# ACE 2008: Cross-Document Annotation Guidelines (XDOC)

## Version 1.6 - 2008.05.06

Linguistic Data Consortium

http://projects.ldc.upenn.edu/ace/

### Overview

The objective of the Automatic Content Extraction (ACE) series of evaluations is to develop human language understanding technology to provide automatic detection and recognition of key information about real-world entities, relations and events in source language text, and to convert that data into a structured form.

For 2008, the ACE evaluation will not only include within-document entity and relation detection and disambiguation for Arabic and English, but also include cross-document global integration and reconciliation of information.

For cross-document and cross-language entity and relation disambiguation tasks, system output will be evaluated only for person (PER) and organization (ORG), and only for documents in which the "target entities" are mentioned by name. Target entities refer to a carefully selected list of pre-defined entities of interest.

LDC/sponsors have identified 50 target entities, and a 400 document corpus, which have the following features:

- each entity should be mentioned 5-100 times in the corpus
- some examples of the alias issue, for instance, <u>Carlos the Jackal aka Ilich</u> <u>Ramírez Sánchez</u>
- some examples of spelling and orthographic variations, for instance, <u>Khadafi example</u>
- some examples of the confusable entities issues, for instance, the six <u>Michael Jordans</u>
- some entities that appear in both English and Arabic data sets

The purpose of the Cross-Document (XDOC) task is to globally coreference these 50 ACE entities, and all ACE relations which contain them, over the 400 document corpus. These 400 documents will have been previously ACE annotated, so all ACE entities and relations will already be coreferenced within the documents.

The XDOC task will be to find the target ACE entities and relations, and coreference them globally, i.e., not within but **between** documents.

The XDOC annotation tool, called Callisto, will allow us to do this effectively, by allowing searching and global co-reference of ACE annotations. The procedure for this annotation is outlined below.

#### 1 Entity and Relation Cross-Document Coreference

The three main tasks for XDOC annotation are:

- Global Entity co-reference on 50 target entities
- Global Relation co-reference on relations containing 50 target entities
- Global Entity co-reference on PER and ORG arguments in those relations that are not one of the 50 target entities

We will also do Global Entity co-reference (GEDR) on the remaining PER and ORG entities in a post-hoc process a few weeks after the main ACE 2008 delivery has been completed.

#### 1.1 Global Entity Co-reference on 50 Target Entities

The procedure for Global Entity co-reference on the 50 target entities is:

- Open the Callisto annotation tool
- Open name profile/list of candidate files for your assigned name (AWS will do this)
- Open an anchor file for your assigned name
- Find your assigned name in the anchor file.
- Search the 400-document ACE annotated corpus for every possible name variant of that entity.
- For each search result, view the document to see if that entity is referring to your target entity.
- If the search result entity is referring to your target entity, link it to your target entity
  - Watch out for confusable entities of your target entity (a different entity with the same name). Do not link a confusable entity to your target entity.
- If you find a confusable entity that falls under the criteria in section 1.1.2, repeat from step 3.
- If you find a mention of your assigned name that is missed, wrongly tagged, or not co-referenced with other mentions in that ACE annotated file, **do not** globally co-reference it. Please mark that name "Broken" in AWS, and include in the comment field a description of the error and the file name it occurs in.

#### 1.1.1 Entity Profiles

During the Structured Data Exploration task, "Entity profiles" were compiled for the 50 target entities. These entity profiles are intended to be a complete list of all names which can refer to that entity, so the annotator can find all the mentions of that entity in the corpus. It is possible some variants in the corpus are not in your entity profile. If you find a name mention not in your entity profile, make sure to do a search for this variant as well.

The entity profiles consist of:

- Entity Handle (main name)
- Aliases and Nicknames
- Transliterations
- Orthographic Variants
- Knowledge/Fact about Entity

You may need to search online for more information about your assigned name if your entity profile seems sparse.

#### **1.1.2** Confusable Entities

The Knowledge/Fact is given to help you differentiate each entity from other similarly named entities.

If you find your **entity handle** refers to multiple entities in the corpus, you must globally co-reference all of the entities during that assignment, **even if the entity profile fact refers to only one of the entities.** 

If the entity handle was "George Bush", you would need to perform GEDR on both "George W. Bush" and "George H. W. Bush" entities, even if the fact was "the 43<sup>rd</sup> U.S. President".

However, if the *entity handle* was "George W. Bush", you would **not** have to perform GEDR on both George Bushes, because that name only refers to one of them.

If you are not sure whether you need to annotate multiple entities for an assigned name, ask your supervisor.

#### 2 Callisto Annotation Tool

#### 2.1 Starting Callisto

Callisto can be run by opening a terminal and running the following command:

/ldc/bin/callisto.sh &

#### 2.2 Setting up Callisto

In order to work properly, you must configure Callisto for your account.

Run Callisto for the first time with the above command, and then close it. This will create a text file called "callisto.prefs" in a subdirectory of your home directory called ".callisto".

Before starting annotation, you will need to close the tool, and edit the callisto.prefs file by opening a new terminal window and running the following command:

emacs .callisto/callisto.prefs &

This opens the callisto.prefs file in the emacs text editor.

Add the following line into the file:

jawbFrame.showEdnaImport=true

Go to "File" and click on "Save current buffer", and close this file. Then you can open Callisto again and start annotation.

#### 2.3 Importing files from Edith database

Callisto requires that you start XDOC annotation with a file mentioning your assigned entity. That file serves as an anchor to which you will link the other mentions in the corpus.

Your assigned entity profile has a list of candidate anchor files at the bottom. Go to the File menu and select "Import from server":

ి <mark>C</mark> a	allisto			
<u>F</u> ile	<u>E</u> dit	F <u>o</u> rmat	<u>T</u> ools	He
<u>N</u> ev	N		Ctrl-	N
<u>O</u> pe	en		Ctrl-	0
<u>C</u> lo:	se		Ctrl-	W
<u>N</u> ev	v Colle	ection		
<u>C</u> or	ntinue	Collection	Ctrl-	L
<u>S</u> av	/e		Ctrl-	s
Sav	/e As			
Dor	1e			
Dor	1e & N	le <u>x</u> t	Ctrl-	D
<u>I</u> mp	ort			
Exp	ort			
Imp	ort fro	om Ser <u>v</u> er	r	
Doc	cumen	nt		•
Hist	tory			•
<u>N</u> ev	w Win	dow		
Clo	se <u>W</u> ii	ndow		
E <u>x</u> it	t			

the Configurable Annotation Workbench

Input the Server URL. The English Server URL is listed below. The Arabic URL will be provided in email or on the wiki.

English Server URL:

Server Login         Server URL:       http://sark.mitre.org:8989/edith/edith.do         User:       day         Password:       •••••         Get List       Get List         Last listing time:       Never         Files listed:       0         Files listed:       Show All Users' File         Current Selection:       None         File       Status         Locked By       Locked Since	Server Login         Server URL:         http://sark.mitre.org:8989/edith/edith.do         User:       day         Password:       eevee         Itast listing time:       Never         Files listed:       0         Files listed:       0         Show All Users' File         Current Selection:       None         File       Status         Locked By       Locked Since					
Server URL:       http://sark.mitre.org:8989/edith/edith.do         User:       day         Password:       •••••         Get List         Last listing time:       Never         Files listed:       0         Files listed:       Show All Users' File         Current Selection:       None         File       Status         Locked By       Locked Since	Server URL: http://sark.mitre.org:8989/edith/edith.do   User: day   Password: ••••••	Server Login				
User:       day         Password:       Get List         Last listing time:       Never         Files listed:       0         Files listed:       0         Show All Users' File         Current Selection:       None         File       Status       Locked By         Locked Since       Image: Comparison of the status       Image: Comparison of the status	User:       day         Password:       Get List         Get List       Get List         Last listing time:       Never         Files listed:       0         File       Status         Locked By       Locked Since	Server URL: http://sark.m	nitre.org:8989/edith/edit	th.do		
Password:       Get List         Last listing time:       Never         Files listed:       0         Files listed:       0         Show All Users' File         Current Selection:       None         File       Status       Locked By         Locked Since       1	Password:       Get List         Last listing time:       Never         Files listed:       0         File       Show All Users' File         File       Status         Locked By       Locked Since	<u>U</u> ser: day				
Last listing time:       Never       Files listed:       0       Files locked by your         Show All Users' File         Current Selection:       None         File       Status       Locked By       Locked Since         File       Status       Locked By       Locked Since	Last listing time:       Never       Files listed:       Files locked by you:         Show <u>A</u> ll Users' Fil         Current Selection:       None         File       Status       Locked By       Locked Since	Password:			Get	_ist
File     Status     Locked By     Locked Since	File       Status       Locked By       Locked Since	Last listing time: Neve	r	Files li	sted: 0 Files locked by	you: s' File
File Status Locked By Locked Since	File     Status     Locked By     Locked Since	Current Selection: None	3			
		File	Status	Locked By	Locked Since	
		File	Status	Locked By	Locked Since	
		File	Status	Locked By	Locked Since	
		File	Status	Locked By	Locked Since	
		File	Status	Locked By	Locked Since	
		File	Status	Locked By	Locked Since	
		File	Status	Locked By	Locked Since	

http://blooie.ldc.upenn.edu:8080/edith/edith.do

Enter your XDOC username/password and hit "Get List"

Within the same session, your XDOC password will be remembered, so next time you can just hit "Get List" again.

"Get List" will bring up the list of all the files in the corpus that can be imported:

Server Login			
Server URL: http://sark.mitre.org:8989/edith/edith	n.do		
User: day			
Password:			G
Last listing time: Mon Apr 07 17:08:30 EDT 20	08	Files listed:	549 Files locked
Current Selection:			
File	Status /	Locked By	Locked Since
AFP_ENG_20030304.0250.sgm	DONE		
CNN_ENG_20030526_133535.4.sgm	DONE		
APW_ENG_20030502.0686.sgm	9/40		
AFP_ENG_20030320.0722.sgm	8/38		
AFP_ENG_20030323.0020.sgm	5/39		
AFP_ENG_20030327.0224.sgm	3/35		
AFP_ENG_20030305.0918.sgm	3/33		
AFP_ENG_20030314.0238.sgm	3/27		
AFP_ENG_20030311.0491.sgm	2/9		
APW_ENG_20030519.0548.sgm	2/7		
APW_ENG_20030318.0689.sgm	2/39		
AFP_ENG_20030413.0098.sgm	2/23		
AFP_ENG_20030508.0118.sgm	1/7		
CNN_ENG_20030624_082841.12.sgm	1/24		
AFP_ENG_20030319.0879.sgm	1/21		
APW_ENG_20030311.0775.sgm	1/21		
AFP_ENG_20030401.0476.sgm	1/15		
AFP_ENG_20030427.0118.sgm	1/14		
	1/14		
AFP_ENG_20030430.0075.sgm		1	
AFP_ENG_20030430.0075.sgm BACONSREBELLION_20050222.0817.sgm	1/14		

Click on any column header to sort. The "Status" column shows how many entities have been annotated out of the number of entities in that file.

Find the file you want, then select that file's row and hit "Import"

If you don't have any candidate files available for importing, tell your supervisor immediately.

#### 2.4 Annotating Files

When you import a file, it will open the annotation function window. You will see the text of the anchor file on the left, and the Entities tab at the bottom, which contains a table of all the ACE annotated entities in the document.

Callisto - CNN_CF_20030303.1900.02.apf.xml	Collator CNI (£ 20030031990.07.2.0fsm)							
File Edit Format Tools Panes Help								
59 Main					Gr Search Results Gr Linked Embles			
							Sourch Docute Summary	
CNN_CF_20030303.1900.02							Search results Summary	
STORY				-	Sort By: Highes	t Ranked 💌		
2003-03-03T19:00:00-05:00						External Ref	Document Entity ID	Text
				-				
New Questions About Attacking Iraq; Is Torturing	g Terrorists Necessary?							
NOVAK								
Welcome back Orders went out today to deploy	17,000 US Army soldiers							
in the Persian Guir region. The army's entire firs	d Calvary division							
already in the region. We're talking about possib	silities of full scale							
war with former Congressman Tom Andrews,	Democration <mark>Maine.</mark> He's now							
pational director of Win Without War, and former	Congressman Bob Dornan,							
Republican of California.								
BEGALA				_				
Roll one of the reasons I think so many America	and are worried about this			•				
🗗 Qurrent Entity					🗗 Search Detail	5		
Current Entity ID(s):								
Show Disambiguated View					View In Context	NA		
Search Canonical Only					Type-Subtype: /	νA		
a search continued only					Name(s): M/A			
12 Tano Subtano					Pronoun(s): N/A			
Type:subype					Nominal(s): N/A			
+ Draman (a)					Comments: N/A Polationchine: M	*		
Pronoun(s)					readourships. 24			
I Nominai(s)					Link result to se	lected entity Link		
Comments								
Relationships								
	Sea	irch						
G Entities G Entity Mentions G Rel	lation Mentions 🗗 Events	Event Mentions     Ø Values	Timex2					
External Ref	Document Entity ID	Primary Reference	Mentions		lype	Subtype	Class	Comments
	E1	Tom Andrews		11 PER		Individual	SPC	
	E2 E4	Our Search Repository fo	r this Entity	27 GPE		Nation	SPC	
	E6	Congressma Populate Search Spe	c with this Entity	13 PER		Individual	SPC	
noncanonical/1150310670782	E7	Min Without V Disambiguate (Uniqu	e in Collection)	1 ORG		Non-Governmental State-or-Province	SPC	
noncanon/cdr1100310010702	E11	owneedd -		4 PER		Individual	SPC	
	E12	Americans		1 PER		Group	USP	
	E13 E14	people		1 PER		Celectel	05P 9PC	
P								

You can sort by type, primary reference, name, etc., by clicking on the header of each column.

Entities in red text have not yet been given an External reference (global) ID.

If your assigned PER or ORG is green, someone else has already linked that entity. In that case, mark the name "Broken" and inform your supervisor.

If your PER or ORG name is in red, double click on it:

Sea	arch				
Entity Mentions	Relation Mentions	Events	Event Mentions	Values	Timex2
Document Entity	Primary Refere	Mentions	Туре	Subtype	Class
E16	members	2	PER	Group	USP
E19	LONDON	1	GPE	Population-Center	SPC
E21	Taliban	1	ORG	Non-Governmental	SPC
E22	Afghanistan	3	GPE	Nation	SPC
E23	Rumsfeld	9	PER	Individual	SPC
				Font: 12pt. D	efault Charset: UTF

ACE08\_XDOC\_1.6.docThen right-click:

Entity Mentions	P Relation Men	tions	D Events	Event Mention	ns	🗗 Values	Ø	Timex2	
Document Entity	Primary Refere		Mentions	Туре		Subtype		Cla	ass
E16	members		2	2 PER		oup		USP	
E19	LONDON		1	ODE	Po	pulation-Cente	er	SPC	
E21	Taliban	Sear	rch Repository for th	nis Entity	No	n-Governmen	tal	SPC	
E22	Afghanistan	Pop	ulate Search Spec w	with this Entity	Na	tion		SPC	
E23 Rumsfeld						lividual		SPC	
Disambiguate (Unique in Collection)									
Font: 12pt. Default Charset: U							set: U		

Select "Populate Search Spec with this Entity". This will automatically fill the search term field with the NAM mentions, and let you manually add more terms before executing the search.

#### 2.4.1 Manual Search Terms

You can manually add in additional search terms into the text box below the NAM mentions. Searching for part of a name will find any names that contain that string. For instance, a search on "dou" will find "Mohammed Al-Douri".

Use this function to search for all variants in your entity's name profile. Make sure to do a manual search for the NAM variants from your anchor file, because the automatic search function can miss some mentions.

If you want to search not just on PER or ORG NAMs, you need to unselect "Type/Subtype", or select the type you want to include.

These searches can be useful to find names spelled completely wrong, like "leo clinhinghof-hoffer" for "Leon Klinghoffer", which can occur in transcribed speech.

Leon Klinghoffer was often referred to as an "American tourist", so you could search for "tourist" under NOM to find articles that didn't match your NAM searches.

When you have entered the terms you wanted, hit "Search".

					102
rgrmat Tools Falles Helb					
	😨 Search Results	Linked Entities			
	A		Search Results Summary		
20030303.1900.02					
	Sort By: Highest Ra	anked			
3119:00:00-05:00		External Ref	Document Entity ID	Text	
	571569	g-5)	/1569	Lucerne in Maine, Lucerne-in-Maine	-
	18551251	g-11	3651251	Maine d'Euche	
lions About Attacking Iraq; Is Torturing Terrorists Necessary?	18551253	9-11	3551253	Maine-Chevalier	
	10550225	Q-11	1000220	Le Maine	
	18558227	g-11	8558220	Le Maine	
	18558228	0-11	8558228	Le Maine	
	18562962	g-11	8562962	La Maine	
ack. Orders went out today to deploy 17,000 U.S. Army soldiers	253730	g-2	53730	Maine Prairie	
ian Gulf region. The army's entire first Calvary division	430310	g-4	30310	Maine, Township of	
ort Hood, <mark>Texas,</mark> would join the quarter million U.S. forces	430311	g-4	30311	Maine, Township of	
he region. We're talking about possibilities of full scale	469332	g-4	39332	Maine, Township of	
mer Congressman Tom Andrews, (Permoration Maine, He's now	6658/8	g-6i	10878	maine, Township of	
ector of Win Without War, and former Congressman Bob Doman,	950162	g-91	30102	East Maine Maine Town of	
n of California.	1584630	0-91	584630	Maine Town of	
	1584631	0-1	584631	Maine Town of	
	1780787	0-1	780787	Maine. State of	
	1038246	0-11	038246	Maine, Township of	
f the reasons I think on many Americano are worried about this	30336767	ا م ما	3005767	Maina Duidan	
t Potty	Search Details				
ity ID(s): CNN_CF_20030303.1900.02-E2					
and batemidme size	View In Contexts	a 4790707			
aannagaaca voor	VIEW III COIREAL	g-1700707			
Canonical Only	Time-Subtime: GPE-	countries. 2nd order divisions			
	type completer at a				
Public and	Name(s):				
Zmukho	Maine, State of				
x"GPE"	Pronoun(s): N/A				
type: "State-or-Province"	Nominal(s): N/A				
	Comments: AVA				
(5)	Relationships: N/A				
10 <sup>2</sup>	Link result to select	ted entity Link			
pun(s)					
nal(s)					
	×				
+(name*maine*)-eldicnn_cf_20030331900.02-e2					
search					
- Leanning the second states and the second states and the second states and					
er cristy microsina er nortaon meritions er events er event meritions er values er Timex2					
External Ref Document Entity ID Primary Reference Mentions	Туре	Subtype	Class	Comments	
EI Tom Andrews 11/PER	Ind	lividual	SPC		-
EZ Maine 1 GPE	Sti	tie-or-Province	SPC SPC		-
E4 007 2/10PE	Na Ind	non Beideal	lepc		
E7 Win Wink War 10PG	No	n-Governmental	SPC		
	95	ate-or-Province	SPC		
California La California La California	Ind	lividual	SPC		
E11 California 2017E	0.0	oup	USP		
E10 California 2007E E11 I 4/PER E12 Americans 1/PER	ION				
An 1103300/07/02         E10         L_amonta         2/07E           E11         I         4/PER           E12         Americas         1/PER           E13         people         1/PER	Gri	oup	USP		

🗗 Sea <u>r</u> ch Results 🔀	Linked Entities						
	Search Results Summary						
Sort By: Highest Ranked							
External Ref	Document Entity ID Text						
	AFP_ENG_20030305.091 London						
	AFP_ENG_20030401.047 LONDON						
	AFP_ENG_20030425.040 London						
	AFP_ENG_20030508.011 LONDON						
	AFP_ENG_20030509.034 LONDON						
	AFP_ENG_20030514.070 London						
	AGGRESSIVEVOICEDAILY London						
	APW_ENG_20030318.068 London						
	APW_ENG_20030325.078London						
	APW ENG 20030415.074 London						
₽ Search <u>D</u> etails ⊠							
View In Context: AFP_ENG_20030305.0918-E117							
Type-Subtype: GPE-Popula	ation-Center						
Name(s): London							
Pronoun(s): N/A							
Nominal(s): N/A							
Comments: N/A							
Relationshins: N/A							
reaction ships. W/A							
Link result to selected entity Link							

If an entity in your search result has an External Ref ID, it has already been globally co-referenced.

You can see from the Text column which mentions were a match for your search. 12

Select the row and click on "View in Context" to see the entity mentions in the file context. They will be highlighted in red text.

If the entities are the same, close the document and select "Link" to link that entity to your "anchor" entity. The tool will generate an External Ref (Global) ID for that entity the first time it is linked.



When it already has an External Ref (XRef) ID, it will show you that ID and the other entities already belonging to it.

🖧 Callisto	- AFP_ENG_20030320.0722.apf.xml*		
<u>File Edit</u>	Format Tools Views Help		
🖙 Main			× es
CAIRO, N	Entities	Detail	Search Result
Tens of 1 East LLT Colling of aillies. Lebanon the US-le warned I benefit or Protests Hussein as in Am  Current E Current E Current Show Searc		View In Context: N/A Type-Subtype: N/A Name(s): N/A Pronoun(s): N/A Comments: N/A Relationships: N/A	ent Entity ID 307-E15 372-E5 372-E64 0555-E1 0314-E41 0775-E35 069-E17 039-E14 030-E14 062-E14 0662-E14 0662-E14 0662-E14
		n aq	

If you don't want to link the entity, you can just close this window.

If you do select Link, that document result will be moved from the Search results tab to the "Linked Entities" tab.

	🗗 Search Results 🛛 🗗 Linked Entit	ies 🛛	
	Linked	Entities Summary	
=	Sort By: Highest Ranked 💌		
	External Ref Doc	cument Entity ID Text	
•		0_20030303.0316 E01001	
	₽ Search Details X		
			<b>A</b>

If you link an entity accidentally, you can select this row and hit "Unlink".

When you are confident there are no more mentions of your target entity in the corpus, then you are done with your collection.

If you did not find any other mentions of your target entity in the corpus, rightclick on your entity and select the "Disambiguate in the Corpus" option. This will give the entity an External Ref ID, without linking any other mentions to it.

Now select File -> Close to close the file. The tool saves automatically, so you never need to use "Save".

After you select "Close", it will ask you if you want to release the lock on this file. Make sure to click "Yes" to release the file if you are done with that nACE08\_XDOC\_1.6.docame.

To get your next assignment, hit "Stop" on AWS as normal, and say Yes to a new name assignment.

Then go to "File > Import from Server" again in Callisto, hit "Get List", and find the anchor file for the new name.

#### 2.5 Entity Linking errors

Due to the fact that multiple annotators can simultaneously search and annotate with the Callisto/Edna task, it is possible that two or more annotators might inadvertently attempt to link to the same entity at the "same time." If Callisto suspects this behavior is happening, this error message will appear.



If this error message happens to you, report it to your supervisor.

You may keep working on the name as long as your mention is correctly tagged in the ACE file. If it's not, mark the name as "Broken" and move on.