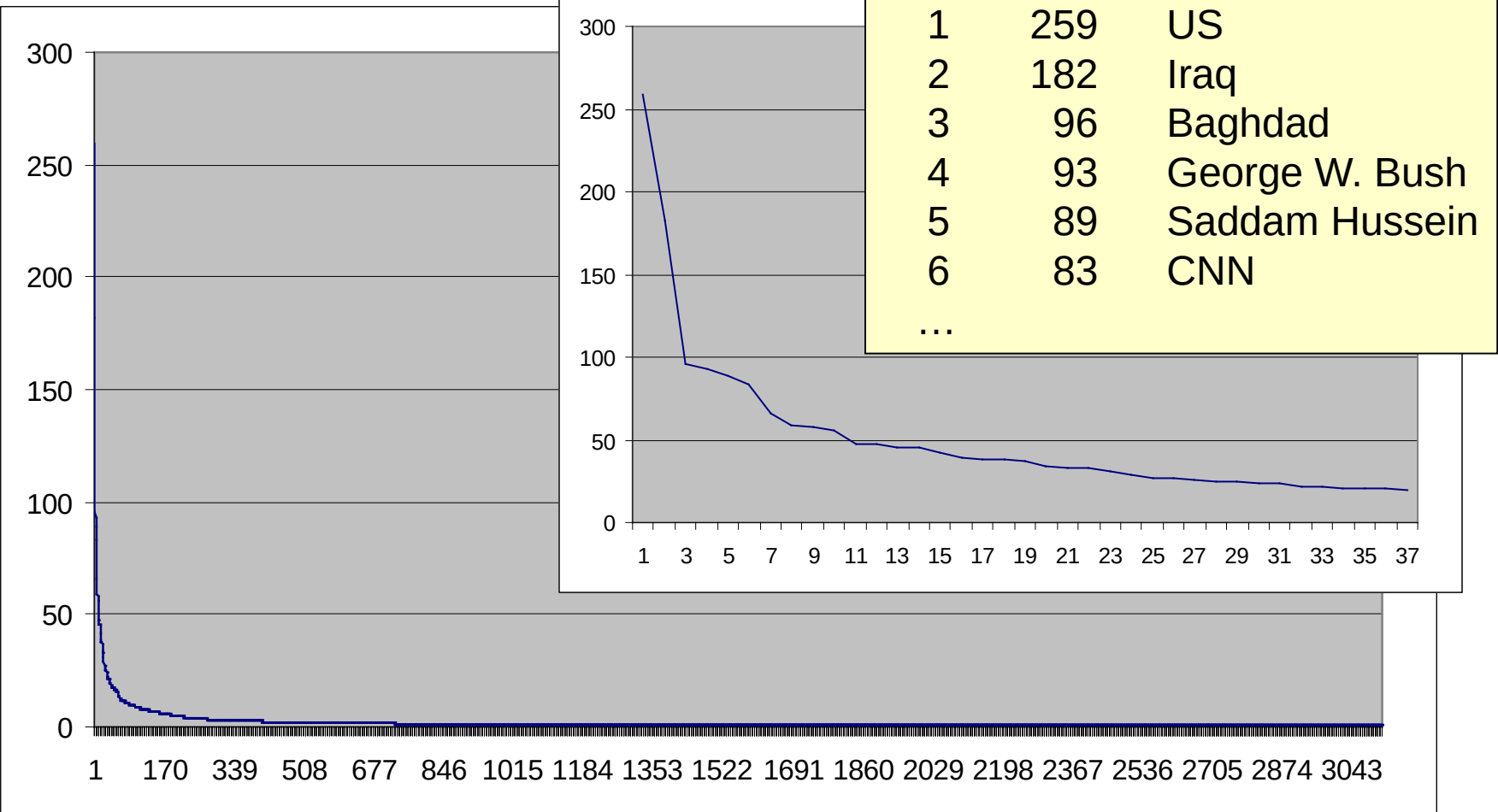


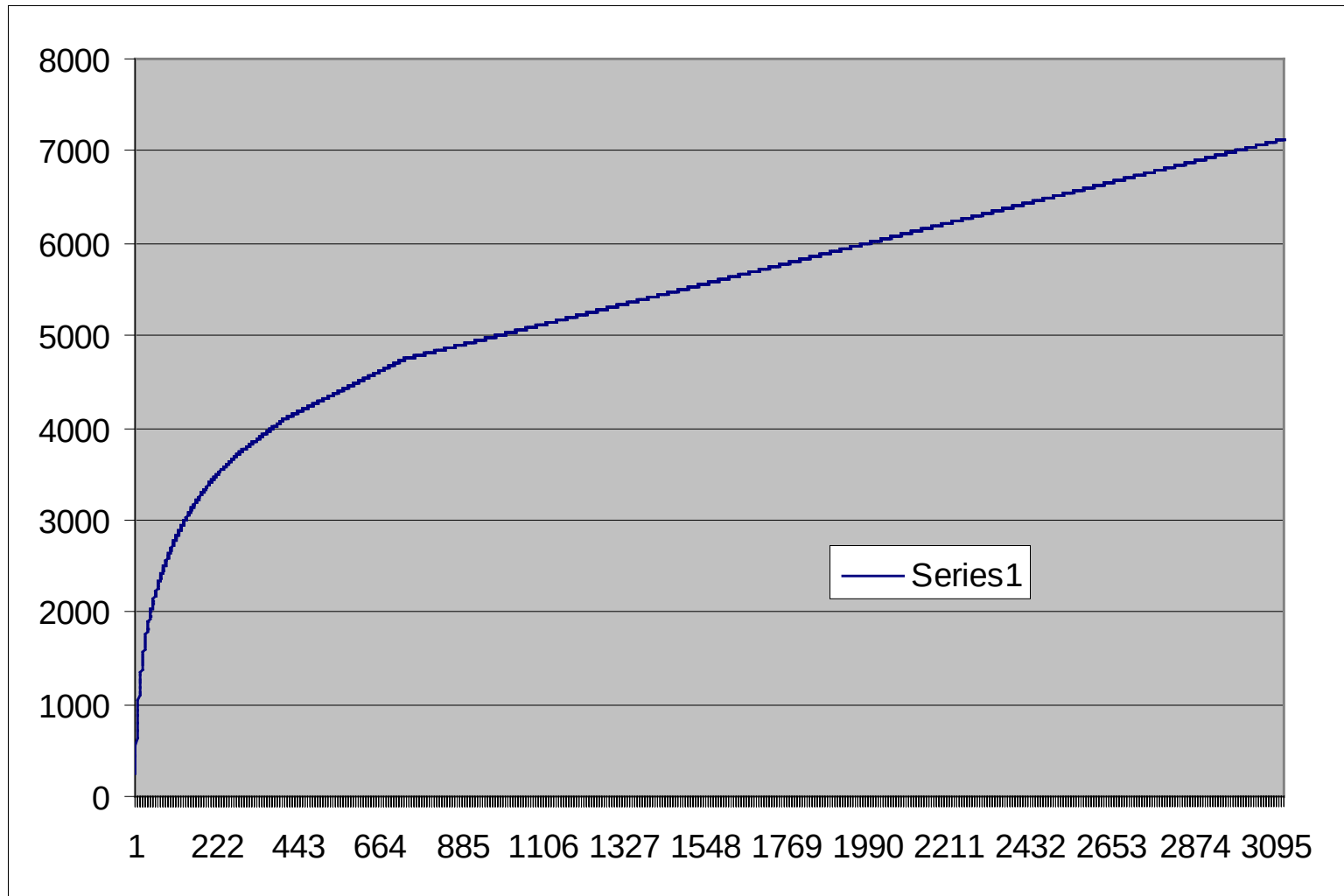
Corpus Statistics

- ACE2005/ACE2007 English EDR
 - Chars: 1.5M Words: 257K
 - Entities: 18K (PER 9.7K, ORG 3K, GPE 3K, FAC 1K, LOC 897, WEA 579, VEH 571)
 - Mentions: 55K (PRO 20K, NAM 18K, NOM 17K)
- CDC Entities (PER, ORG, LOC, GPE)
 - IDC Entities 7,129 (Entities with at least one name)
 - CDC Entities 3,660 (after manual linking)
 - 2,390 singleton entities
- CDC Annotation Effort
 - Approximately 2 staff weeks
 - Annotated after automatic pre-linking of entities that shared at least one identical (case-sensitive) name string

Cross-Document Entity Mention Count Histogram



Total Mentions Covered by Frequency-Sorted Entities



Callisto/EDNA

- Entity Disambiguation and Normalization Annotation (EDNA) tool
 - A plug-in for Callisto client
 - Multiple annotators supported with single Tomcat server (with document locking)
 - Document set indexed by APF-customized Lucene search engine
- Assumes documents annotated for ACE EDR (entity mentions and intra-document coreference)

Logging onto the Server

Import from Server...

Server Login...

Server URL:

User:

Password:

Get List

Last listing time:

Files listed: Files locked by you:

☐ Show All Users' Files

Current Selection:

File	Status	Locked By	Locked Since
------	--------	-----------	--------------

Import **Cancel**

File Selection, Locking & Status

Import from Server...

Server Login...

Server URL:

User:

Password:

Last listing time:

Files listed: Files locked by you:

☐ Show All Users' Files

Current Selection:

File	Status	Locked By	Locked Since
CNN_CF_20030303.1900.02.sgm	2/26		
CNN_CF_20030303.1900.06-1.sgm	2/20		
CNN_CF_20030303.1900.06-2.sgm			
CNN_CF_20030304.1900.02.sgm			
CNN_CF_20030304.1900.04.sgm			
CNN_CF_20030304.1900.06-2.sgm			
CNN_CF_20030305.1900.00-1.sgm			
CNN_CF_20030305.1900.00-2.sgm			
CNN_CF_20030305.1900.00-3.sgm			
CNN_CF_20030305.1900.02.sgm	1/24	edna3 ☒ Locked	Wed Jun 14 13:06:35 EDT 2006
CNN_CF_20030305.1900.06-1.sgm			
CNN_CF_20030305.1900.06-2.sgm			
CNN_IP_20030328.1600.07.sgm			
CNN_IP_20030329.1600.00-2.sgm			
CNN_IP_20030329.1600.00-3.sgm			
CNN_IP_20030329.1600.00-4.sgm			
CNN_IP_20030329.1600.00-5.sgm			
CNN_IP_20030329.1600.00-6.sgm			
CNN_IP_20030329.1600.01-1.sgm			
CNN_IP_20030329.1600.01-3.sgm			
CNN_IP_20030330.1600.05-2.sgm			

Highlighted Mentions and ACE Annotations

The screenshot displays the Callisto application interface, which is used for managing and analyzing entity mentions. The interface is divided into several panes:

- Source document:** The top-left pane shows a news article snippet from CNN. It contains text about military operations in the Persian Gulf and mentions of U.S. Army soldiers. A red box highlights a portion of the text, and a green box highlights another portion.
- ACE Annotations:** The top-right pane displays the search results summary, including a table with columns for External Ref, Document Entity ID, and Text. A red box highlights a specific row in the table.
- Current Entity ID(s):** The bottom-left pane shows a list of entity IDs and their corresponding mentions. It includes checkboxes for various entity types (Name(s), Pronoun(s), Nominal(s), Comments, Relationships) and a search button.
- Entities:** The bottom-right pane shows a detailed table of entity mentions, including columns for External Ref, Document Entity ID, Primary Reference, Mentions, Type, Subtype, Class, and Comments. A red box highlights a specific row in the table.

Annotations are highlighted in the source document and the ACE Annotations pane. The source document shows mentions of "U.S. Army soldiers" and "former Congressman Tom Andrews". The ACE Annotations pane shows a list of entity mentions, including "Tom Andrews" and "former Congressman Tom Andrews".

External Ref	Document Entity ID	Primary Reference	Mentions	Type	Subtype	Class	Comments
	E1	Tom Andrews	11	PER	Individual	SPC	
	E2	Maine	1	GPE	State-or-Province	SPC	
	E4	Our	27	GPE	Nation	SPC	
	E6	Congressman	13	PER	Individual	SPC	
	E7	Win Without	1	ORG	Non-Governmental	SPC	
noncanonical-1150310670782	E10	California	2	GPE	State-or-Province	SPC	
	E11	I	4	PER	Individual	SPC	
	E12	Americans	1	PER	Group	USP	
	E13	people	1	PER	Group	USP	
	E14	world	1	LOC	Geographical	SPC	

Default and Customizable Entity Search

The screenshot displays the EDNAFACE Event Task application interface. It features a search results window with a table of entities, a 'Selected Entity Details' panel, and a 'Current Entity' configuration panel. The search results are sorted by 'Highest Ranked' and show a list of entities with their external references, document entity IDs, and text. The 'Selected Entity Details' panel shows the details for the entity 'Maine, State of' (g-1780787), including its type, subtype, name, and relationships. The 'Current Entity' panel shows the configuration for the current entity, including the entity ID, search criteria, and search options.

Entity-based Search Criteria

Search Results

Selected Entity Details

External Ref	Document Entity ID	Text
571569	g-571569	Lucerne in Maine, Lucerne-in-Maine
18551251	g-18551251	Maine d'Euche
18551253	g-18551253	Maine-Chevalier
18558225	g-18558225	Le Maine
18558226	g-18558226	Le Maine
18558227	g-18558227	Le Maine
18558228	g-18558228	Le Maine
18562962	g-18562962	La Maine
253730	g-253730	Maine Prairie
430310	g-430310	Maine, Township of
430311	g-430311	Maine, Township of
469332	g-469332	Maine, Township of
665878	g-665878	Maine, Township of
950162	g-950162	East Maine
980184	g-980184	Maine, Town of
1584630	g-1584630	Maine, Town of
1584631	g-1584631	Maine, Town of
1780787	g-1780787	Maine, State of
1038246	g-1038246	Maine, Township of

Current Entity

Current Entity ID(s): CNN_CF_20030303.1900.02-E2

☐ Show Disambiguated View

☐ Search Canonical Only

☒ Type/Subtype

☒ Type:"GPE"

☐ Subtype:"State-or-Province"

☒ Name(s)

☒ Maine

☐ Pronoun(s)

☐ Nominal(s)

*type:"gpe" +(name:"maine") -eid:cnn_cf_20030303.1900.02-e2

Search Details

View In Context: g-1780787

Type-Subtype: GPE-countries, 2nd order divisions

Name(s): Maine, State of

Pronoun(s): N/A

Nominal(s): N/A

Comments: N/A

Relationships: N/A

Link result to selected entity [Link](#)

External Ref	Document Entity ID	Primary Reference	Mentions	Type	Subtype	Class	Comments
	E1	Tom Andrews	11	PER	Individual	SPC	
	E2	Maine	1	GPE	State-or-Province	SPC	
	E4	Our	27	GPE	Nation	SPC	
	E6	Congressman	13	PER	Individual	SPC	
	E7	Win Without War	1	ORG	Non-Governmental	SPC	
noncanonical-1150310670782	E10	California	2	GPE	State-or-Province	SPC	
	E11	I	4	PER	Individual	SPC	
	E12	Americans	1	PER	Group	USP	
	E13	people	1	PER	Group	USP	
	E14	world	1	LOC	Celestial	SPC	

Font: 12pt. Default | Charset: UTF-8 | Task: EDNAFACE Event Task

Color Coding Entity Status & Type

Callisto - CNN_CF_20030303.1900.02.apf.xml*

File Edit Format Tools Panes Help

Main

Search Results Linked Entities

Search Results Summary

Sort By: Highest Ranked

External Ref	Document Entity ID	Text
1657382	g-1657382	Maine Prairie Station, Vale
18498230	g-18498230	Clonmany, Cluain Maine
18900208	g-18900208	Maine, Mai?ne? ??
18900032	g-18500032	Chevaigne?, Chevaigne, Chevaigne?-du-Maine, Ch...
18582142	g-18582142	Chalons, Cha?lons, Cha?lons-du-Maine, Chalons...
33699	g-33699	??, Parks, Park, Maine
990248	g-990248	??, Maine, Oak Grove
18551250	g-18551250	De?partement de Maine-et-Loire, Maine-et-Loire, D...
18634067	g-18634067	Saint-Hilaire, Saint-Hilaire-du-Maine
		Saint-Fiacre, Saint-Fiacre-sur-Maine
		Saint-Ellier, Saint-Ellier-du-Maine
		Longecourt, Longecourt-en-Maine, Longecourt-en-P...
		La Ferrie, La Ferrie au Maine
		Le Bignon, Le Bignon-du-Maine
		Saint-Denis, Saint-Denis-du-Maine, ???
		Saint-Ouen, Saint-Ouen-sur-Maine, Saint-Ouen-sur...
		Mai?ne?-Soroa, Arondissement de Mai?ne?-Soroa...
		??, Desplaines Station, Des Plaines, Maine, Rand
		????, Maine, ??, Dunning, Leyden Center, Monroe

NOYAK

Tom Andrews, I think we all realize that a government doesn't go to war a nation goes to war. And so, would like you to take a look at the CNN/USA TODAY/Gallup poll, taken last week, should U.S. troops to go to Iraq to remove Saddam Hussein from power. Take a look at it. Favor 59% opposed 37%, that's a vastly larger support than President Bush Senior had in getting the U.S. troops out of Kuwait before that war started. That's pretty good support isn't it?

ANDREWS

Now, Bob, come on you don't really buy this. I mean, listen, this is the oldest trick in the book. You can have a general question like this that could mean anything and ask people and they give you what comes off the top of their head. But, ask them another question, ask them what they think about spending \$1.3 trillion in destroying this economy. Ask them about going and not just a war, Bob, but an invasion and occupying for us to 10 years a sovereign Arab nation in the midst of one of the most

Current Entity

Current Entity ID(s): CNN_CF_20030303.1900.02-E2

☐ Show Disambiguated View

☐ Search Canonical Only

☒ Type/Subtype

☒ Type:"GPE"

☐ Subtype:"State or Province"

☒ Name(s)

☒ Maine

☐ Pronoun(s)

☐ Nominal(s)

*type:"gpe" +(name:"maine") -eid:cnn_cf_20030303.1900.02-e2 -eref:1780787

Search

Entities Entity Mentions Relation Mentions Events Event Mentions Values Timex2

External Ref	Document Entity ID	Primary Reference	Mentions	Type	Subtype	Class	Comments
	E14	world	1 LOC		Celestial	SPC	
	E15	administration	3 ORG		Government	SPC	
	E16	president	8 PER		Individual	SPC	
	E17	whom	15 PER		Individual	SPC	
noncanonical-1150304841239	E20	Saddam Hussein	10 PER		Individual	SPC	This is the bad guy, all right.
	E21	troops	1 PER		Group	USP	
	E22	Iraq	4 GPE		Nation	SPC	
	E24	program	1 ORG		Government	SPC	
	E25	soldiers	1 PER		Group	USP	
	E26	weapons	3 MFA		Chemical	USP	

Font: 12pt Default | Charset: UTF-8 | Task: EDNAACE EventTask

Start Catalina Shortcut to callisto.bat Callisto - CNN_CF_200... Microsoft PowerPoint - ... 6:19 PM

Reviewing Target Link Target in Context of Source Document

Document:
Entity ID:
NOTE: This document is intended for reference only and the underlying file has not been locked against change while you view it.

CNN_ENG_20030612_072835.2
NEWS STORY
2003-06-12 07:39:00

want to get to florida. a massive manhunt underway where a rapist is targeting children. dna testing has conclusively linked one man. the two most recent victims, age 11 and 13. they also believed one of those attacked in this string is age 79 as well. we talked with the lieutenant in miami and there is big time concern in the shenandoah neighborhood, southwestern portion of miami. overseas in **iraq**, a major operation underway.

2003-06-12 07:39:28

☒ Type/Subtype
☒ Type:"GPE"
☒ Subtype:"Nation"
☒ Name(s)
☒ Iraq
☐ Pronoun(s)
☐ Nominal(s)
+type:"gpe" +subtype:"nation" +(name:"iraq") -eid:cnn_cf_20030303.1900.02-e22

Search Results Summary

Sort By: **Highest Ranked**

External Ref	Document Entity ID	Text
	CNN_ENG_20030612_072835.2-E1	iraq
	CNN_ENG_20030612_072835.2-E1	iraq
	CNN_CF_20030303.1900.00-E75	iraq
	CNN_CF_20030305.1900.02-E90	iraq, a defenseless nation, They, them, there
	CNN_CF_20030305.1900.06-E29	iraq
	CNN_IP_20030402.1600.00-4-E43	iraq, a new government, they
	CNN_IP_20030412.1600.05-E4	iraq, iraqi
	CNN_LE_20030504.1200.02-2-E39	iraq, the country, there
	CNNHL_ENG_20030304_142751.10-E35	iraq
	CNNHL_ENG_20030415_193729.5-E19	iraq
	CNNHL_ENG_20030523_221118.14-E6	iraq
	CNN_ENG_20030304_173120.16-E19	iraq
	CNN_ENG_20030305_170125.1-E32	iraq
	CNN_ENG_20030306_083604.6-E18	iraq
	CNN_ENG_20030328_150609.10-E42	iraq, there
	CNN_ENG_20030401_073033.14-E19	iraq
	CNN_ENG_20030407_170605.7-E12	iraq
	CNN_ENG_20030411_233701.11-E15	iraq
	CNN_ENG_20030418_063040.1-E28	iraq
	CNN_ENG_20030431_120500.12-E2	iraq, the country,

Search Details

View In Context: **CNN_ENG_20030612_072835.2-E1**

Type/Subtype: GPE-Nation
Name(s):
iraq
Pronoun(s): N/A
Nominal(s): N/A
Comments: N/A
Relationships: N/A

Link result to selected entity [Link](#)

External Ref	Document Entity ID	Primary Reference	Mentions	Type	Subtype	Class	Comments
	E14	world	1	LOC	Celestial	SPC	
	E15	administration	3	ORG	Government	SPC	
	E16	president	8	PER	Individual	SPC	
	E17	whom	15	PER	Individual	SPC	
noncanonical-1150304841239	E20	Saddam Hussein	10	PER	Individual	SPC	This is the bad guy, all right.
	E21	troops	1	PER	Group	USP	
	E22	iraq	4	GPE	Nation	SPC	
	E24	program	1	ORG	Government	SPC	
	E25	soldiers	1	PER	Group	USP	
	E26	weapons	3	MA	Chemical	USP	

Font: 12pt Default | Charset: UTF-8 | Task: EDNA/ACE EventTask

Type Restrictions in Search Can Be Relaxed

Callisto - CNN_CF_20030303.1900.02.apf.xml*

File Edit Format Tools Panes Help

Main

city on the earth in a minute mean get the hold of the means to do it.

NOVAK

Tom Andrews. I think we all realize that a government doesn't go to war a nation goes to war. And so I would like you to take a look at the CNN/USA TODAY/Gallup poll, taken last week, should U.S. troops to go to Iraq to remove Saddam Hussein from power. Take a look at it. Favor 59% opposed 37%, that's a vastly larger support than President Bush Senior had in getting the U.S. troops out of Kuwait before that war started. That's pretty good support isn't it?

ANDREWS

Now, Bob, come on you don't really buy this. I mean, listen, this is the oldest trick in the book. You can have a general question like this that could mean anything and ask people and they give you what comes off the top of their head. But, ask them another question, ask them what they think about spending \$1.3 trillion in destroying this economy. Ask them about going and not just a war, Bob, but an invasion and occupying for us to 10 years a sovereign Arab nation in the midst of one of the most

Search Results

Linked Entities

Search Results Summary

Sort By: Highest Ranked

External Ref	Document Entity ID	Text
26067667	g-26067667	Al 'Ira?q, Al 'Iraq
16932751	g-16932751	Iragh, Iraj, I'ra?i, 'Iraq, 'Ira?q
19028124	g-19028124	'Ira?q al Ami?r, 'Iraq al Amir, 'Iraq el Ami?r, 'Iraq el A...
19029851	g-19029851	Al 'Ira?q, Al 'Iraq, Arak, El 'Ara?q, El 'Araq, El 'Ira?q, ...
19028125	g-19028125	'Ira?q al 'Abd, 'Iraq al 'Abd, 'Ira?q el 'Abd, 'Iraq el 'Abd
16897518	g-16897518	Iraq, Al 'Ira?q, Al 'Iraq, Irak, Republic of Iraq, Al Jumh...
16946481	g-16946481	Arak, Ara?k, 'Iraq, 'Ira?q, Soltanabad, Solt'a?na?ba...
16817472	g-16817472	Al 'Ara?q, Al 'Araq, Al 'Arq, Al 'Ira?q, Al 'Iraq

Current Entity

Current Entity ID(s): CNN_CF_20030303.1900.02-E22

Show Disambiguated View

Search Canonical Only

Type/Subtype

Type:"GPE"

Subtype:"Nation"

Name(s)

Iraq

Pronoun(s)

Nominal(s)

Search

Search Details

View In Context: g-16897518

Type-Subtype: GPE-countries

Name(s):

Iraq

Al 'Ira?q

Al 'Iraq

Irak

Republic of Iraq

Al Jumhu?ri?yah al 'Ira?qi?yah

Al Jumhuriyah al 'Iraqiyah

???

??????

????

Pronoun(s): N/A

Nominal(s): N/A

Comments: N/A

Relationships: N/A

Link result to selected entity Link

Entities

Entity Mentions

Relation Mentions

Events

Event Mentions

Values

TimeX2

External Ref	Document Entity ID	Primary Reference	Mentions	Type	Subtype	Class	Comments
	E14	world	1 LOC		Celestial	SPC	
	E15	administration	3 ORG		Government	SPC	
	E16	president	8 PER		Individual	SPC	
	E17	whom	15 PER		Individual	SPC	
noncanonical-1150304841239	E20	Saddam Hussein	10 PER		Individual	SPC	This is the bad guy, all right.
	E21	troops	1 PER		Group	USP	
noncanonical-1150324010215	E22	Iraq	4 GPE		Nation	SPC	
	E24	program	1 ORG		Government	SPC	
	E25	soldiers	1 PER		Group	USP	
	E26	was none	3 MFA		Chemical	USP	

Font: 12pt. Default | Charset: UTF-8 | Task: EDNAACE EventTask

Start

Catalina

Shortcut to callisto.bat

Callisto - CNN_CF_200...

Microsoft PowerPoint - [...]

6:30 PM

Annotator Comments can be Added and Retained

Callisto - CNN_CF_20030303.1900.02.apf.xml*

File Edit Format Tools Panes Help

Main

Search Results Linked Entities

Search Results Summary

Sort By: Highest Ranked

External Ref	Document Entity ID	Text
	FLOPPINGACES_20050217.1237.014-E1	Bush
	FLOPPINGACES_20050217.1237.014-E1	Bush
	CNN_CF_20030303.1900.02-E16	Bush, Our president, our president, he, him
	CNN_CF_20030305.1900.00-2-E2	Bush
	CNN_LE_20030504.1200.02-2-E29	Bush
	CNN_ENG_20030508_170552.18-E22	bush
	fsh_29782_2-E11	Bush, an idiot
	FLOPPINGACES_20041114.1240.039-E10	Bush, our president, him, his
	FLOPPINGACES_20041116.0833.027-E49	Bush
	MARKETVIEW_20050215.1858-E2	Bush
	CNN_CF_20030305.1900.00-2-E2	Bush
	CNN_LE_20030504.1200.02-2-E29	Bush
	CNN_ENG_20030508_170552.18-E22	bush
	fsh_29782_2-E11	Bush, an idiot
noncanonical-1150253546758	FLOPPINGACES_20041114.1240.039-E10	Bush, our president, him, his
	FLOPPINGACES_20041116.0833.027-E49	Bush

Current Entity

Current Entity ID(s): CNN_CF_20030303.1900.02-E28

☐ Show Disambiguated View

☐ Search Canonical Only

☒ Type/Subtype

☒ Name(s)

☒ Bush

☐ Pronoun(s)

☐ Nominal(s)

+type:"per" +subtype:"individual" +(name:"bush") -eid:cnn_cf_20030303.1900.02-e28

Search

Search Details

View In Context: N/A

Type-Subtype: N/A

Name(s): N/A

Pronoun(s): N/A

Nominal(s): N/A

Comments: N/A

Relationships: N/A

Link result to selected entity [Link](#)

Entities Entity Mentions Relation Mentions Events Event Mentions Values TimeX2

External Ref	Document Entity ID	Primary Reference	Mentions	Type	Subtype	Class	Comments
	E26	weapons	3	WEA	Chemical	USP	
	E27	Newsweek* magazine	3	ORG	Media	SPC	
	E28	Bush	3	PER	Individual	SPC	George Herbert Walker Bush -- father of George Walker Bush
	E30	CIA	1	ORG	Government	SPC	
	E31	White House	1	FAC	Building-Grounds	SPC	
	E32	men	1	PER	Group	USP	
	E33	person	1	PER	Individual	GEN	
	E34	us	3	PER	Group	SPC	
	E35	Clinton	2	PER	Individual	SPC	
	E36	Rangers	4	PER	Group	SPC	
	E37	18	1	PER	Group	SPC	
	E38	Matt Reersen	1	PER	Individual	SPC	
	E39	families	1	PER	Group	SPC	
	E40	Walter Reed	1	FAC	Building-Grounds	SPC	
	E44	city	1	GPE	Population-Center	GEN	
	E45	earth	2	LOC	Celestial	SPC	
	E46	I	5	PER	Individual	SPC	

Font: 12pt. Default | Charset: UTF-8 | Task: EDNAACE EventTask

Start Catalina Shortcut to callisto.bat Callisto - CNN_CF_200... Microsoft PowerPoint - ... 6:44 PM