





Simplifying Data Access: The Energy Data Collection Project

José Luis Ambite University of Southern California Information Science Institute









The problem and the solution

- Problem: FedStats has thousands of databases in over 70 Government agencies:
 data is overlapping, sometimes near-duplicated
 even Government officials cannot find it!
- Solution: Create a system to provide easy standardized access:
 - need multi-database access engine,
 - need powerful user interface,
 - need terminology standardization mechanism.





Research challenges

- How do you scale to incorporate <u>many</u> databases? ... try to build data models automatically
- How do you integrate their data models, to allow querying across sources and agencies?
 ... take a large ontology, link the models into it automatically, and provide query reformulator
- How do you incorporate additional information that is available from <u>text</u> sources? ... use language processing tools to extract it
- How do you handle <u>footnotes</u> in the databases? ... extract them from the tables automatically









Information Integration in SIMS

[Arens+96] [Ambite+00]

To enable query access, SIMS needs to:

address semantic heterogeneity: => describe sources in common domain model

address syntactic (format) heterogeneity:
 standardize access to sources:
 Structured (DBMS): Oracle, MS Access ...
 Semistructured: *wrappers* for html, text, pdf









Wrappers

provide uniform mechanism for extracting data from semi-structured sources (HTML, text, ...)







Wrapper Building Tools

[Muslea+99]

Creating Wrappers (semi-)automatically:

- Demonstration-oriented interface enables users to show system what to extract by example
- System automatically induces extraction rules
- Common extraction engine

Benefits:

- Rapid wrapper creation
- Simplified wrapper maintenance
- Fetch.com
 - Start-up that comercializes the technology





Wrapper Building Tools

🔘 Wrapper Builder - detail.trn - [Teaching) and Learning]						
>>>> File Edit Mark-Up Tools View Wind	low Help						
📄 🖀 New 😂 Open 🔚 Save	📕 🛄 Fetch 🖧 Define 🧠 1	Teach 🧟 Learn 🗹 Verify 🤇	🔓 Link				
@ detail ☐ - ♥ Page3.html	Yahoo! Weather - Beijing		@ <u>Ad</u>	d to My Y.	ahoo!		Web E
🌾 Mostly Cloudy 🌾 3	Today	<u>F</u> ° or C°	Sat	Sun	Mon	Tue	plore
♥ -11 ♥ 80% ♥ NE/3 km/h ♥ 6 km ♥ -6 ♥ 1022.00 mb	-3° at: 10:00pm GMT (loudy LO -11° 4 4 4 4 4 4 4 4 4 4 4 4 4	Wind Chill: -2° Humidity: 80% Wind: NE/3 km/h Visibility: 6 km Dewpoint: -6° Barometer: 1022.00 mb Records & Averages Current conditions and forecas	Partly Cloudy Hi 8 Lo -6	Partly Cloudy Hi 8 Lo 1	Partly Cloudy Hi 10 Lo -4	Partly Cloudy Hi 7 Lo -1	r Raw HTML
					227 618 28 28 28	69 M2 / C. S. (162 M	
Node Name Rule Type Rule	Content						
temp Extract Be	gin_Rule _FE_ [0]ST 2 " face = " Arial " > <b d_Rule _FE_ [1] ST °</b 	or> <nobr></nobr>					
outlook Extract Be	gin_Rule _FE_ [0]ST > <font si<br="">d Rule FE	ize = " - 2 " face = " arial " >					
Pages Rules Messages							
		Project: weather; Site: ya	ahoo Mark	up sample	es and trai	in. Created or	n 2/16/2001 //









Ontology, models, data

- SENSUS ontology (90000 concepts)
- LKB glossary extracts (6000 concepts)
 - Purpose: provide rich domain knowledge
 - Sources: Census, EIA, EPA texts
- SIMS domain model (500 dimension conce, 50000 series)
 - Purpose: describe data, query planning and optimization
 - Sources: database metadata, extracted by hand
- Database sources (25000 time series)
 - EIA OGIRS: CD with MS Access database; no footnotes
- Web data sources (25000 time series)
 - EIA Multi-State: 25000 time series; formatted text
 - BLS: 60 time series; web form; 40 have footnotes
 - EIA Petroleum Supply Monthly: 20 time series in PDF; converted to HTML and then wrapped; all with footnotes
 - CEC: 12 series; static HTML tables; no footnotes





ISI's SENSUS ontology



- Taxonomy, multiple superclass links
- Approx. 90,000 concepts
- Top level: Penman Upper Model (ISI)
- Body: WordNet (Princeton), rearranged
- Used at ISI for machine translation, text summarization, database access



DESCRIPTION OF INT



Extracting metadata from text

[Klavans+01]

Problems:

INSTITUTE

- Proliferation of terms in domain
- Agencies define terms differently
- Many refer to the same or related entity
- Lengthy term definitions often contain important information which is buried

Example input:

Motor Gasoline Blending Components: Naphthas (e.g., straight-run gasoline, alkylate, reformate, benzene, toluene, xylene) used for blending or compounding into finished motor gasoline. These components include reformulated gasoline blendstock for oxygenate blending (RBOB) but exclude oxygenates (alcohols, ethers), butane, and pentanes plus. Note: Oxygenates are reported as individual components and are included in the total for other hydrocarbons, hydrogens, and oxygenates.





Lexical Knowledge Base (LKB) Tool

Analysis Combines statistical and linguistic methods: Regular Gasoline Gross-reference Gasoline Grous Term: Giasoline Head Genus Word: Gasoline Head Genus Word: Gasoline Head Genus Word: Gasoline Properties: having an antiknock index Excludes-Includes: Quantifiers: less than 88 Note: Octane requirements may vary by altitude. See Gasoline Grasoline (Regular Gasoline (Regular Gasoline (isa "Gasoline") (contains "an antiknock index") Combines statistical and linguistic methods: Indentifies topics with high accuracy Indentifies topics with high accuracy Indentifies topics with high accuracy Indentifi	Definition: Regular Gasoline: Gasoline having an antiknock index, i.e., o requirements may vary by altitude. See Gasoline Grades.	<mark>ctane rating</mark> , greater than or equal to <mark>85</mark> and less than <mark>88</mark> . Note: Octane
	Analysis Head Term: Regular Gasoline Cross-reference Gasoline Grades Genus Term: Gasoline Head Genus Word: Gasoline Properties: having : an antiknock index Excludes-Includes: Quantifiers: less than 88 Note: Octane requirements may vary by altitude. See Gasol (Regular Gasoline (isa "Gasoline") (contains "an antiknock index")	 Combines statistical and linguistic methods: trainable Finite State Recognizer identifies topics with high accuracy provides complete coverage useful for any subject area produced over 6,000 gasoline concepts from EIA and BLS



Linking Domain Models to Ontology [Hovy+01] Match DM concepts against SENSUS concepts

- 1. Name Match
 - Variations: substrings; reward for contiguous endings; stemming; variant word forms; case sensitivity; etc.
 - Implemented string match from molecular biology for DNA matching

2. Definition Match

- Variations: stemming; stop words; words or letter trigrams; word weights (*itf, tf.idf*), etc.
- Used IR vector space; cosine measure
- 3. Dispersal Match
 - Def: Given a cluster of concepts, match them all and then select one match for each concept to form tightest cluster in SENSUS
 - Variations: greedy search; number of candidates; weightings





Select Attributes and Staththites agency Trainity Trainity Trequency Trequency Trequency Trequency Trequency Trequency Trequency Trequency Treose value Troose va	Select Attributes and Stut-Attributes agency choose value family choose value family choose value protuct choose value provenance quality choose value provenance quality choose value provenance due provenance choose value provenance choose value provenance duality choose value choose value choose value choose value provenance duality choose value browenance choose value provenance choose value provenance choose value provenance choose value choose value choose value choose value choose value choose value choose value choose value choose value provenance choose value choose value choo	Select Attributes and Stub-Attributes alendy alend	Select Attributes and Studients	Select Attributes and Sub-Attributes agency agency area Torose value Torose value Torose value Prototuct Choose value Cho	Select Attributes and Sub-Attributes aleny aleny area Choose value Family Frequency Fr	Select Attributes and Sub-Attributes agency reading family frequency fre	Select Attributes and Sub-Attributes alera attributes and attributes area attributes area attribute area attrib	Select Attributes and Stu-Attributes area area finity frequency fr	Select Attributes and Stu-Attributes alera area area famiy frequency frequen	Select Attributes and Sub-Attributes alency Choose value Tamiy Choose value Tamiy Choose value Choose value point of sale product product provenance choose value product provenance choose value product choose value choose valu	state Query Browse Ontology	diate		
agency choose value unit choose value choose	agency choose value Definition area family Choose value C	agency choose value offinition and family choose value family frequency choose value choose valu	agency area family frequency frequen	agency choose value oefinition area Tamily Choose value oefinition frequency Choose value Choose value choose value or choose value or choose value choose value or choose value choose value choose value choose value choose value choose value or choose va	agency choose value of family choose value for area choose value for any choose value choose val	agency choose value Choose valu	agency choose value choose valu	agency area family forger watue for choose value frequency frequency protruct protru	agency Choose value Choose valu	agency area family family family family family family family frequency protected protect	Select Attributes and S	ub-Attributes		
area Choose value famity Choose value famity Choose value Choose value Choose value protuct Choose value protuct Choose value protuct Choose value choose value duality Choose value Choose value Unit Choose value C	area Choose value	area Choose value famity Choose value protuct of sale product product provenance quality unt choose value provenance quality choose value choose value choose	area (noose value	area Choose value (Choose valu	atea Choose value famity requency frequency Choose value Choose value Choose value product Choose value product Choose value Choose val	area famity frequencry product product product duality choose value choose value ch	atea Tamby Tequency Frequency Protect	ata Iranity Choose value Choose value Dotit of sale protett pr	ata Iamity famity fequency footse value product produc	atea Choose value famity Choose value frequency Choose value product of sale Choose value provenance Choose value and dijustment Choose value unt Choose value Unt Choose value Choose valu	agency	Choose value		Definition
familyChoose valuefrequencyChoose valuepoint of saleChoose valueproductChoose valueproductChoose valueprovenanceChoose val	family Choose value frequency Choose value point of sale Choose value product Choose value product Choose value provenance Choose value quality Choose value quality Choose value out Choose value choose value Choose va	family Choose value frequency Choose value point of sale Choose value protuct Choose value provenance Choose value quality Choose value unit Choose value choose value Submit Query	family Choose value frequency Choose value product Choose value provenance Choose value provenance Choose value provenance Choose value quality Choose value unt Choose value unt Choose value Choose value Choose value Choose value Choose value unt Choose value Submit Queny Reads Suggested Queries 2. PPI Revision Curr •	family Choose value frequency Choose value point of sale Choose value product Choose value provenance Choose value provenance Choose value quality Choose value out of sale Choose value provenance Choose value quality Choose value out of sale Choose value choose value Choose value choose value Choose value unit Choose value choose value Choose value choose value Meset supprested Queries 2. PPI Revision Curr ▼	family Choose value frequency Choose value product Choose value provenance Choose value unit Choose value unit Choose value Choose value Choose value unit Choose value Submit Query Reset Subgested Quertes 2. PPI Revision Curr	Amily Choose value frequency Choose value point of sale Choose value product Choose value product Choose value provenance Choose value provenance Choose value quality Choose value choose value Subose value	Amily requency Choose value product Choose value product Choose value product Choose value provenance Supprovenance	fanity Choose value frequency Choose value point of sale Choose value product Choose value promenance Choose value choose value Choose value promenance Submit Query promenance Z. PPI Revision Cur	Amily Choose value frequency Choose value point of sale Choose value product Choose value provenance Supprevenance	fanity Choose value frequency Choose value point of sale Choose value product Choose value quality Choose value unit Choose value Loose value Choose value Choose value Choose value unit Submit Cuery Submit Cuery Reset Submit Cuery 2. PPI Revision Curr V	area	Choose value		
frequency Choose value point of sale Choose value product Choose value provenance Choose value provenance Choose value quality Choose value unit Choose value unit Choose value seasonal adjustment Choose value unit Choose value Submit Query Reset	frequency Choose value product Choose value product Choose value provenance Choose value provenance Choose value quality Choose value out Choose value provenance Choose value quality Choose value out Choose value choose value Submit Query Subgested Queries 2. PPI Revision Cur	frequency point of sale product product provenance provenance provenance provenance provenance provenance provenance choose value choose value choo	frequency Choose value point of sale Choose value product Choose value provenance Submit Query Subgreated Queries 2. PPI Revision Cur V	frequency Choose value point of sale Choose value product Choose value provenance Choose value provenance Choose value quality Choose value unit Choose value Submit Ouery Reset Suggested Queries 2. PPI Revision Cur	If requency Choose value point of sale Choose value product Choose value provenance Choose value provenance Choose value quality Choose value quality Choose value quality Choose value quality Choose value unt Choose value Choose value Choose value Choose value Choose value unt Choose value Submit Query Reset Submit Query 2. PPI Revision Curt	frequency point of sale product provemance quality quality choose value choose valu	If requency Choose value product Choose value proventance Choose value proventance Choose value proventance Choose value proventance Choose value choose value Choose value unit Choose value unit Choose value choose value Choose value	Irequency Choose value Choose v	If equency Choose value point of sale Choose value provenance Choose value provenance Choose value quality Choose value quality Choose value choose value Choose value quality Choose value choose value Choose value unit Choose value choose value Submit Query Subgreated Queries Z. PDI Revision Curr	If equency point of sale product produ	family	Choose value		
point of sale product product provenance quality choose value choose value choo	point of sale product provenance duality eassonal adjustment unt tunt Suggested Queries Choose value Choose value Cho	point of sale product Choose value provenance Choose value quality ending choose value choose	point of sale Choose value product Choose value provenance Choose value quality Choose value quality Choose value unit Choose value Unit Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curt •	point of sale product product Provenance quality quality easonal adjustment unit unit Choose value Choose value Subini Query Choose value Choose val	point of sale Choose value product Choose value provenance Choose value provenance Choose value quality Choose value quality Choose value unit Choose value Lunit Choose value Submit Query Reset Suggested Queries 2. PDI Revision Curr.	print of sale Choose value prototuct Choose value provenance Choose value provenance Choose value quality Choose value choose value Choose value unit Choose value choose value Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curr •	point of sale Choose value product Choose value provenance Choose value quality Choose value quality Choose value quality Choose value unt Choose value unt Choose value choose value Choose value unt Choose value choose value Suppresend value	point of sale Choose value product Choose value proventance Choose value quality Choose value quality Choose value unt Choose value unt Choose value Choose value Choose value Unt Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curt	point of sale Product Product Product Provenance Proven	point of sale Choose value product Choose value provenance Choose value quality Choose value quality Choose value choose value Choose value unt Choose value choose value Choose value unt Choose value choose value Submit Query Subgested Queries 2. PPI Revision Curr	frequency	Choose value		
product provenance quality easonal adjustment unt Choose value Choose value	product provenance quality quality choose value thit choose value choose value	product Provenance quality quality choose value timt timt timt Choose value Choose value	protuct Proventance quality equality choose value choose value cho	product Choose value provenance Choose value quality Choose value quality Choose value unit Choose value unit Choose value Suggested Queries 2. PPI Revision Curr	product provenance quality easonal adjustment unit Unit Choose value Choose value C	product provenance quality quality choose value choose va	protuct proventance quality quality evalue Unit Unit Choose value Choose value Choo	product Proventance quality choose value unit choose value choose	product Choose value proventance Choose value quality Choose value quality Choose value unit Choose value unit Choose value choose value Choose value choose value Choose value unit Choose value choose value Choose value choose value Choose value submit Query Reset Submit Queries 2. PPI Revision Cur	product Provenance quality choose value unt Lint Lints Choose value Choose val	point of sale	Choose value		
provenance quality seasonal adjustment unt Choose value Choose value Choose value Choose value Choose value Choose value Choose value	provenance Choose value quality Choose value seasonal adjustment Choose value Unit Choose value Choose value Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curr ♥	provenance (uaity choose value choose val	proventance quality reasonal adjustment unit Choose value Choose va	provenance Choose value 2. PPI Revision Cur	provenance Choose value quality Choose value seasonal adjustment Choose value Unit Choose value Choose value Choose value Submit Query Reset Suggested Queries 2. PPI Revision Cur	provenance Choose value Choose	proventance quality choose value unt tunt choose value Choose valu	provenance quality reasonal adjustment tint tint tint Choose value Choose value	provenance duality seasonal adjustment Unit Unit Choose value Choose value Choo	proventance Choose value quality Choose value seasonal adjustment Choose value unit Choose value Lunit Choose value Subgested Ouerles 2. PPI Revision Curr	product	Choose value		
quality seasonal adjustment unit Choose value Choose value Reset	quality Choose value seasonal adjustment Choose value unit Choose value Choose value Image: Submit Query Suggested Queries 2. PPI Revision Curr	quality Choose value seasonal adjustment Choose value unit Choose value Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curr ◆	quality Choose value seasonal adjustment Choose value unit Choose value Choose value Ender Submit Query Reset Suggested Queries 2. PPI Revision Curr	quality Choose value easonal adjustment Choose value unit Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curr	quality reasonal adjustment unt Choose value Choose va	quality Choose value seasonal adjustment Choose value unit Choose value choose value Submit Query Submit Query Reset Suggested Queries 2. PPI Revision Curr	quality Choose value seasonal adjustment Choose value unt Choose value Submit Query Reset Suggested Quertes 2. PPI Revision Curr ◆	quality Choose value seasonal adjustment Choose value unt Choose value Choose value Submit Query Submit Query Reset Suggested Queries 2. PPI Revision Curr	quality Choose value seasonal adjustment Choose value unit Choose value Choose value Choose value Submit Query Reset Suggested Queries 2. PPI Revision Cur V	quality Choose value unit Choose value unit Choose value Choose value Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curr	provenance	Choose value		
seasonal adjustment Choose value Choose value Choose value Submit Query Reset	seasonal adjustment Choose value Choose value Choose value Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curr	seasonal adjustment unit Choose value Choose value Submit Query Suggested Queries 2. PPI Revision Curr▼	seasonal adjustment Unit Choose value Choose value Submit Query Suggested Queries CPI Revision Curr	seasonal adjustment Unit Choose value Choose value Submit Query Suggested Queries C.PPI Revision Cur	seasonal adjustment Unit Choose value Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curr	seasonal adjustment unit Choose value Choose value Submit Query Reset Suggested Queries 2. PPI Revision Cur ▼	seasonal adjustment Unit Choose value Choose value Submit Query Suggested Queries CPPI Revision Curr V	seasonal adjustment unt Choose value Choose value Submit Query Suggested Queries 2. PPI Revision Cur	easonal adjustment Linit Choose value Choose value Submit Query Reset Choose value Submit Query Reset Chorse value Choose value Cho	seasonal adjustment Unit Choose value Choose value Submit Query Submit Query Suggested Queries 2. PPI Revision Curt ✓	quality	Choose value		
unit Choose value Submit Query Reset	unit Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curr	unit Submit Query Suggested Queries 2. PPI Revision Curr	unt Submit Query Reset Suggested Queries 2. PPI Revision Curr	unit Choose value Submit Query Reset Reset 2. PPI Revision Curr V	unit Submit Query Suggested Queries 2. PPI Revision Curr	Unit Submit Query Baseted Queries 2. PPI Revision Curr	Choose value Choose value Submit Query Reset Suggested Queries 2. PPI Revision Curr ♦	unit Choose value Submit Query Reset Beet Suggested Quertes 2. PPI Revision Curr V	Introduction Choose value Submit Query Reset Subgreated Queries 2. PPI Revision Curr	Choose value Submit Query Reset Utr	seasonal adjustment	Choose value		
Submit Query Reset	Submit Query Reset Suggested Queries 2. PPI Revision Curr	Submit Query Reset Suggested Queries 2. PPI Revision Curr	Suggested Queries 2. PPI Revision Curr	Suggested Queries 2. PPI Revision Curr	Suggested Queries 2. PPI Revision Curr	Suggested Queries 2. PPI Revision Curr	Submit Query Reset Suggested Oueries 2. PPI Revision Curr	Submit Query Reset Suggested Queries 2. PPI Revision Curr	Submit Query Reset Suggested Queries 2. PPI Revision Curr V	Submit Query Reset Suggested Queries 2. PPI Revision Curr	unit	Choose value		
	Suggested Queries 2. PPI Revision Curr	Suggested Queries 2. PPI Revision Curr 👻	Suggested Queries 2. PPI Revision Curr	Suggested Queries 2. PPI Revision Curr	Suggested Queries 2. PPI Revision Curr	Suggested Oueries 2. PPI Revision Curr	Suggested Oueries 2. PPI Revision Curr V	Suggested Queries 2. PPI Revision Curr	Suggested Queries 2. PPI Revision Curr	Suggested Queries 2. PPI Revision Curr		Submit Query	Reset	
Suggested Queries 2. PPI Revision Curr												Suggested Queries 2	2. PPI Revision	1 Curr 👻



http://vigor.isi.edu:8005/swing/work/swing.html - Microsoft Int	ernet Explorer		
Main Display Generate Query Browse Ontology Help			
History	Results and Footnotes		
Query 1	19981100	53.2	•
	19981200	45.4	
	19990000	70.3	
	19990100	49.5	
	19990200	47.0	
	19990400	69.4	
Terms Used	19990500	712	
BLS	19990600	68.9	
USA PC	19990800	81.8	
tuesday 13	19990900	85.6	
point of sale unleaded regular	19991000	80.4	
BLS	19991200	82.6	
revision current series not seasonally adjusted	20000100	83.0 (P)	
dollars per dallon	2000200	93.4 (P)	
Original Source Data	2000300	107.4 (P)	
	20000400	101.6 (P)	
	P : Preliminary. All indexes are subi	ect to revision four months after original pub	lication.
		•	
			100

(\mathcal{O})	X	(III)		ks »									1									•	1
X	i			ہ																			
D			~	S.G.								-					-						met
X			Discuss	×							Ann		49.3	68.1	104.7	121.4	112.7	101.7	95.7	94.0	60.6	64.5	🚺 Inte
			+ # العا								Dec	47.4	52.8	84.9	108.7	120.9	110.0	99.2	93.6	95.9	51.8	62.7	
											Nov	17.4	51.7	32.8	108.2	121.3	111.7	101.1	95.2	95.6	51.3	56.5	-
			Print								Oct	7.6	1.8	1.3	07.9	21.7	13.2	03.4	4.6	3.1	1.0	5.8	
			Mail •								en	7.8	1.3 5	7.9 8	08.7 1	21.9 1	15.1	04.4	2.9	4.8	5.7 5	7.0 6	
											BIL	<u>م</u>	.5 5	.4	8.6 1	2.2 1:	6.7 1	15.3	8.8	6.9	.4 5:	1.7 6	
			s Hist								P In	0.	.4 50	0.72	8.2 10	3.4 12	5.3 11	4.5 10	7 93	.1 96	.1 51	.2 70	
	rer		Favorite		ies							6 48	4 49	8 69	7.7 10	4.7 12	7.3 11	3.3 10	8 96	0.0 99	0 52	4 68	
	et Explo		Search		ut Ser						Υ Γ	4	4	3 64	8.2 10'	5.3 124	2.8 10	8 10,	0 98.	8 10	8 65.	4 66.	
	Interne		 		urrei						M		47.	61.	108	125	102	98.	.66	97.	62.	64.	
	Aicrosoft	Heb	Esh Hor	rgate	sion-C						Apr	•	46.9	58.0	106.3	126.2	107.5	95.1	96.7	92.4	56.2	64.0	
	s Data - N	s Iools	E B	{4/cgi-bin/	ex Revi		34		ar g		Mar		46.8	55.9	101.5	124.3	114.7	98.0	96.0	87.4	61.9	50.5	
	Statistic	Favorite	St St	146.142.4.2	ce Inde		2911#13	Adjusted	eum Refi led regul 6		Feb		46.9	54.3	93.8	114.7	117.9	101.7	95.3	85.8	79.0	60.7	
	of Labo	t <u>V</u> iew	Forma	http://	er Pri	atalog	: PCU	sonally.	: Petrol Unlea te : 850		Jan		7.3	3.5	8.6	10.3	20.1	0.0	5.2	9.5	9.2	7.1	
	Bureau	jie <u>E</u> d	가 ⁸⁸	dress 🧟	roduc	eries (eries III	ot Sea	dustry oduct ase Dat	ata:	ear	977	978 4	979 5	980 8	981	982	983 1	984 9	985 8	986 8	987 5	Done
	ių.			P4	18 18	ŭ I	ي الــــــ		មកក	<u> </u>												_	
			t Interne [ools <u>F</u>	t) Befresh	i-bin/srga	Contraction of the	Ser		ty Spec	TT.													
			incrosof ontes	8 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2.4.24/cg	<mark>b</mark>			d Surve	elow:	U.											*	
			eport - M W Fav	vard +	//146.142	formatio		d Help	ram an	d(s) be													
		all	dit ⊻ie	∎ ee	🛃 http:/	bility In		tion an	S Prop	D Forr	** • • • • • • • • • • • • • • • • • • •												
S	SCIENCE	INTIN	Elle E	↓	A <u>d</u> dress	Accessi		Informa	ABL	Series I Enter s	67004											- 18] Done
		1	2																				10

×			1	(ĺ
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
					l
-					l
lore					l
ц×					l
et					l
en					l
Ē					l
soft					l
Ĩ.	1				l
Ĭ	lelp	arch	¥ 8		l
	는	Sei	ä		l
1	8				l
E	fol				l
ls.	õ				l
¥.	NSE	e			l
3	20	soli			l
- E		8			l
/sw	fen	E I			
S	õ	l je			
18	ate	ade			
edi	nei	le			
isi.	Ge		8		
ğ	8	Tern	Seel		
2	spl	For	Lm		
à	1 Di	ch.	t Te		
Ŧ	đaù	Sear	Las		
12		05			

\geq			1		< >
-					
				×	
					SIMS)
					ElAGU
					lineade (from the first second s
					gasoline.c
					regular
				cepts.	IS : pro
ier				al con	3U: unlar 3U: unl
et Explo				sever	ne ^A EIA
Intern				turnec	r gasoli
icrosoft	0	L		Concep arch re elect c	regula let Winn
tml - Mi	IN Hel	Searc	Back	<mark>Pick a</mark> Dur sea ease s	oduct.g ning: Apr
swing.h	Ontolo	Π			
/work/	rowse				
/swing.	leny B				
u: 8005	rate Qu				
r.isi.ed	Genel	;w	:Ua		
://vigo	Display	h For Ter	ferm See		
tttp	Main	Search	Last		

🛔 http://vigor.isi.edu:8005/swing/work/swing.html - Microsoft	oft Internet Explorer	×
Main Display Generate Query Browse Ontology Help		
Search For Term: Search Last Term Seen: Search Back		
Unleaded regular gaso Gasoline having an antiknock index (R+M/2) greater tha lead or 0.005 grams of phosphorus per gallon. OPI: EI-4	oline (from EIAGU) han or equal to 85 and less than 88, and containing not more than 0.05 grams of 1-40 Sources: EIA-782A, EIA-782B, EIA-782C, DOE/EIA-0487	
Core Definition Gasoline having an antiknock index (R+M/2) greater tha ead or 0.005 grams of phosphorus per gallon	han or equal to 85 and less than 88, and containing not more than 0.05 grams of	
source Agency EIA		
un word: unleaded regular gasoline femainder of definition: OPI: EI-40 Sources: EIA-782A, EIA-782A, EIA-782A source URL: http://www.cs.columbia.edu/~bwhitman/filb/reia_large_femanonestreaded regular (from SIMS) source URL: http://www.cs.columbia.edu/~bwhitman/filb/reia_large_femanonestreaded regular gasoline source URL: http://www.cs.columbia.edu/~bwhitman/filb/reia_large_femanonestreaded regular gasoline	unteaded regular gasoline (from EIA©L) S2C. DOE/EIA:0487 S2C. DOE/EIA:0487	()





USC INFORMATION SCIENCES INSTITUTE

Conclusion

Large-scale information integration Major research challenges: Semi-automated domain modeling, information extraction, linkage to ontology Efficient query processing Footnote extraction and display Major practical challenges: Understanding users' needs: interface, information ... Getting a lot of data into the system