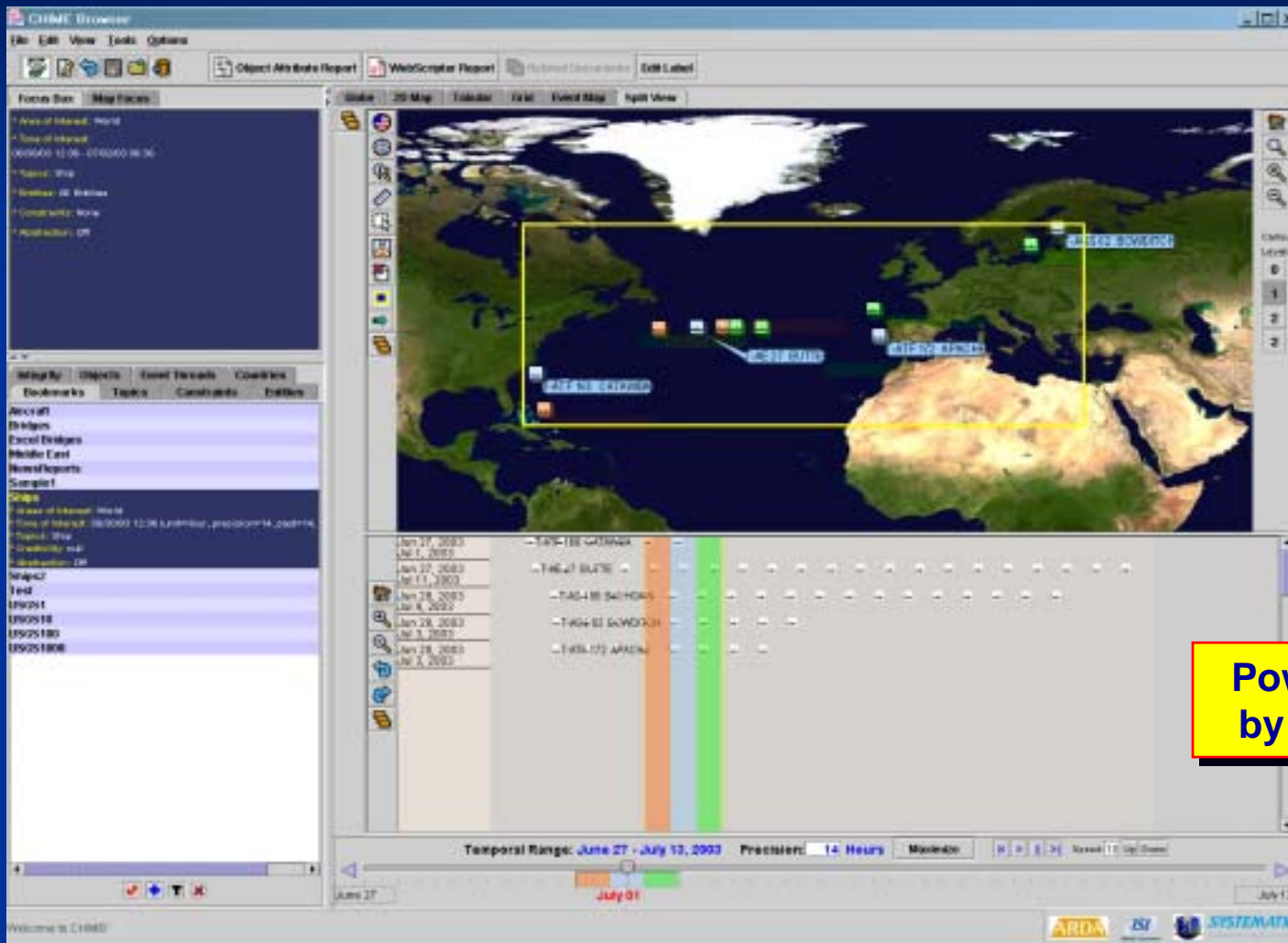
Grassroots Approach

- New breed of tools that can function without centralized ontologies and alignments
- Carrots instead of sticks – reward users for:
 - creating vocabulary
 - adopting each others' vocabularies
 - creating alignment catalogs
 - adopting each others' alignments
 - annotations

Key End-user Activities

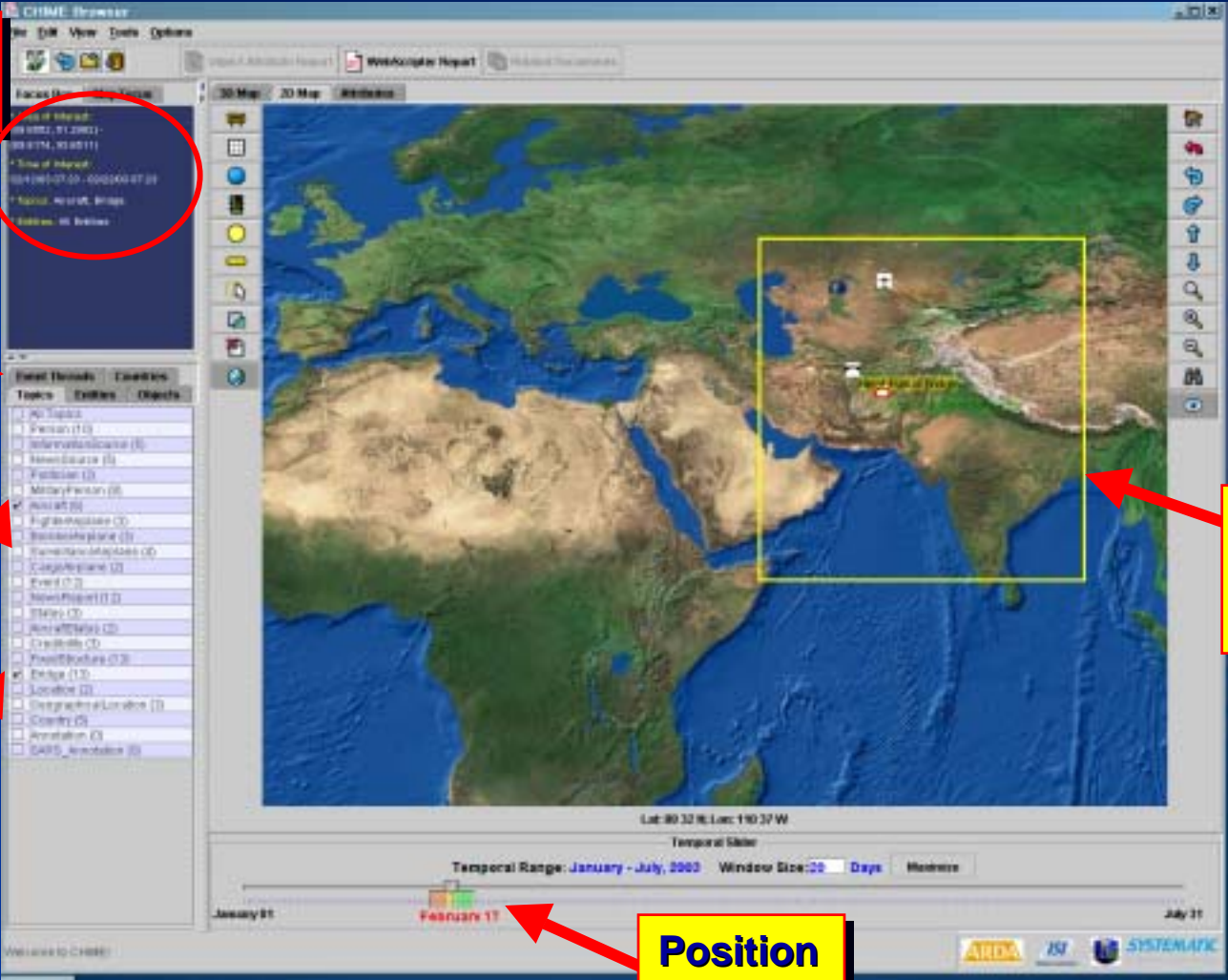
- Authoring
 - eNotes to self, others, annotations
 - Markup (authoring + advertisement)
 - of maps, text documents, Web pages
 - Visualization
 - of annotated maps, reports, etc.
 - Retrieving
 - eDocuments, eNotes, communications, annotations
 - Organizing
- ➔ Ontologies should be a side-effect of each of these activities

CHIME: GeoSpatial Visualization



Powered by OWL

Grassroots Spatio-temporal Queries



The screenshot shows the CHIME Browser interface. On the left, there is a sidebar with a "Focus Box" settings section (circled in red), a list of topics/entities, and a "Position Time Slider" at the bottom. The main area is a 3D map of the world with a yellow box highlighting a region in the Middle East. A red arrow points from the "Position Time Slider" to the date "February 11" on the timeline.

"Focus Box" settings

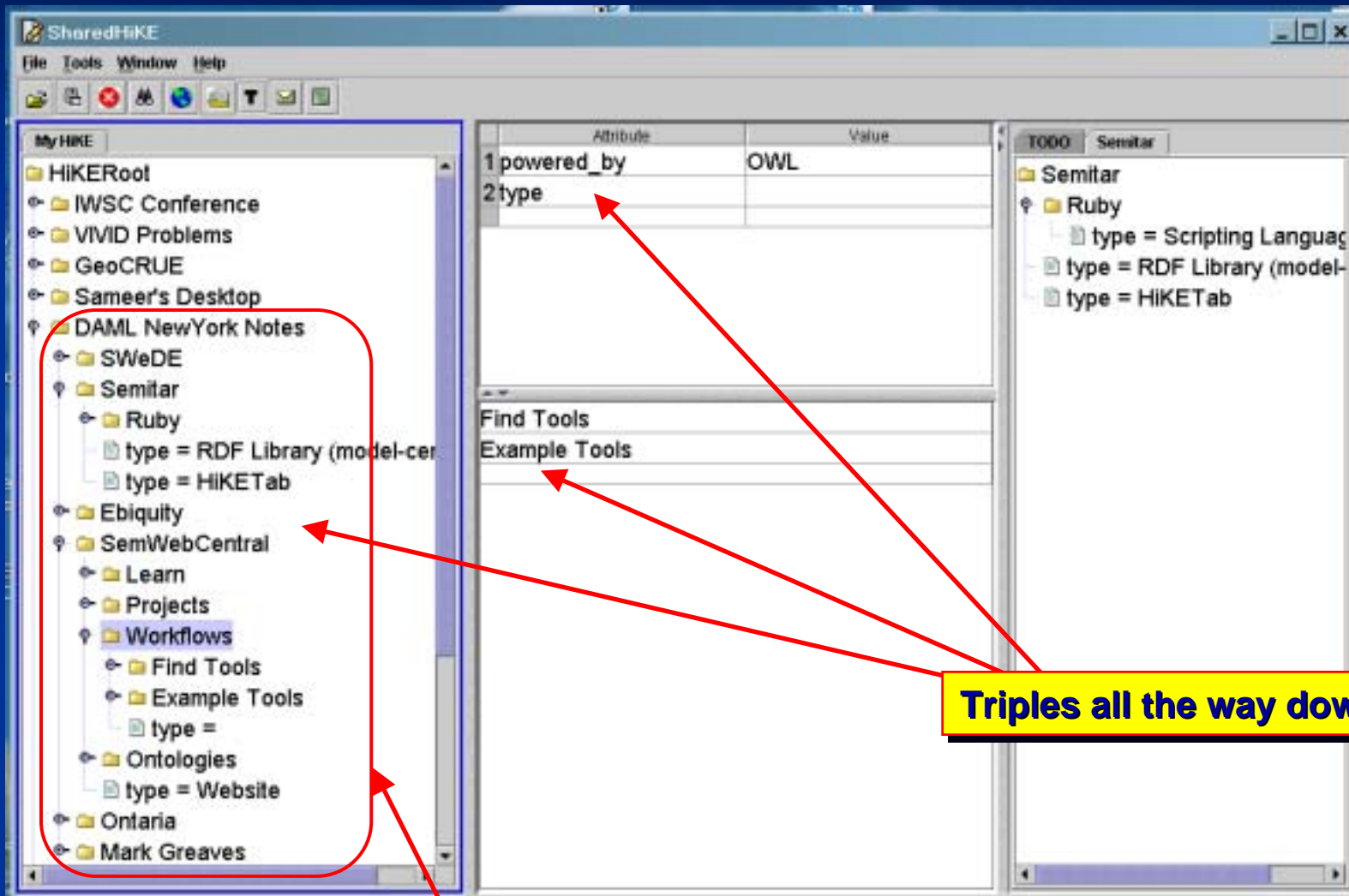
Select Integrity Level

Choose Topics (or Entities, Event Threads, ...)

Identify Spatial Region

Position Time Slider

HiKE Semantic Desktop



The screenshot shows the SharedHiKE application window. On the left is a file tree under 'MyHiKE'. The center pane displays a table of triples:

Attribute	Value
1 powered_by	OWL
2 type	

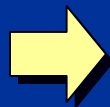
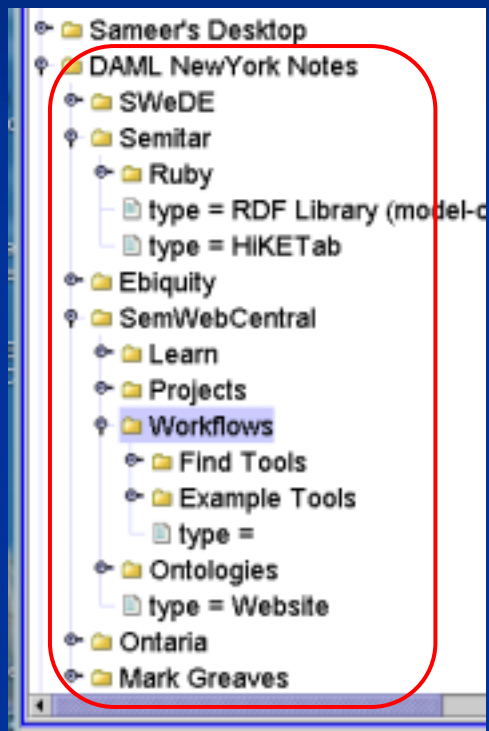
Below the table are sections for 'Find Tools' and 'Example Tools'. On the right, a 'TOOO Semitar' pane shows a tree structure for 'Semitar' with sub-items like 'Ruby' and various 'type' attributes.

Triples all the way down

Notes taken yesterday

Report Generation

- How do I print out a section of a HiKE network?



- WebScripter can't handle nested/hierarchic data
 - ➔ inspired development of **VIVID**

VIVID: Visual Variable-Depth Information Displays

- Semantic Web data is
 - frequently non-uniform
 - often deeply nested
 - not easily viewed or browsed using conventional tools

- "Navigating a semantic web is like looking at data through a soda straw." – Joe Rockmore

How to Display SWeb data

■ Problems

- Hierarchic displays waste most of the screen real estate
- Conventional tabular displays space-efficient, but handle only “uniform” data

Sigmod Record Example of Low-Density Display

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<!--
XML-Page generated by PERSEUS:
1234 Academic Group and University "Fox Inc", Rome, ITALY
-->
<!DOCTYPE SigmodRecord (View Source for full doctype...)>
<SigmodRecords>
  <issues>
    <issue>
      <volume>15</volume>
      <number>2</number>
    </issue>
  </issues>
  <articles>
    <article>
      <title articleCode="152033">A DBMS prototype to support extended NF2
relations: an integrated view on flat tables and hierarchies</title>
      <authors>
        <author AuthorPosition="03">F Andersen</author>
        <author AuthorPosition="04">H Blaske</author>
        <author AuthorPosition="02">K Kuespert</author>
        <author AuthorPosition="01">P Dadam</author>
        <author AuthorPosition="05">R Erbe</author>
      </authors>
    </article>
    <article>
      <title articleCode="152014">A message passing framework for logical query
evaluation</title>
      <authors>
        <author AuthorPosition="01">Allen Van Gelder</author>
      </authors>
    </article>
    <article>
      <title articleCode="152026">A rule-based object/task modelling approach</title>
      <authors>
        <author AuthorPosition="01">Qiming Chen</author>
      </authors>
    </article>
    <article>
      <title articleCode="152004">A snapshot differential refresh algorithm</title>
      <authors>
        <author AuthorPosition="01">Bruce Lindsay</author>
        <author AuthorPosition="03">C Mohan</author>
        <author AuthorPosition="04">Hamid Pirahesh</author>
        <author AuthorPosition="02">Laura Haas</author>
      </authors>
    </article>
  </articles>

```

VIVID Display (same dataset)

Home Class Page Show Format Tree Style: Original Levels: 1 Highlight:

title data						
		authors [Authors]			[Articles]	
		author data [Author]			issue [Issue]	
title	articleCode	authorPosition	author	value	count	isRootNode
Data warehousing and OLAP for decision support	262686	01 02	Surajit Chaudhuri Umeshwar Dayal	26	2	true
Foundations of Statistical Natural Language Processing - by C. Manning et al	313194	00	Reviewer: G. Weikum	31	3	true
Communication Efficient Distributed Mining of Association Rules	302071	00	Assaf Schuster	30	2	true
Performance Evaluation of a New Distributed Deadlock Detection Algorithm	233211	00 02 01	Chim-fu Yeung Kam-yiu Lam Sheung-lun Hung	23	3	true
A New Join Algorithm	234223	01 00	Arnold Charles Heltzer Dong Keun Shin	23	4	true
Online Association Rule Mining	282030	01	Christian Hidber	28	2	true
Constructing the Next 100 Database Management Systems	231196	00 01	Andreas Geppert Klaus R. Dittrich	23	1	true
Left-deep vs. bushy trees: an analysis of strategy spaces and its implications for query optimization	202235	01 02	Yannis E. Ioannidis Younkyung Cha Kang	20	2	true
Template-based wrappers in the TSIMMIS system	262694	04 05 03 01 06 02	Ramana Yerneni Marcus Breunig Svetlozar Nestorov Joachim Hammer Vasilis Vassalos Héctor García-Molina	26	2	true
Lore: A Database Management System for Semistructured Data	263363	01 02 04 00 03	Serge Abiteboul Roy Goldman Jennifer Widom Jason McHugh Dallan Quass	26	3	true
The Impact of Database Research on Industrial Products (Panel Summary)	233214	02 04 03 04	Daniel H. Fishman Michael Stonebraker David B. Lomet Jeffrey D. Bratton	23	3	true

Hierarchic Header

Tabular Data

How to Manage Header Information (CIA Fact Book dataset)

Home Class Page Show Format Tree Style: Small Type Levels: 4 Highlight:

yield (yield)												road (road)	
introduction		international_dispute				people (people)						people (people)	
background	country	international_dispute	total_dispute	death_rate	inf_population_rate	inf_mortality	health (health)				total_fertility_rate	birth_rate	
							years-05_and_over	total_population	at_birth	years-15_to_64	under_15_years		
Pitcairn... 50 today	Pitcairn Islands	international_noc		NA death...opstlbr	NA migra...opstlbr	NA death...rate						NA total_fertility_rate	NA birth...opstlbr
Slovene... in 1990	Slovenia	international_noc	dispute...with Europe	6.21 dec...01 est	MEDL	6.39 dec...01 est	0.71 male/female	0.93 mal...001 est	1.05 male/female	0.93 male/female	1.07 male/female	1.8 chl...001 est	12.64 bi...001 est
After is... a crowd	Belarus	international_noc	dispute...with Europe	13.07 dec...01 est	1.00 mig...001 est	14.30 dec...01 est	0.48 male/female	0.88 mal...001 est	1.05 male/female	0.94 male/female	1.04 male/female		
The is... in 1825	Clipperton Island	international_noc											
New Zeal... in 1907	Samoa	international_noc		8.20 dec...01 est	-11.82 m...01 est	31.78 dec...01 est	0.27 male/female	1.08 mal...001 est	1.08 male/female	1.71 male/female	1.03 male/female		
Since 18... see only	Sierra Leone	internat... Liberia		18.18 dec...01 est	15.23 m...returning	148.32 k...001 est	0.86 male/female	0.94 mal...001 est	1.03 male/female	0.91 male/female	0.95 male/female	8.01 chl...001 est	48.11 bi...001 est
The boat... in 1940	Falkland Islands	internat... by the US											
in 1974... an year	Tuvalu	international_noc		7.55 dec...01 est	0 migran...001 est	22.85 dec...01 est	0.72 male/female	0.94 mal...001 est	1.04 male/female	0.92 male/female	1.04 male/female	3.08 chl...001 est	21.58 bi...001 est
The full... embarks	Brazil	internat... the island	dispute...with Paraguay	3.30 dec...01 est	4.07 mig...001 est	14.4 dec...01 est	0.92 male/female	1.1 male...001 est	1.05 male/female	1.14 male/female	1.04 male/female	2.44 chl...001 est	20.45 bi...001 est
This arc... of Torres	Faeroe Islands	internat... of Western											
The town... in 1980	Central African Republic	international_noc		18.63 dec...01 est	0 migran...001 est	106.25 d...001 est	0.78 male/female	0.88 mal...001 est	1.03 male/female	0.98 male/female	1.21 male/female	4.88 chl...001 est	37.05 bi...001 est
Sakombe... in 1881	Marshall Islands	international_noc	dispute...with the US	14.61 dec...01 est	0 migran...001 est	89.88 dec...01 est	0.75 male/female	0.97 mal...001 est	1.03 male/female	0.94 male/female	1.02 male/female	8.23 chl...001 est	48.23 bi...001 est
Uninhab... status	Saint Helena	international_noc		0.33 dec...01 est	0 migran...001 est	22.30 dec...01 est	0.78 male/female	1.04 mal...001 est	1.04 male/female	1.08 male/female	1.02 male/female	1.53 chl...001 est	13.48 bi...001 est
Occupied... status	Sri Lanka	international_noc		6.42 dec...01 est	-1.42 m...001 est	16.09 dec...01 est	0.82 male/female	0.97 mal...001 est	1.05 male/female	0.95 male/female	1.05 male/female	1.95 chl...001 est	18.58 bi...001 est

Full header is ~18 feet wide

Header Formatting Tool

Home Class Page Hide Format Tree Style: Small Type Levels: 4 Highlight:

Go! Reset Select Column Width

VIVID Header Tree

- record [Record]
- introduction [Introduction]
- background
- country
- transnational_disputes [Transnational_disputes]
 - international_disputes
 - illicit_drugs
- people [People]
 - death_rate
 - net_migration_rate
 - infant_mortality_rate
 - sex_ratio [Sex_ratio]
 - years-65_and_over
 - total_population
 - at_birth
 - years-15_to_64
 - under_15_years
 - total_fertility_rate
 - birth_rate
 - religions
 - ethnic_groups
 - population
 - languages
 - population_growth_rate
 - hIV_AIDS-deaths
 - hIV_AIDS-adult_prevalence_rate
- literacy [Literacy]
 - female
 - definition
 - total_population
 - male

	years-65_and_over	total_population	at_birth	years-15_to_64	under_15_years	total_fertility_rate	birth_rate
163 south_... ethi						6.82 (ml_001 est.)	38 (ml_001 est.)
530 eth_001 est.)	0.71 male@female	0.22 (ml_001 est.)	1.05 male @female	0.02 male @female	1.07 male@female	1.8 (ml_001 est.)	12.04 (l_001 est.)
1420 (ml_001 est.)	0.40 male@female	0.80 (ml_001 est.)	1.05 male @female	0.04 male @female	1.04 male@female	1.20 (ml_001 est.)	4.87 (l_001 est.)
2175 (ml_001 est.)	0.87 male@female	1.30 (ml_001 est.)	1.05 male @female	1.71 (ml_001 est.)	1.03 male@female	3.4 (ml_001 est.)	15.59 (l_001 est.)
14652 (l_001 est.)	0.20 male@female	0.04 (ml_001 est.)	1.02 male @female	0.01 male @female	0.95 male@female	6.01 (ml_001 est.)	45.11 (l_001 est.)
2205 (ml_001 est.)	0.72 male@female	0.94 (ml_001 est.)	1.04 male @female	0.02 male @female	1.04 male@female	3.09 (ml_001 est.)	21.50 (l_001 est.)
144 (ml_001 est.)	0.92 male@female	1.1 (ml_001 est.)	1.05 male @female	1.14 male @female	1.04 male@female	2.44 (ml_001 est.)	20.45 (l_001 est.)
10825 (l_001 est.)	0.71 male@female	0.94 (ml_001 est.)	1.02 male @female	0.02 male @female	1.01 male@female	4.80 (ml_001 est.)	37.05 (l_001 est.)
6605 (ml_001 est.)	0.75 male@female	0.97 (ml_001 est.)	1.02 male @female	0.04 male @female	1.02 male@female	6.23 (ml_001 est.)	44.22 (l_001 est.)
2236 (ml_001 est.)	0.75 male@female	1.04 (ml_001 est.)	1.04 male @female	1.05 male @female	1.02 male@female	1.93 (ml_001 est.)	13.40 (l_001 est.)
1608 (ml_001 est.)	0.82 male@female	0.87 (ml_001 est.)	1.05 male @female	0.05 male @female	1.05 male@female	1.90 (ml_001 est.)	19.80 (l_001 est.)

Header Formatter is ~2.5 feet tall

View dataset from 'Author' perspective

Home Class Page Show Format Tree Style: Original Levels: 8 Highlight:

author data [Author]		title data [Title]		articles [Articles]		
author	authorPosition	articleCode	title	number	volume	isRootNode
Omran Bukhres	04	232424	Ensuring relaxed atomicity for flexible transactions in multidatabase systems	2	23	true
Hasan Davulcu	01	282077	A Layered Architecture for Querying Dynamic Web Content	2	28	true
Eric N. Hanson	01	212277	Rule condition testing and action execution in Ariel	2	21	true
Cevdet Dengi	07	252635	METU interoperable database system	2	25	true
Asuman Dogac	01	252635	METU interoperable database system	2	25	true
Ouri Wolfson	04	313288	Research Activities in Database Management and Information Retrieval at the University of Illinois at Chicago	3	31	true
A. Singh	03	262676	Efficient view maintenance at data warehouses	2	26	true
M. Hauswirth	02	314049	A Framework for Semantic Gossiping	4	31	true
Peter Z. Kunszt	01	292038	Designing and Mining Multi-Terabyte Astronomy Archives: The Sloan Digital Sky Survey	2	29	true
Alex Pang	04	282031	Optimization of Constrained Frequent Set Queries with 2-variable Constraints	2	28	true
Rafael Alonso	00	231206	Trade Press News	1	23	true
M. Haynie	01	172112	A DBMS for large design automation databases	2	17	true
Chim-fu Yeung	00	233211	Performance Evaluation of a New Distributed Deadlock Detection Algorithm	3	23	true
T Lougenia	01	153032	Graph user interfaces and database management systems	2	15	true

Pivot on 'Title'

Pivoting (cont.)

View dataset from 'Title' perspective

Home Class Page Show Format Tree Style: Original Levels: 8 Highlight:

title data [Title]		article [Article]				
articleCode	title	authors [Authors]	articles [Articles]			
		author dete [Author]	issue [Issue]	issues [Issues]	[SignedRecord]	isRootNode
		author	authorPosition	number	volume	
282139	The WASA2 Object-Oriented Workflow Management System	Mathias Weske Gottfried Vossen	02 01	2	28	true
192193	ACTA: a framework for specifying and reasoning about transaction structure and behavior	Panayiotis K. Chrysanthis Krithi Ramamritham	01 02	2	19	true
294022	Cache Invalidation Scheme for Mobile Computing Systems with Real-time Data	E. Chan K.-Y. Lam H. Leung J. Yuen	02 03 04 01	4	29	true
292061	Self-Organizing Data Sharing Communities with SAGRES	Ewa Jaslikowska Zachary G. Ives Jing Su Alon Y. Levy Qiong Chen Shiori Betzler Wai Tak Theodora Yeung Rachel Pottinger Stefan Saroiu	08 00 09 01 07 06 10 03 04 02	2	29	true

Users easily generate complex queries

article [Article]			
authors [Authors]	title data [Title]	articles [Articles]	
author data [Author]		issue [Issue]	
author	title	number	volume
	Bibliography	4	22
Y. A. Aslandogan K. Vadaparty G. Ozsoyoglu	Towards a unified visual database access	2	22
Bhavani M. Thuraisingham Hai-Ping Ko	Concurrency Control in Trusted Database Management Systems: A Survey	4	22
Dhamir N. Mannai Khaled Bugrara	Enhancing inter-operability and data sharing in medical information systems	2	22
Tony Schaller	The INTERsect concept for multidatabase system integration in the pharmaceutical industry	2	22
Lois M. J. Delcambre			

Simple query with one column constraint

RDQL-like formulation of the query

```

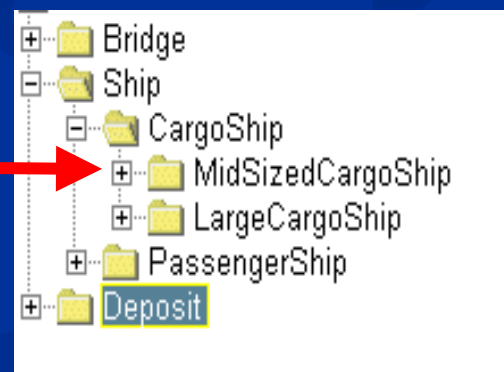
SELECT ?author, ?title, ?number, ?volume
WHERE
  (?article, rdf:type, cf:Article) AND
  (?article, cf:title_data, ?td) AND
  (?td, cf:title, ?title) AND
  (?article, cf:authors, ?authors) AND
  (?authors, cf:author_data, ?ad) AND
  (?ad, cf:author, ?author) AND
  (?articles, cf:article, ?article) AND
  (?issue, cf:articles, ?articles) AND
  (?issue, cf:number, ?number) AND
  (?issue, cf:volume, 22)
USING cf For <http://www.isi.edu/webscripiter/CIAFactBook#>
  
```

Using Queries to Build Vocabulary

ship [Ship]						ship [Ship]			
date	name	oparea	procode	time	length	[Fleet]	inventory [Inventory]	location [Location]	
						isRootNode	item	latitude	longitude
-	-	-	-	-	GREATER_OR_EQUAL 70 LESS_OR_EQUAL 150	-	-	-	-

- Retrieve cargo ships with length between 70 and 150
- Name query “MidSizedCargoShip”
- Automatically positioned in class hierarchy

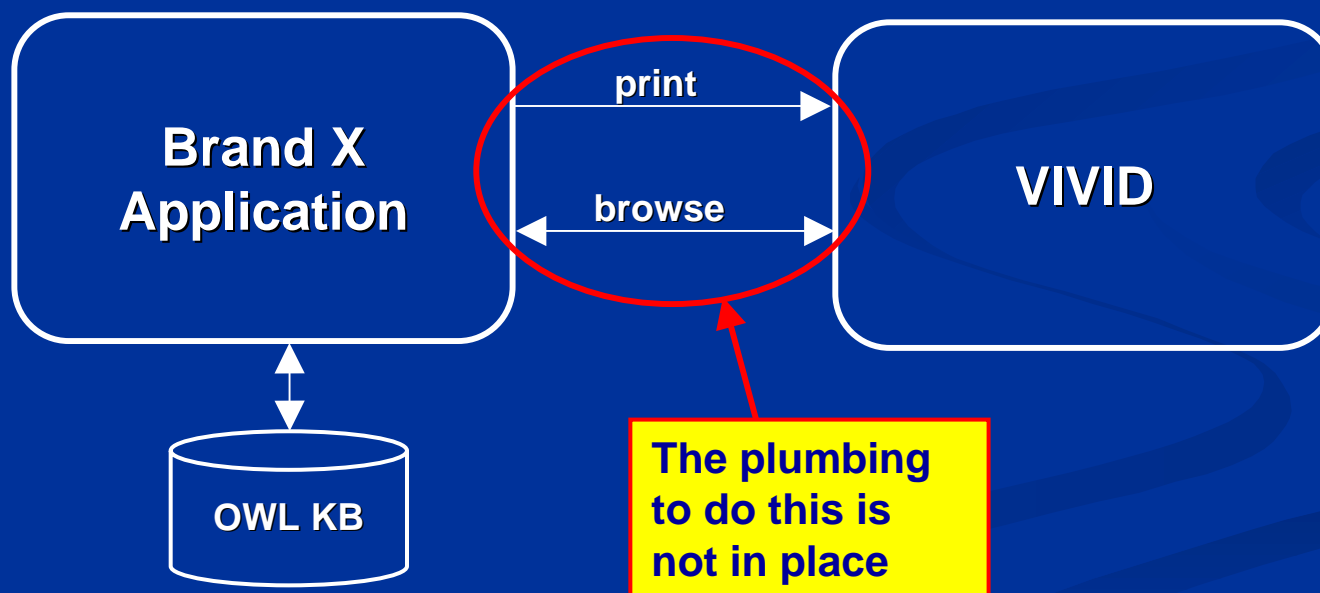
MidSizedCargoShip =
Ship and
length > 70 and
length < 150



How can I do this in OWL or SWRL?

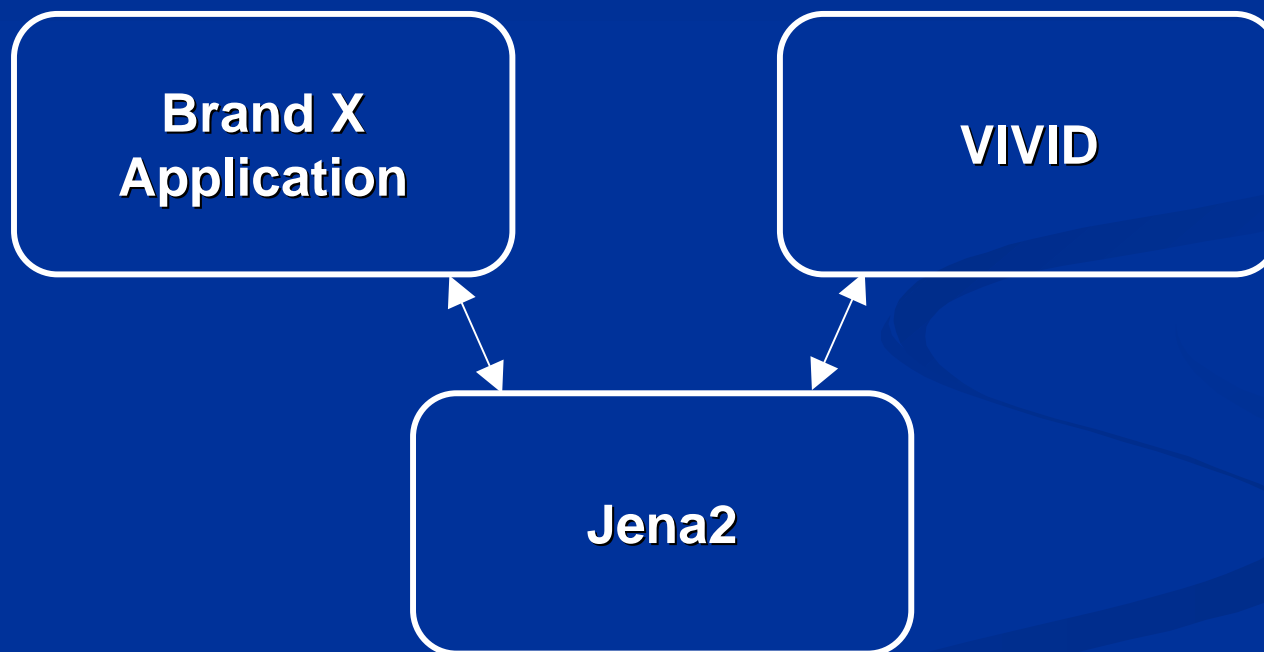
Transition to SemWebCentral

- VIVID would like to be tightly-coupled to an app
 - needs high bandwidth to browse knowledge base
- Problem: No standard API for interfacing RDF/OWL tools

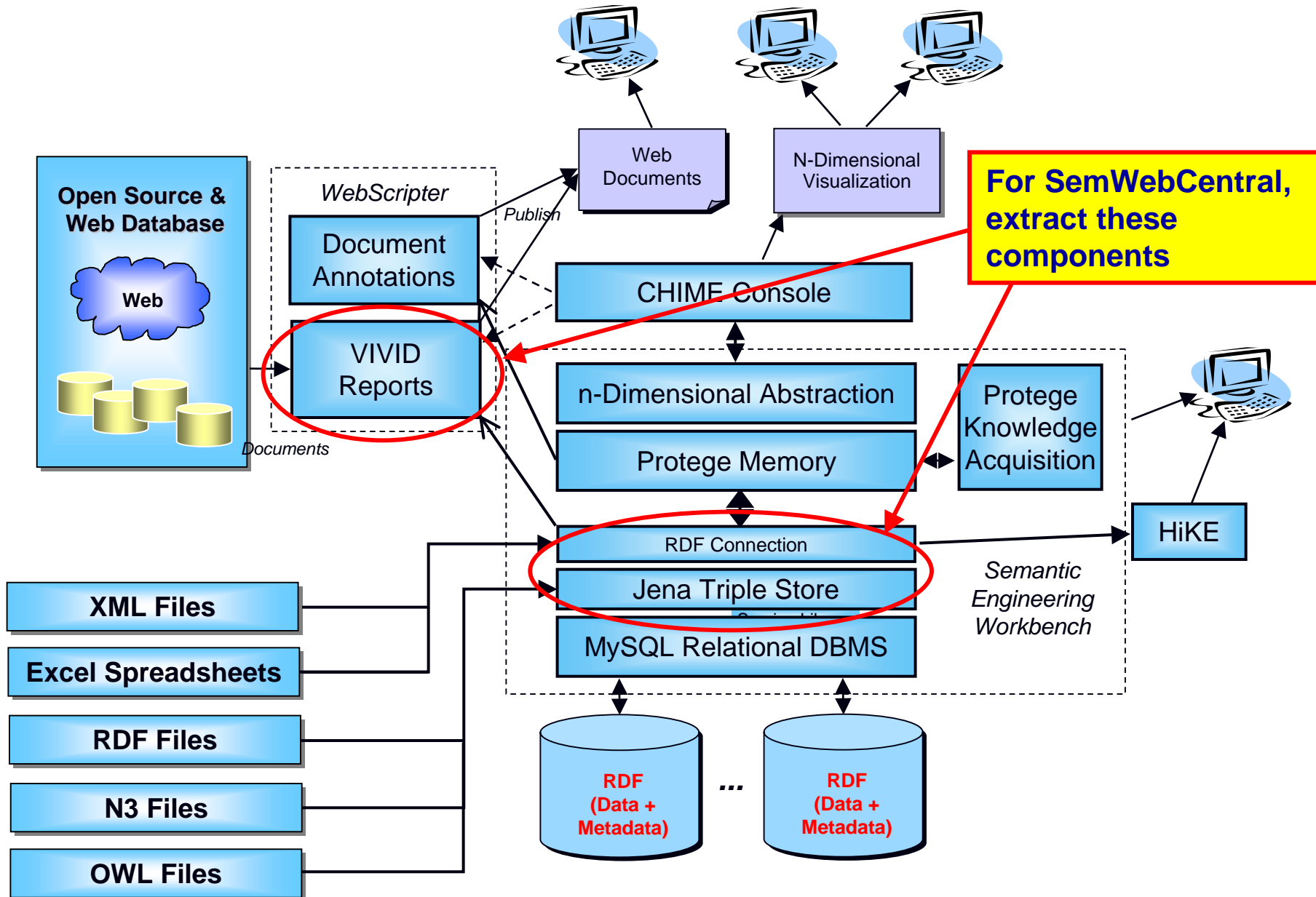


Transition to SemWebCentral (cont.)

- For now, use Jena2 as bridge between app and VIVID



Semantic Engineering Workbench



Tasks for Remainder of '04

- Round-out first VIVID release
 - Printing/Report Generation
 - Scale-up
- First release to SemWebCentral
- Additional features
 - Updating
 - Bookmarking/User Profiles
- Second release to SemWebCentral

END OF TALK ONE

Ontoagents

(Stefan Decker, Siegfried Handschuh,
Steffen Staab, Rudi Studer)

TRIPLE: <http://triple.semanticweb.org> (ISI)

■ Goals/Status:

- ✓ Inference Engine API
- ✓ Grid Reasoning Service (Grid Matchmaker)
- ✓ Data integration in NSF Argos project (integrating transport data)
- ✓ Many users and contributors: WU Vienna, University of Hannover, Fraunhofer Berlin, TU Berlin, ISI, DFKI, DERI,...

■ Work in Progress

- ↪ Re-Integrate work from outside contributors
- ↪ “Decker Problem”: Specifying Input/Output of Web Services – using this for validation
- ↪ Implementing OWL semantics (AMAP)

■ SemWebCentral

- ↪ Currently on SourceForge: Date: After DAML Meeting , end of June 2004
-

Ontoagents

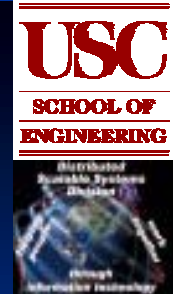
(Stefan Decker, Siegfried Handschuh,
Steffen Staab, Rudi Studer)

Annotation/Crawling (AIFB, Karlsruhe):

- Goals/Status:
 - ✓ OWL compatible Annotation Tool (CREAM/OntoMat)
 - ✓ OWL-Crawler (standalone and OntoMat-plugin)
 - ✓ Annotation/Meta-Ontology
 - ✓ Extended Information-Extraction (The Self-Annotating Web)
 - Work in Progress
 - ↵ Annotation & Authoring Web Pages
 - ↵ Ontology Engineering & Annotation
 - ↵ Date: Oct 2004
 - SemWebCentral
 - ↵ OWL-Crawler (project approved for release)
 - ↵ Annotation-Tool
 - ↵ Date: After DAML Meeting , June 2004
-

■ BACKUPS

For more information

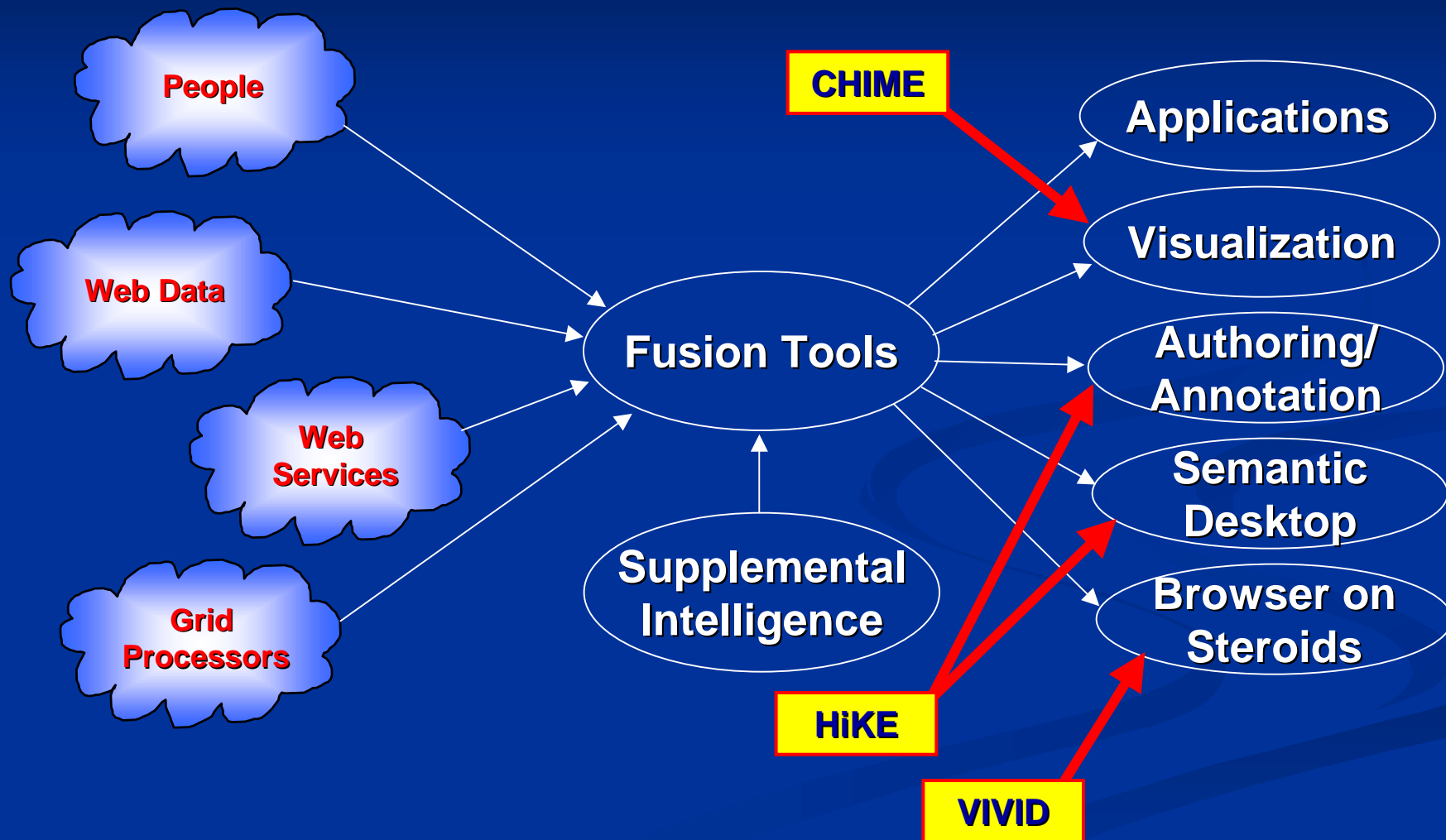


- Dr Robert MacGregor
 - University of Southern California/Information Sciences Institute
 - 310-448-8423
 - macgregor@isi.edu

VIVID Challenges

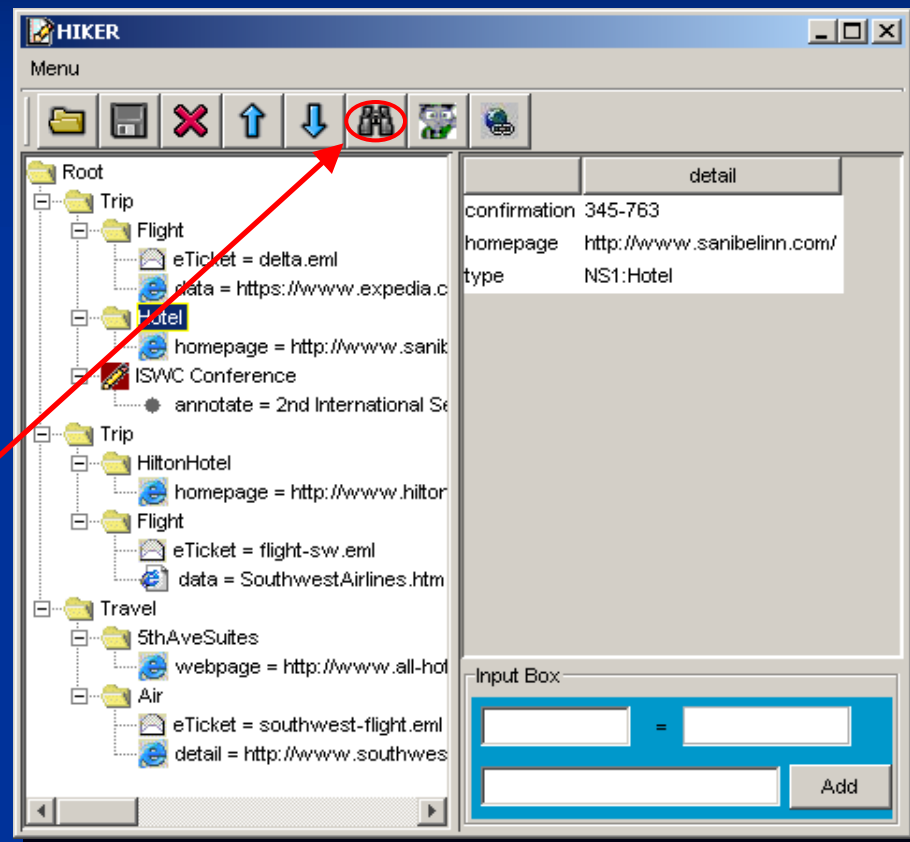
- Solved:
 - Data compaction
 - How to manage very large headers
 - Pivoting
 - Queries
- Current Focus:
 - Data formatting
 - Displaying recursive structures
 - Printing/Report Generation
- Next:
 - Scaling to very large datasets
 - Updating
 - Bookmarks, User-Profiles
 - Inverse links
 - Annotation (of rows and cells)
 - ...

Semantic Web Vision



HiKE: Semantic Desktop

- Simple, friendly Semantic Web authoring tool
 - Take notes, annotate, organize documents
- Semantic Web interrelates desktop entities
 - Drag desktop objects into hierarchy
 - Attach attribute values anywhere
 - Semantic annotations enable precise searches



- 1) Select dataset
- 2) Enhance VIVID to handle dataset
- 3) Repeat

-
- CIA Factbook
 - Handle very wide fanout
 - Sigmod Record
 - Pivoting
 - SATURN dataset
 - Recursion

What We Need From Definition Language

- Something that supports rule-based (if-and-only-if) definitions
(recall the 'MidSizedShip' example)

- Precedents:
 - Loom 'satisfies'
 - KIF 'kappa'