

**AFRL-RI-RS-TR-2008-125**  
**Final Technical Report**  
**April 2008**



## **TERRORISM KNOWLEDGE BASE (TKB)**

**Cycorp, Inc.**

*APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.*

**STINFO COPY**

**AIR FORCE RESEARCH LABORATORY**  
**INFORMATION DIRECTORATE**  
**ROME RESEARCH SITE**  
**ROME, NEW YORK**

## NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the Air Force Research Laboratory Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RI-RS-TR-2008-125 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/s/

/s/

NANCY A. ROBERTS  
Work Unit Manager

JOSEPH CAMERA, Chief  
Information & Intelligence Exploitation Division  
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

**REPORT DOCUMENTATION PAGE***Form Approved*  
**OMB No. 0704-0188**

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> APR 08		<b>2. REPORT TYPE</b> Final		<b>3. DATES COVERED (From - To)</b> Nov 02 – Nov 07	
<b>4. TITLE AND SUBTITLE</b>  TERRORISM KNOWLEDGE BASE (TKB)				<b>5a. CONTRACT NUMBER</b> F30602-03-C-0007	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b> 31011G	
<b>6. AUTHOR(S)</b>  Douglas Lenat and Chris Deaton				<b>5d. PROJECT NUMBER</b> GENO	
				<b>5e. TASK NUMBER</b> A0	
				<b>5f. WORK UNIT NUMBER</b> 05	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Cycorp, Inc. 7718 Wood Hollow Dr, Ste 250 Austin TX 78731-1645				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>  AFRL/RIED 525 Brooks Rd Rome NY 13441-4505				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSORING/MONITORING AGENCY REPORT NUMBER</b> AFRL-RI-RS-TR-2008-125	
<b>12. DISTRIBUTION AVAILABILITY STATEMENT</b> <i>APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA# WPAFB -08-0448</i>					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> The objective of this project was to support intelligence analysts by developing a comprehensive Terrorism Knowledge Base (TKB) which included information about terrorist events and terrorist groups and their members and activities, as well as information captured by the analyst's use of the tool. Using that knowledge base, plus the knowledge base and inference engine of our company's Cyc(r) technology, the TKB was to exhibit sophisticated reasoning using domain knowledge, externally-stored data, common sense knowledge and knowledge about what the analyst has considered relevant or irrelevant and templated question-answering with explanations. Analysts were to be able to use the TKB to pose terrorism-related queries, and to help them derive answers to those questions, integrate data, correlate observations, compose explanations, and, in general, augment their ability to effectively complete the reasoning tasks that they need to perform.					
<b>15. SUBJECT TERMS</b> Terrorism knowledge base, representation and reasoning, intelligence analyst tool, question answering, query and knowledge entry, terrorist events and terrorist groups					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>  UU	<b>18. NUMBER OF PAGES</b>  72	<b>19a. NAME OF RESPONSIBLE PERSON</b> Nancy A. Roberts
<b>a. REPORT</b> U	<b>b. ABSTRACT</b> U	<b>c. THIS PAGE</b> U			<b>19b. TELEPHONE NUMBER (Include area code)</b> N/A

# Table of Contents

<b>1</b>	<b>OVERVIEW OF THE TKB PROJECT .....</b>	<b>1</b>
<b>2</b>	<b>CAPABILITIES OF TKB AND CAE .....</b>	<b>3</b>
<b>3</b>	<b>TECHNICAL ACCOMPLISHMENTS.....</b>	<b>13</b>
3.1	SPATIO-TEMPORAL REASONING IN THE TKB .....	16
3.2	EXPERIMENTS .....	19
3.2.1	<i>Project Arete</i> .....	19
3.2.2	<i>Project Leviathan</i> .....	21
3.3	THE DEVELOPMENT OF THE FACT ENTRY TOOL (FET) AND THE CYC ANALYTIC ENVIRONMENT (CAE) .....	23
3.3.1	<i>Improvements to natural language understanding</i> .....	25
3.3.2	<i>The CAE and the Query Library</i> .....	28
3.3.3	<i>Other Important CAE functionality</i> .....	31
<b>4</b>	<b>COLLABORATION.....</b>	<b>36</b>
<b>5</b>	<b>OUTSIDE EXPERT EVALUATION .....</b>	<b>38</b>
<b>6</b>	<b>CONCLUSION .....</b>	<b>40</b>
<b>7</b>	<b>APPENDIX I: INITIAL TERRORISM REPRESENTATION SCHEMA.....</b>	<b>41</b>
<b>8</b>	<b>APPENDIX II: LISTING OF ENGLISH GLOSSES OF REPRESENTED TERRORISM DOMAIN QUERIES.....</b>	<b>50</b>
<b>9</b>	<b>APPENDIX III: CHARACTERIZATION OF TKB CONTENT.....</b>	<b>64</b>

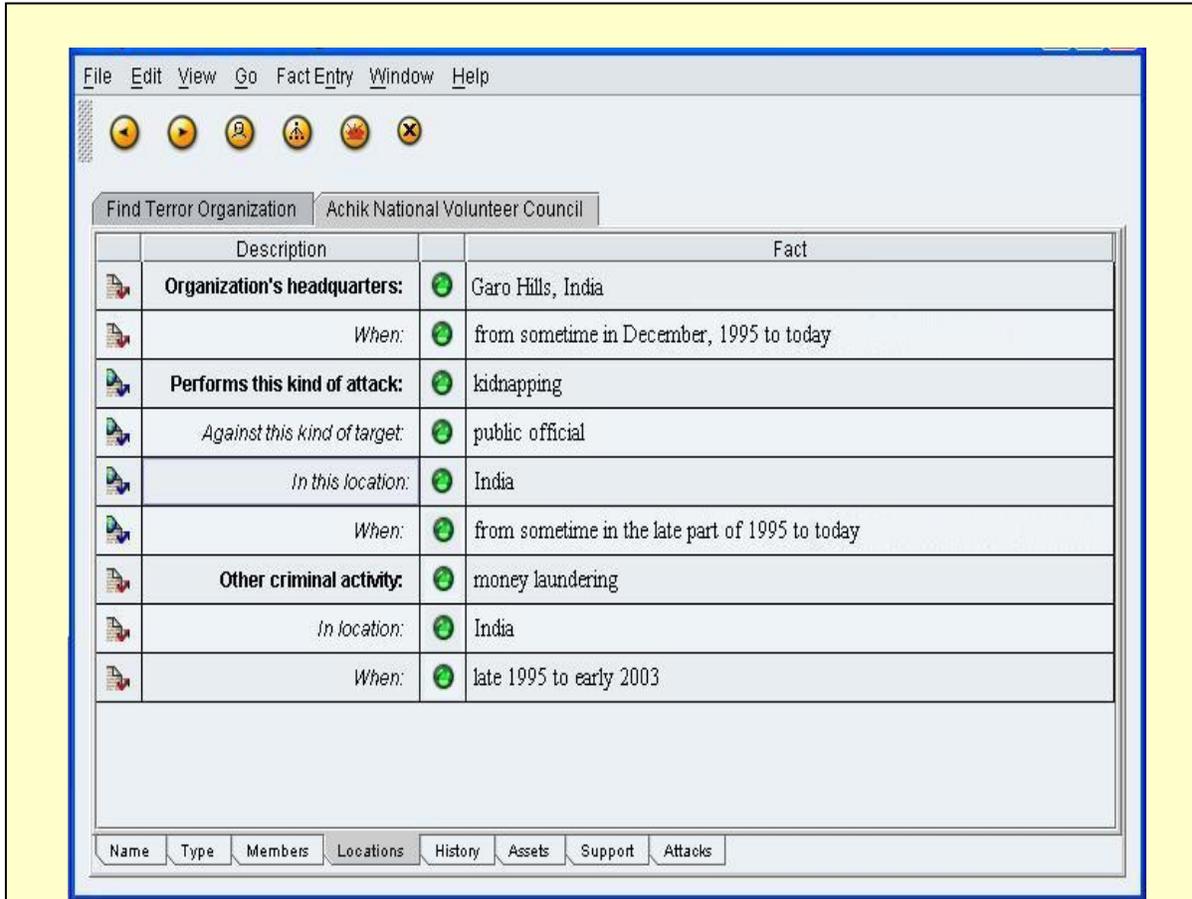
# 1 Overview of the TKB project

The objective of this project was to support intelligence analysts by developing a comprehensive Terrorism Knowledge Base (TKB) which included information about terrorist events and terrorist groups and their members and activities, as well as information captured by the analyst's use of the tool. Using that knowledge base, plus the knowledge base and inference engine of our company's Cyc® technology, the TKB was to exhibit sophisticated reasoning using domain knowledge, externally-stored data, common sense knowledge and knowledge about what the analyst has considered relevant or irrelevant and templated question-answering with explanations. Analysts were to be able to use the TKB to pose terrorism-related queries, and to help them derive answers to those questions, integrate data, correlate observations, compose explanations, and, in general, augment their ability to effectively complete the reasoning tasks that they need to perform.

The TKB is an augmentation of the existing Cyc® Knowledge Base (Cyc KB), which has been under intensive construction for the past 23 years. The Cyc KB contains a formalized representation of large tracts of consensus reality, encoded in hundreds of thousands of terms and millions of hand-axiomatized assertions organized into hundreds of contexts (called "microtheories"). Most of the current content of the Cyc KB consists of general facts about kinds of everyday objects and activities. It also contains "almanac-style" facts about individual countries, ethnic groups and organizations. Prior to launching the development of the TKB, the Cyc KB already had some knowledge relevant to terrorist activity, such as knowledge about geopolitical events, WMD, biowarfare, etc., from various specialized previous U.S. Government projects using it (and, the process, expanding it) that had been performed by Cycorp.

Indeed, that was the essence of the motivation behind TKB: the United States did not have a comprehensive knowledge base of information about terrorists, terrorist groups, and terrorist events. There are a plethora of scattered fragmentary slivers of such a knowledge base in existence – e.g., databases such as the PGIS and MIPT efforts, which contain a dozen or so structured fields about each entity and then "bottom out" in English paragraphs about the individuals, groups, and events; or the comprehensive Iraq terrorism DB, which by definition has a very narrow scope. Now, it has one such KB: the TKB. As a KB, rather than just a DB, the TKB can be used for deductive inference, for helping analysts pose complex *ad hoc* queries, and reason logically to answer them.

The TKB effort has so far added to the Cyc KB knowledge of over thirty-seven hundred individual terrorists, over one thousand different terrorist groups, and over fourteen thousand terrorist attacks. The representations of these individuals, groups and events are involved in over three hundred thousand assertions such as "Xavier Djaffor participated in the Jihad from 1996 to 2000" and "Lashkar-e-Taiba is an Islamic terror group founded in 1990". These terrorism-specific assertions have been acquired via a knowledge entry effort that involved representing facts from websites in a structured format, mapping existing databases and spreadsheets into our representation schema and manual knowledge entry using an application we created for this purpose, called the "Fact Entry Tool" or "FET":



**The FET:** This is a screen shot of the “Fact Entry Tool” (FET), which is the primary knowledge-entry tool that terrorism experts use to populate the TKB. Subject-matter experts reading wire service reports, newspaper articles, etc., record information in the fields of the FET, which operates very much like a web form. The FET user looks up or creates some particular individual – a terrorist, terrorist attack, or terrorist group – and then enters information about that individual by filling in particular FET fields. The strings to the left of the green dots in the interface are field headings, indicating what kind of knowledge should be entered to the right of the green dots. In some cases values for the fields can be selected by choosing them from a drop-down menu. The FET user can type ordinary English into the fields, and the system will parse that English text into a representation of the proper logical form. For example, if the user is prompted to enter the name of a person, then if a representation of the individual already exists in the TKB, the system parses their name to the unique TKB knowledge structure that represents them. If a representation of the individual does not already exist in the TKB, then the new individual is created. Lightly trained subject-matter experts can enter knowledge at rates of over 100 assertions per hour using the FET.

Each assertion has its source (such as a web page, an expert, a newspaper article, etc.) associated with it, and the source itself is represented as a first class object with assertions describing e.g. its name, date and place of issue, and its author or publisher (if relevant). Further, we keep track of which expert entered which data from each source, and when. Indeed, different experts may enter conflicting knowledge in different microtheories. A subject-matter expert working from open source data has entered every fact that has been represented thus far in the TKB. Querying and knowledge entry are achieved via access

to an interactive graphical and text-forms user interface developed during this project called the Cyc Analytic Environment (CAE).

## 2 Capabilities of TKB and CAE

Every fact represented in the TKB (and, indeed, every fact in the Cyc KB) is codified in CycL, which is a form of higher-order predicate logic. The Cyc inference engine is optimized for reasoning over assertions written in CycL. The inference engine consists of a growing regiment of over 1000 special-case reasoning modules as well as a general theorem prover. These “Heuristic level” (HL) modules are efficient implementations of particular patterns of reasoning, such as a technique for calculating the transitive closure of a transitive binary predicate, without having to resort to general theorem proving. Certain specialized modules enable TKB users to receive answers to queries based not merely on a particular syntactic form, but also on the logical properties of the relations involved in the query. For example, a query for all the individuals known to be answerable to Hezbollah during 1998 will return not only all those explicitly asserted to be members (via the CycL predicate “hasMembers”) during that time frame, but will also return all those individuals that can be *deduced* to be answerable based on other assertions (e.g., they were known to lead some group which, in hindsight, we now realize was covertly an arm of Hezbollah.)

The advantages to having information pertaining to terrorism represented in this structured fashion are numerous. With this information captured in CycL, the Cyc inference engine can use it to compute very quickly what might take a human a non-trivial amount of time to calculate. A good example of this that arose recently in response to a user’s query is asking the TKB to calculate the percentage of Hamas attacks between June 1, 2002 and May 15, 2004 that fall into a number of different classes – the percentage that are bombings, the percentage that are homicides, etc. The advantage of the TKB over a standard structured database, in this case, is that Cyc knows that any attack that results in the death of a civilian is a homicide. So, even if the attack is classified explicitly in the TKB as a bombing, so long as Cyc knows that at least one civilian was killed as a result of the attack, then it knows to count that attack as a homicide when calculating the percentage of Hamas attacks that are homicides. The reasoning often involves multiple sources. E.g., Cyc has information about the Khobar Towers bombing entered from two sources, one of which says that there were 19 casualties, and one of which says that 19 U.S. soldiers were killed in the attack. From these, Cyc can conclude that there were no civilian casualties in that attack, that all the casualties were Americans, and so on.

In this next section we describe, step by step, how a moderately-trained analyst – someone who has been given 3 days’ training in the currently running version of our tool, the Cyc Analytic Environment or “CAE” – conveys to the TKB the information in a terrorist incident news report (in this case a CNN report about a car bombing).

The context is that the analyst has been tasked with entering information about Imad F. Mughniyah. Let’s suppose they do a Google-like keyword search through classified and unclassified sources, and among other “hits” they come across the following CNN report:

**CNN online news report (*fictional*)**  
**British Embassy in Rio de Janeiro Bombed.**  
**May 1, 2004**

A car bomb exploded earlier today at the British Embassy in Rio de Janeiro. Three German tourists were killed by the blast along with a local boy who was skating on the street in front of the embassy building when the bomb detonated. Several embassy security guards were wounded in the attack. The embassy itself was damaged but not seriously. The police suspect that Imad F. Mughniyah was involved in the attack.

How exactly do they enter that information into the TKB, in a way that the machine will actually understand (not just memorize the words)? Here is the step-by-step process:

(1) The user starts up the Cyc Analytic Environment. They open a Fact Entry Template by going to the “Tools” menu at the top of the window and choosing “Enter Facts About Existing Terrorist”. Once loaded the “Find Terrorist” tab will open.

(2) The user types “Mughniyah” in the Last Name Field. A colored circle to the left of the field turns green to indicate the entry has been understood. The system finds Mughniyah and displays all the information about him that the TKB already knows. On the next page is a screenshot of about 20% of that scrollable “Fact Sheet” on Mughniyah. Each of the sentences there is generated automatically, by Cyc, from the underlying logical assertions that are already known about that individual. These are not beautiful prose, but they are understandable by the analyst, especially after they have seen several similar fact sheets.

# Imad Fayeز Mugniyah



Two small portions of the TKB Fact Sheet on Imad Mugniyah. The scrollable fact sheet contains about 10 times this much information, in toto. Note that each of these English sentences is generated automatically, by Cyc, from the underlying formal assertion which is written in predicate calculus (CycL). The assertion footnotes point to the original source(s) for each fact. Dozens of sources mention this person, hence are integrated into the TKB, thence into the fact sheet.

## Biographical data:

- Imad Fayeز Mugniyah

## Location data:

- Imad Fayeز Mugniyah

- From 1962 to 1982, 1

- From sometime in 19

Fayeز Mugniyah resid

- In 2002, Imad Fayeز

- In the early part of th



- (3 similar sentences)

- In the late part of the

- In the middle part of

[Sudan](#)

- In 1985, Imad Fayeز

- In 1985, Imad Fayeز

- (6 similar sentences)

## Documentation data:

- Note: Tied to Khobar

## Affiliation data:

Imad Fayeز Mugniyah was a naturalized citizen of [Iran](#) in the late part of the 1980's. In 2004, Imad Fayeز Mugniyah had the responsibility of providing intelligence for [Lebanese Hezbollah](#). In June 1985, Imad Fayeز Mugniyah and [Feridun Mehdi Nezhad](#) were accomplices. From sometime in the 1970's to 1982, Imad Fayeز Mugniyah was employed by Salah Khalaf aka Abu Ayad. On May 28, 1986, Imad Fayeز Mugniyah was a key member of [Lebanese Hezbollah](#). Since the middle part of the 1980's, [Hezbollah Operational Security](#) has been headed by Imad Fayeز Mugniyah. [Sheikh Hussein Fadlallah](#) employed Imad Fayeز Mugniyah in the early part of the 1980's. In 2004, Imad Fayeز Mugniyah was a senior member of [Lebanese Hezbollah](#).

- Imad Fayeز Mugniyah is head of security for [Lebanese Hezbollah](#) since June 1982.
- Imad Fayeز Mugniyah is head of security for [Lebanese Hezbollah](#) at 15:22:37, February 17, 2005.
- Imad Fayeز Mugniyah is a citizen by birth of [Lebanon](#).
- Imad Fayeز Mugniyah was [Jihad Shuman's](#) follower in 2001.
- [Sheikh Hussein Fadlallah](#) was Imad Fayeز Mugniyah's follower in the early part of the 1980's.
- [Ahmad Assad Barakat](#) was Imad Fayeز Mugniyah's follower before 21:57:30, August 31, 2004.

- In the early part of the 1980's, Imad Fayeز Mugniyah was the bodyguard of [Sheikh Hussein Fadlallah](#).
- In the early part of the 1980's, Imad Fayeز Mugniyah was the bodyguard of [Yasser Arafat](#).

- Imad Fayeز Mugniyah was a head of operations for [Lebanese Hezbollah](#) before 12:52:11, January 21, 2005.
- Imad Fayeز Mugniyah was a head of operations for [Lebanese Hezbollah](#) from sometime in the 1980's to 14:43:40, August 22, 2004.
- Imad Fayeز Mugniyah was a head of operations for [Lebanese Hezbollah](#) before

(3) Since the user wants to tell the system about a new attack that Mughniyah is a suspect in, they click the “Attacks” tab. This brings up a list of attacks Mughniyah is already known or believed (according to various sources) to have been involved in.

(4) The user clicks on any of the “Participant in Attack” glosses and chooses “Add Similar Entry” from the menu that pops up. This tells the CAE that the user wants to tell it about a new attack that this person may have participated in. At this point a blank set of fields will appear – one that states “Participant in Attack” and another that allows one to enter the “Role Played in Attack” since in general there are many different ways that one could participate (e.g., being the provider of the bomb, being a lookout, being the bomber, luring the target to a particular location, driving the car to the bomber, etc.)

The screenshot shows the Cyc Analytic Environment interface. The title bar reads "Cyc Analytic Environment: Research Imad Fayed Mughniyah". The menu bar includes "File", "Edit", "Tools", "Window", and "Help". Below the menu bar are three icons: a yellow arrow, a yellow arrow with a red 'X', and a purple globe. The main window has a tabbed interface with tabs for "Task Info", "Concepts", "Queries", "Find Terrorist", and "Imad Fayed Mughniyah". The "Find Terrorist" tab is active, displaying a table of attack records.

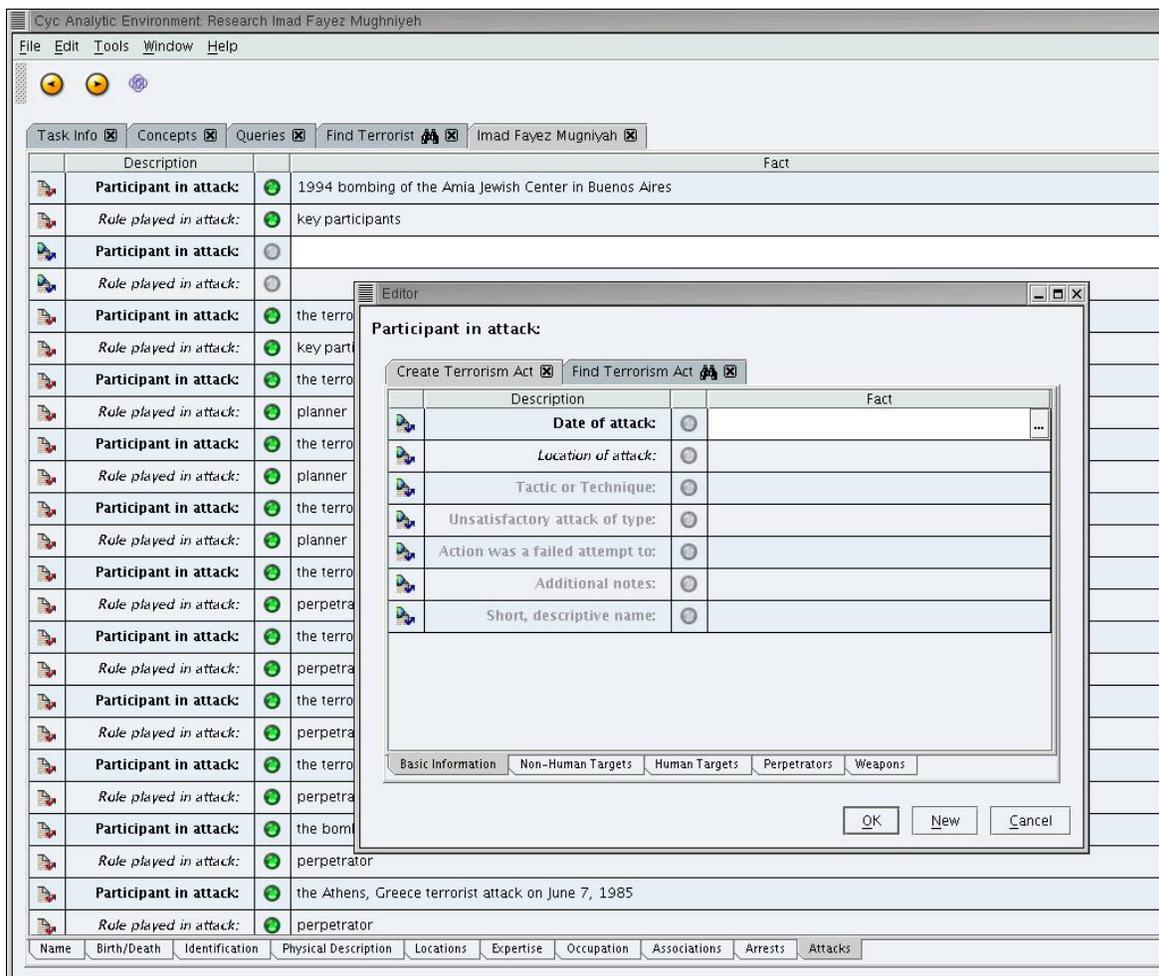
	Description	Fact
	<b>Participant in attack:</b>	1994 bombing of the Amia Jewish Center in Buenos Aires
	<i>Role played in attack:</i>	key participants
	<b>Participant in attack:</b>	the terrorist attack on October 15, 2000 in the United Arab Emirates
	<i>Role played in attack:</i>	key participants
	<b>Participant in attack:</b>	the terrorist attack on July 27, 1994 in London, England
	<i>Role played in attack:</i>	planner
	<b>Participant in attack:</b>	the terrorist attack on July 26, 1994 in London, England
	<i>Role played in attack:</i>	planner
	<b>Participant in attack:</b>	the terrorist attack in Lebanon on October 23, 1983
	<i>Role played in attack:</i>	planner
	<b>Participant in attack:</b>	the terrorist attack in Lebanon on February 17, 1988
	<i>Role played in attack:</i>	perpetrator
	<b>Participant in attack:</b>	the terrorist attack in Lebanon on March 16, 1984
	<i>Role played in attack:</i>	perpetrator
	<b>Participant in attack:</b>	the terrorist attack on April 18, 1983 in Beirut
	<i>Role played in attack:</i>	perpetrator
	<b>Participant in attack:</b>	the terrorist attack on April 18, 1983 in Beirut
	<i>Role played in attack:</i>	perpetrator
	<b>Participant in attack:</b>	the bombing in Beirut on October 23, 1983
	<i>Role played in attack:</i>	perpetrator
	<b>Participant in attack:</b>	the Athens, Greece terrorist attack on June 7, 1985
	<i>Role played in attack:</i>	perpetrator
	<b>Participant in attack:</b>	the terrorist attack on December 3, 1984 in Kuwait
	<i>Role played in attack:</i>	perpetrator

At the bottom of the window, there are several tabs for detailed information: "Name", "Birth/Death", "Identification", "Physical Description", "Locations", "Expertise", "Occupation", "Associations", and "Arrests".

(5) To the left of those empty fields is a source icon. It will have a red “X” to indicate that a source has not yet been selected. The user would click on that source icon to select a source to associate with the information they are entering, i.e., what is the pedigree or provenance of this information? In this case, the provenance is CNN (as of that date/time, since CNN might later alter or retract or contradict that first news report).

(6) A menu of known sources appears. If the source in this case has not previously been represented, then the user can (now or later) go off and tell the CAE about that source. In this case, of course, CNN is well known to the CAE. The user points to the specific CNN article, and clicks a check- box on the bottom left that states default source. This means that, until told otherwise, all the things the user is about to tell the CAE should be assumed to have this very same source.

(7) Now it’s time to actually start to represent information about the attack. The user clicks on the ellipsis ( “...” ) located just to the right of the new “Participant in Attack” field. This brings up the attack template. Here is what the screen looks like at this point:



(8) The user enters the date of the attack in the “Date of Attack” field by typing a date expression in the field. The system understands many different date formats. This is a

good example of where the state of the art of natural language understanding is adequate for the job at hand: almost every short phrase and notation for stating a date is parsable by the CAE (e.g., “last Tuesday”, “early June of this year”, “March 19” etc.) In cases of ambiguity, such as “3/1/05”, the user might be prompted to choose between “March 1, 2005” and “January 3, 2005”, if there were no user model to indicate how they generally type in their dates. Such a model enables “3/1/05” to be parsed as, say, “March 1, 2005”, which is then displayed in that cell – overwriting what they typed – so the user can see if the system has correctly understood the entry or not.

(9) The next field is “location of attack”. The user simply types in “Rio” and the CAE rewrites this to “Rio de Janeiro, Brazil”.

(10) The user doesn’t need to proceed through this field-by-field, exhaustively, or in order. But in this case the next field is relevant, and the user fills it in. That field is “Tactic or Technique”. The user tells the system that the attack was a car bombing by typing “car bombing” into the “Tactic or Technique” field. Instead of typing that in, they could have clicked on the upside-down triangle in the Tactic or Technique field, and a menu of known tactics/techniques would have appeared (including “car bombing” as one choice.) When they type in “car bombing”, the circle next to the field turns green, meaning the information was asserted into the TKB. This is useful in a case like this where the paraphrasing of the meaning is exactly the same as what the user typed in, so the user knows (once the circle turns green) that this is the paraphrase of “car bombing”.

(11) To represent that a British embassy was damaged, the user clicks on the tab at the bottom of the screen that says “Non-Human Targets”. In the field that says “Number of Targets of a Particular Kind” they enter the number “1”. In the “Type of Target” field, they type “British Embassy”. The user could then use the drop down menu In the “Status of Target” field to select “damaged object” to represent the fact that the embassy was damaged (vs. destroyed, etc.). Here is how this looks at that moment:

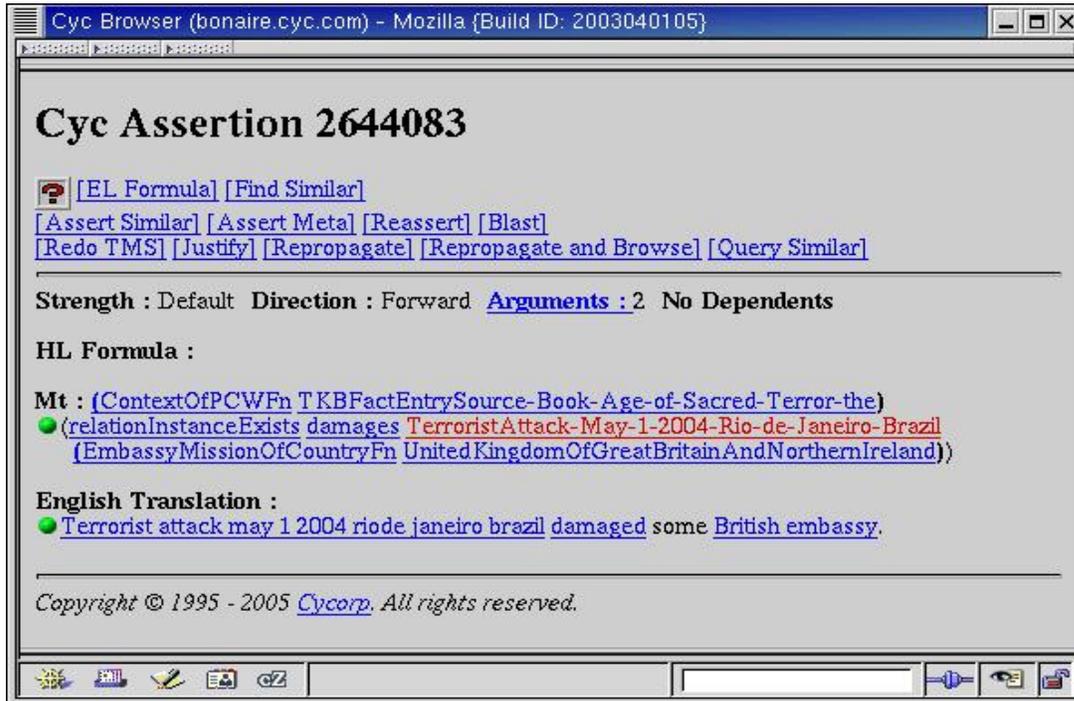
terrorist attack may 1 2004 riode janeiro brazil  Find Terrorism Act

Description	Fact
Number of targets of a particular kind:	1
Type of target:	British embassy
Status of target:	damaged object
Name of individual target:	
Status of target:	

Basic Information Non-Human Targets Human Targets Perpetrators Weapons

OK New Cancel

The user never sees the bulky precise logical form that the TKB uses – the CycL language, which is essentially the same as first order predicate calculus with equality and contexts, etc. Here is some of what has been generated, internally, automatically, from what the user just typed in about the British Embassy being damaged in the attack:

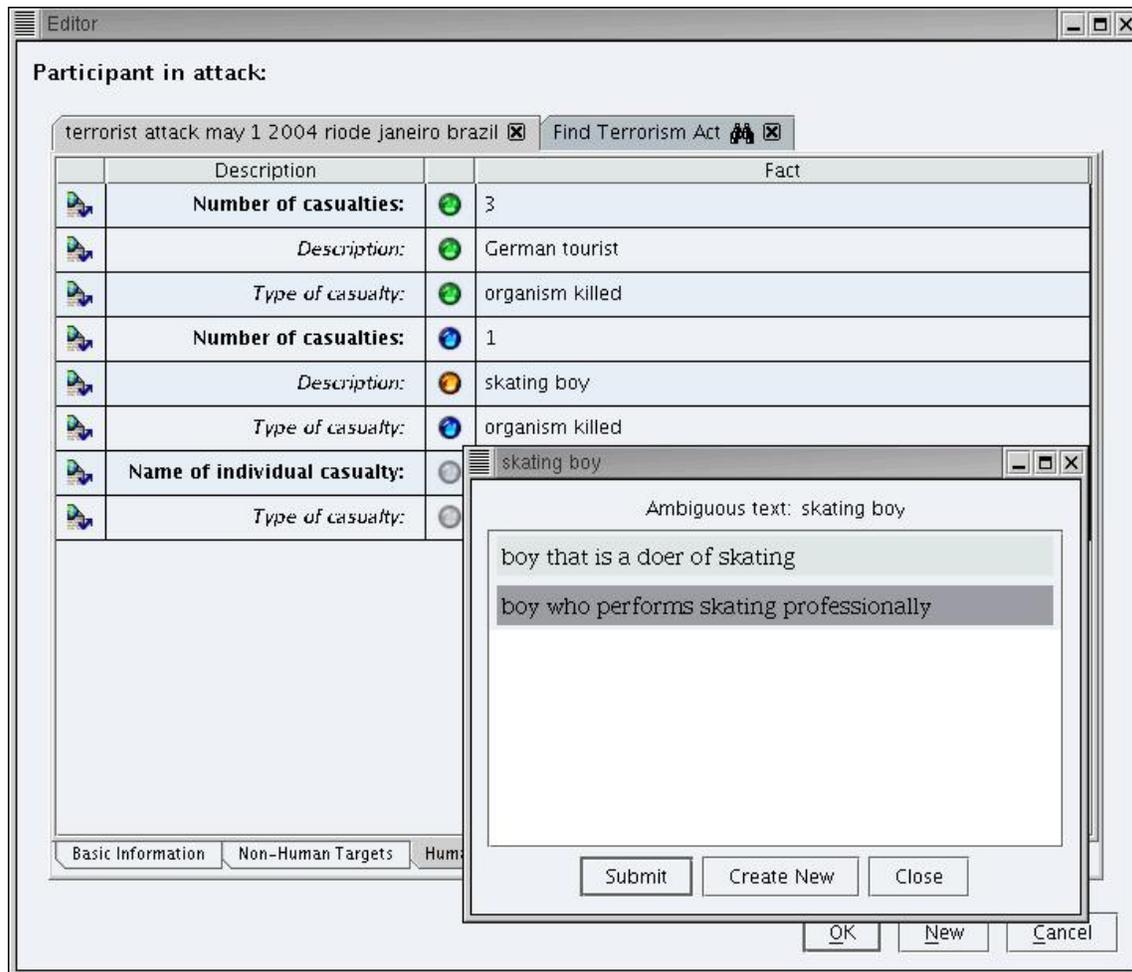


As you can see, there is a logical expression of the fact, and also a paraphrase of it in English that the TKB generated (that the car bombing “damaged some British embassy”).

(12) The user can enter human casualty information by going to the “Human Targets” tab. They type “3” in the “Number of Casualties” field, “German tourists” in the “Description” field, and select “organism killed” from the drop down menu in the “Type of Casualty” field to represent the fact that three German tourists were killed in the attack.

(13) To add a another piece of information of the same sort, namely that one boy was killed in the attack while he was skating, they click on the “Number of Targets of a Particular Kind” field and select “Add Similar Entry”. They type “skating boy” in the description field, “1” in the “Number of Targets of a Particular Kind” field.

(14) After typing “skating boy”, the circle next to that field turns orange. This means that the user needs to disambiguate the meaning of “skating boy” to the system. The user clicks on the orange circle to get the disambiguation menu to pop up. They have a choice between “a boy who performs skating professionally” and “a boy who is a doer of skating”. See the following screenshot.



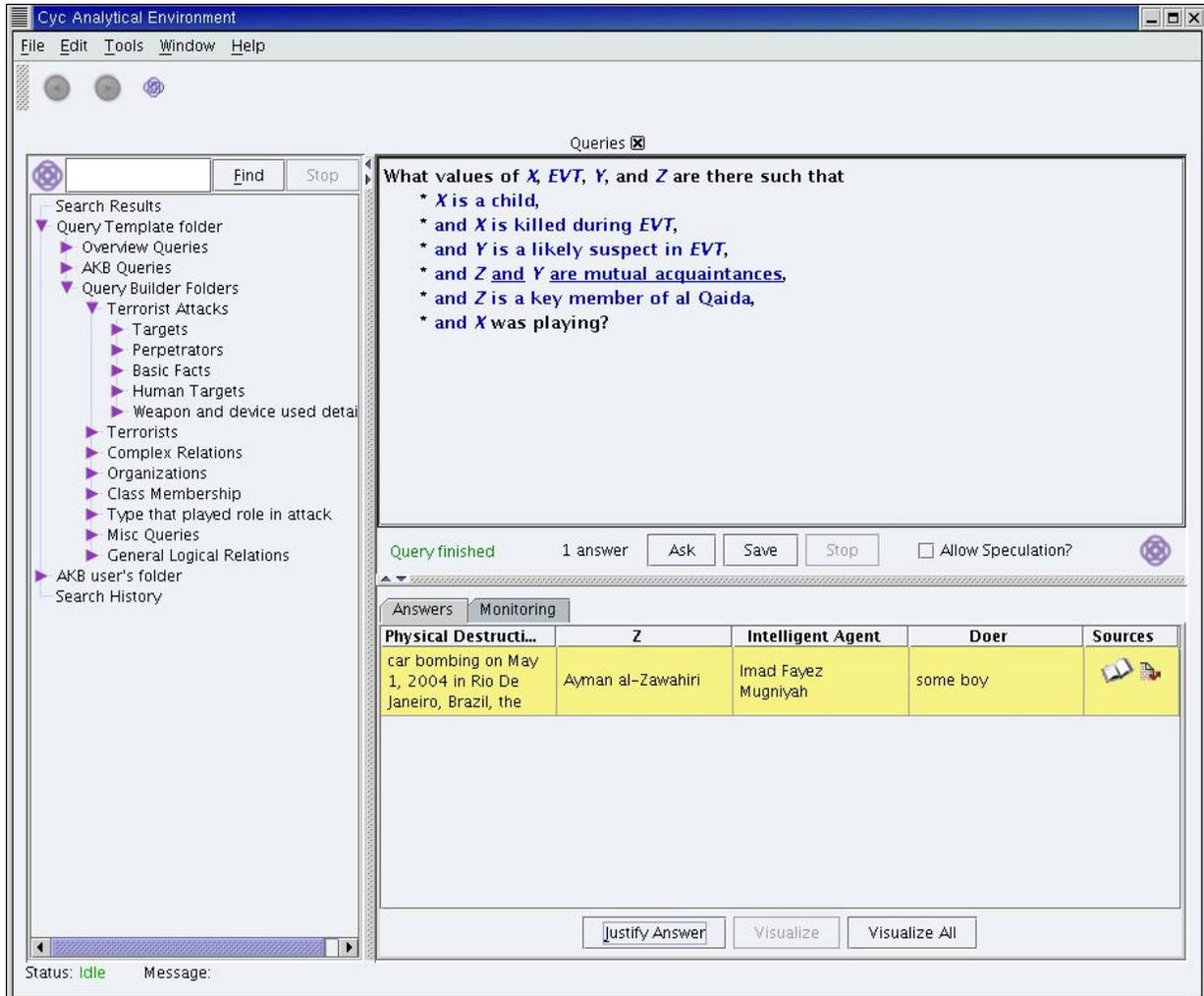
(15) The user repeats this process to represent that several embassy security guards were wounded in the attack. When the user types “embassy security guards” in the description field, the circle next to the field will again turn orange. Cyc has a few interpretations for what that phrase means. The most likely candidate for the intended meaning is “security person who works at some embassy”; this is indeed what the user meant, and they click on that choice.

(16) To finish entering information about the attack, the user clicks on “Role played in the attack”, and selects “likely suspect” from the drop down menu. This tells the TKB that (according to this CNN article) Mughniyah is a likely suspect in that attack.

### **An example of the analyst posing a query to the TKB and getting an answer**

The TKB now “knows about” that attack. If someone asks the TKB a query for which this attack should come up as an answer, the system will return the newly created term, assuming that the TKB (and underlying Cyc knowledge base) have enough domain knowledge and common sense knowledge to do the necessary deduction.

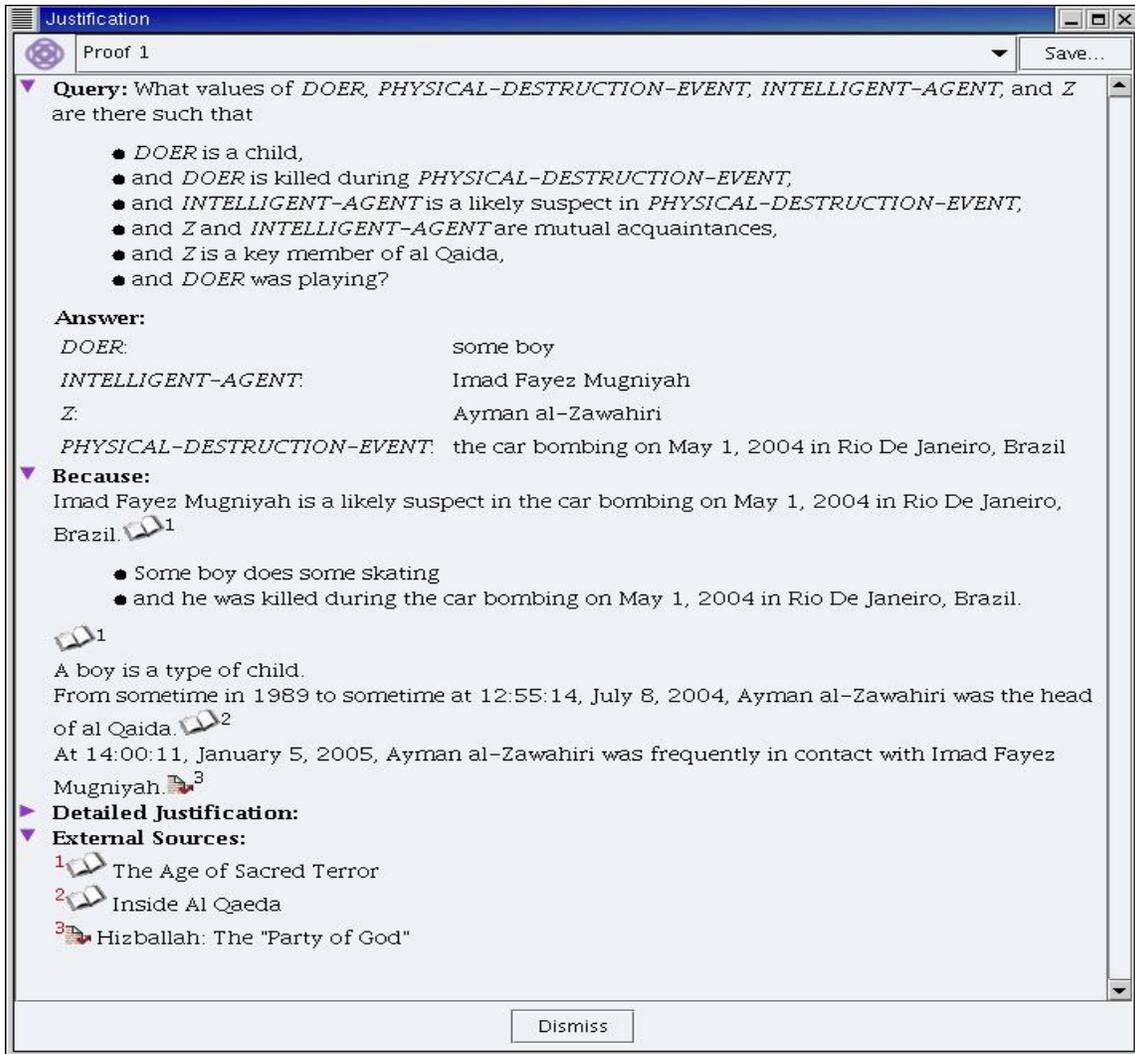
For example, suppose the analyst asks the TKB to “list known attacks which killed any children at play, for which a likely suspect is someone who is mutually acquainted with a key member of Al Qaida”. They can find a similar query and modify it to pose it to the TKB or they can construct the query from scratch using a provided library of query templates. The system will return the attack we just entered into the system, as one of the



answers to this query, based on the following line of reasoning:

- TKB knows that a “skating boy” was killed during the attack and Cyc knows from “common knowledge” that this means a child, not an adult, and that they were playing (they were not, e.g., in a skating tournament at the time).
- TKB knows that Imad Mugniyah is a likely suspect for that attack and – from earlier CNN reports about Mugniyah – knows that he is in frequent contact with al-Zawahiri. Cyc infers, then, that they must be mutual acquaintances.
- Cyc knows that al-Zawahiri is one of the leaders of one of the suborganizations of Al-Qaida and is therefore a key member of that Al-Qaida.

See the following screenshot for this particular argument or line of reasoning, as it is presented to the user, if they click on the “Justify Answer” button:



This illustrates that analysts (with 3 days of training) can use the system to tell it information, and to pose queries and understand the detailed justifications that come back with the answers. The interface still requires some work, but it is usable if not as elegant or simple to use as a typical text search interface.

The work on this project resulted in one of the largest open source knowledge bases on terrorism in existence. Because of various reductions in funding over the course of the project, the TKB is not complete. Because of the reduced resources it was decided to try and generate very complete coverage for a chosen terrorism entity. Given the recent resurgence of its activities, Lebanese Hezbollah was selected as that entity to focus on.

### 3 Technical Accomplishments

When dealing with event-like entities, terrorist attacks, meetings, etc., a standard representational approach is to generate terms to represent the events which are then used to group various pieces of information relating to the events such as dates and times, locations, and the people or objects involved in the events. For example, when representing a terrorist attack that occurred in Beirut, Lebanon on July 21, 1998, the system would generate a term with a name suggestive of the type of thing it represents (but the name itself has no semantic significance).

TerroristAttack-435678

The above term could then be used when representing knowledge about the attack

*(isa TerroristAttack-435678 TerroristAttack)*  
*(eventOccursAt TerroristAttack-435678 CityOfBeirutLebanon)*  
*(dateOfEvent TerroristAttack-435678 (DayFn 21 (MonthFn July (YearFn 1998))))*

As additional information about the event becomes known, the term can be updated with further information.

*(intendedAttackTargetType TerroristAttack-435678 GovernmentBuilding)*  
*(assistingAgent TerroristAttack-435678 SOME-TERROR-GROUP)*

This method of representing events and related entities is called *Davidsonian*. While this method works well for representing facts about events, other types of information are harder to represent in this fashion. Consider the following.

“Rafik was in Beirut throughout July 2003.”

“Ronald was a member of the Cincinnati Better Business Bureau sometime during 1995”

The facts described by the above sentences are less “event” like and more “state” like. They describe a particular state of the world during a period of time. It is awkward at best to force the above sorts of information into the Davidsonian framework. Doing so results in strange constructions like

“The situation of Rafik being in Beirut’s temporal extent has July 2003 as a subinterval.”

“Ronald’s being a member of the Cincinnati Better Business Bureau’s temporal extent intersects 1995.”

These could be rendered in CycL, respectively, as

*(thereExists ?EXTENT*  
*(and*  
*(isa RafikInBeirut-01*  
*(SituationTypeSuchThatFn*

*(TheSet (objectFoundInLocation Rafik Beirut))))  
(temporalExtent RafikInBeirut-01?EXTENT)  
(temporallySubsumes ?EXTENT (MonthFn July (YearFn 2003))))*

*(thereExists ?EXTENT  
(and  
(isa RonaldInCBBB-01  
(SituationTypeSuchThatFn  
(TheSet (hasMembers CincinnatiBBB Ronald))))  
(temporalExtent RonaldInCBBB-01 ?EXTENT)  
(temporallyIntersects ?EXTENT (YearFn 1995))))*

While such representations can be made to work, making use of the information in inference is quite complex.

Since the Cyc Knowledge Base and inference engine already had a quite sophisticated context mechanism. It was decided we would treat sentences like those above as simple statements whose context was restricted instead of complex statements true over a broad array of contexts.

We therefore have implemented a full context reasoning system using the meta-predicate *ist* and Cyc contexts which we call “microtheories”. This system includes temporal reasoning via temporal contexts. A large number of the assertions in the TKB are asserted in temporal contexts. This allows us to reason about when individuals have certain properties or relations without having to construct special case predicates that have additional arguments for dates or time intervals. For example, the fact that Rafik was in Beirut in July of 2003 could be represented with a special case predicate, as in the following:

*(objectFoundInLocationDuring Rafik Beirut (MonthFn July (YearFn 2003)))*

If true, the above will be true in a particular context timelessly. However, we have found that instead of using that special ternary predicate *objectFoundInLocationDuring*, it is much better to represent that information with the simpler, binary, *objectFoundInLocation* and place that assertion in the proper temporal context as in the following example:

*(ist (MtSpace ExampleMt (MtTimeDimFn<sup>1</sup> (MonthFn July (YearFn 2003))))  
(objectFoundInLocation Rafik Tulsa))*

---

<sup>1</sup> Since there are several dimensions of context-space, not just “Time/Date”, we have adopted a more general way of specifying a region of that n-dimensional context-space, rather than having a dozen nested special-case predicates. In the case of one single criterion, such as month/year, this is slightly more cumbersome-looking, but averaged over all context-specifications is actually quite compact. In this case, *(AtTime July2003)* would be written *(MtSpace EMt (MtTimeDimFn July2003))*. For similar reasons, July2003 is not reified with its own named term, but rather is specified functionally as *(MonthFn July (YearFn 2003))* which means the month of July in the year 2003. Finally, it is important to indicate the degree of granularity of the assertion: was Rafik in Beirut *at least once for a moment* during July, 2003? For some portion of every calendar day? Every second of the entire month? etc. We have adopted predicates for representing this granularity as a separate argument, explicitly; examples are below.

This is more inferentially efficient and allows us to more quickly return the logical consequences of that piece of information. Alternatively, one could use a predicate like `holdsIn` to represent the same fact as follows.

*(holdsIn (MonthFn July (YearFn 2003)) (objectFoundInLocation Rafik Tulsa))*

However, supporting a predicate `holdsIn` either requires supporting a different “holds in” predicate for each dimension of evaluation or adding argument places to specify the dimension in which the sentence holds. In general, taking advantage of our existing context mechanism to represent these dimensions is the most natural way to capture this information. This also allows for the easy application of further context dimensions such as geospatial location, who believes the claim, classification level, and so on.

The vocabulary we have created for context allows us to abstract out the important features of a context like the dimension and granularity. For example, when implementing a *location* dimension one has to account for the fact that while it is true that there are about 700,000 medical doctors throughout the United States, it doesn’t follow that there are 700,000 medical doctors in Maryland even though Maryland is a proper sub-region of the United States. Yet, if it is snowing throughout Maine, then it is snowing in every sub-region of Maine. The difference can be captured with the notion of the granularity of a dimension – in the doctor example, the granularity is the whole region while in the snowing example the granularity would probably be an acre. There are analogous issues with implementing temporal contexts. For example, while it may be true that the gross national product of Canada was  $N$  in 1999, it doesn’t follow that the gross national product of Canada was  $N$  in May of 1999. However, if an individual resided in Toronto during 1999, then that individual resided there at each sub-interval of 1999. Again, the notion of granularity comes into play here. The granularity for a sentence expressing the gross national product of some country will be the entire interval at which it was asserted while the granularity of a sentence describing someone’s residence will likely have a granularity of time point, meaning that the sentence will be true at each time point that is subsumed by the time index in which the sentence is asserted. For example, to state that  $P$  holds in some microtheory  $M$  at some time index  $T$  to time point granularity we write:

*(ist (MtSpace M (MtTimeWithGranularityDimFn T TimePoint<sup>2</sup>)) P).*

What follows are examples of some inferences that this sort of context mechanism enables and a discussion of some of the problems encountered when trying to reason cross-contextually.

---

<sup>2</sup> The argument “TimePoint” refers to what we have called a “granularity”. If an assertion,  $P$ , is true at some interval  $T$  to the granularity TimePoint, then  $P$  is true at each sub-interval of  $T$  that subsumes some timepoint.

### 3.1 Spatio-Temporal Reasoning in the TKB

Knowing that someone can't be in two different places that are spatially disjoint (have no shared area or parts) at the same time can, among other things, help determine questions of identity – given current information, this “Rafik Smith” can't be the same as that “Rafik Smith” since the first was in Tulsa on October 1<sup>st</sup> of this year and the second was in Los Angeles at the same time. So the TKB would be able to whittle down some list of “Rafik Smiths” to just those for which it is consistent with known facts that they be identical. Or, the TKB could prove that Rafik wasn't in Los Angeles at that time based on its knowledge that he was in Tulsa. Conversely, of course Rafik was in Oklahoma, since he was in Tulsa then. And he was not on a large ship, since Tulsa is inland. These sorts of inferences are trivial for humans, but they can't be performed by machines that lack the commonsense knowledge that makes them valid. In the following, we examine in detail how this knowledge can be formalized, using context logic, to enable basic spatio-temporal inference in the TKB.

Assume we are trying to prove that Rafik is not in Los Angeles at some time, T, given the knowledge that he was in Tulsa at that time. In the TKB, a specialized microtheory, say *RafiksTripMt*, would hold all of the *atemporal* data about Rafik's trip. For example, this theory might contain an assertion to the effect that:

Rafik's visit to Tulsa happened after his visit to Orlando.

Assertions like this, stating the temporal order of events, are *atemporal* in the sense that they are true at all times if they are true at any time. In contrast, an assertion to the effect that *Rafik is in Tulsa* is *temporal* – it is true at some times and false at others. Temporal assertions must be represented in temporal microtheories in the TKB. A temporal microtheory is typically defined as a composite of an atemporal microtheory<sup>3</sup>, a time interval, and a temporal granularity. For example, the temporal Microtheory used in describing everything that happened in Rafik's trip throughout the interval T is denoted by the following expression:

**C: (*MtSpace RafiksTripMt (MtTimeWithGranularityDimFn T TimePoint)*)**

This microtheory can be thought of as describing a part of the whole context of Rafik's, trip, namely the part of that context that occurs throughout the time T. One of the assertions true in that context, i.e. in C, is that Rafik is in Tulsa:

**1. (*objectFoundInLocation Rafik Tulsa*)**

We represent that fact that (1) is true in that context by relating the context term, C, to (1) via the relation *ist*:

**(*ist (MtSpace RafiksTripMt (MtTimeWithGranularityDimFn T TimePoint)*)**

---

<sup>3</sup> Atemporal microtheories are also, more properly, called “monadic” microtheories or “monads”. For any dimension specific microtheory of which they are a part, they can be thought of as contributing all the truths that are true independently of a particular dimension.

*(objectFoundInLocation Rafik Tulsa)*

The TKB needs to be able to infer spatio-temporal consequences of this knowledge. For example, it is easy to imagine that an analyst's query about Rafik's activities will require proving that according to the known data about Rafik's trip, it was *not* the case that Rafik was in Los Angeles at T. In CycL, this amounts to proving the following:

2. *(not (objectFoundInLocation Rafik CityOfLosAngeles))*

in the context:

C: *(MtSpace RafiksTripMt (MtTimeWithGranularityDimFn T TimePoint))*

To prove 2 is true in the context C, the TKB's inference engine can use any CycL assertions that are asserted in context C, i.e. explicitly represented as being true in the context C, such as 1, and also any assertions that are asserted in contexts more general than C. The predicate *genlMt* is the CycL predicate that expresses the relation of context generalization. *genlMt* is a transitive and reflexive relation that is partially defined by the following axiom:

$$(genlMt C1 C2) \leftrightarrow (\forall P)((ist C2 P) \rightarrow (ist C1 P))^4$$

The following rule is asserted in the TKB's NaiveSpatialMt:

R1. *(not*  
*(and*  
*(objectFoundInLocation ?OBJECT ?PLACE1)*  
*(objectFoundInLocation ?OBJECT ?PLACE2)*  
*(spatiallyDisjoint ?PLACE1 ?PLACE2)))*

The expression *(objectFoundInLocation X Y)* means that X is wholly located at Y. So this rule means that if two places have no common part, then no object is wholly located at both X and Y. The microtheory NaiveSpatialMt is a generalization of RafiksTripMt:

*(genlMt RafiksTripMt NaiveSpatialMt)*

In turn, RafiksTripMt is a generalization of C:

*(genlMt (MtSpace RafiksTripMt (MtTimeWithGranularityDimFn T TimePoint)) RafiksTripMt)*

So R1 can be used in C in order to prove that Rafik is not in Los Angeles, and, given that claim 1, above and the rule R1 are both true in C, it only remains to be shown that:

*(spatiallyDisjoint Tulsa CityOfLosAngeles)*

is true in C in order to show that *Rafik is not in Los Angeles* is true in C.

---

<sup>4</sup> Any statements S with free variables  $x_1 \dots x_n$  is treated as equivalent to  $(\forall x_1) \dots (\forall x_n) S$ .

Many types of naturally or conventionally defined types of geospatial regions, such as continent, country and city, have the following property: Distinct members of that type do not spatially overlap. For example, someone who knows nothing about London and Paris except that they are not the same city can infer that they are spatially disjoint. The TKB represents this property with the collection *SpatiallyDisjointRegionType*. The collections *TrueContinent*, *Country*, *City* and 241 others are instances of *SpatiallyDisjointRegionType*. *SpatiallyDisjointRegionType* is defined by the following rule in the *NaiveSpatialMt*:

R2. (*implies*  
*(and*  
*(isa ?REG-TYPE SpatiallyDisjointRegionType)*  
*(isa ?REG-1 ?REG-TYPE)*  
*(isa ?REG-2 ?REG-TYPE)*  
*(different ?REG-1 ?REG-2))*  
*(spatiallyDisjoint ?REG-1 ?REG-2))*)

So, the above rule can be used by the TKB to prove that Los Angeles and Tulsa are spatially disjoint in the context C. So the system knows, or can prove, each of the following in context C:

*(spatiallyDisjoint Tulsa CityOfLosAngeles)*

*(objectFoundInLocation Rafik Tulsa)*

*(not*  
*(and*  
*(objectFoundInLocation ?OBJECT ?PLACE1)*  
*(objectFoundInLocation ?OBJECT ?PLACE2)*  
*(spatiallyDisjoint ?PLACE1 ?PLACE2)))*

and these are sufficient to prove that:

*(not (objectFoundInLocation Rafik CityOfLosAngeles))*

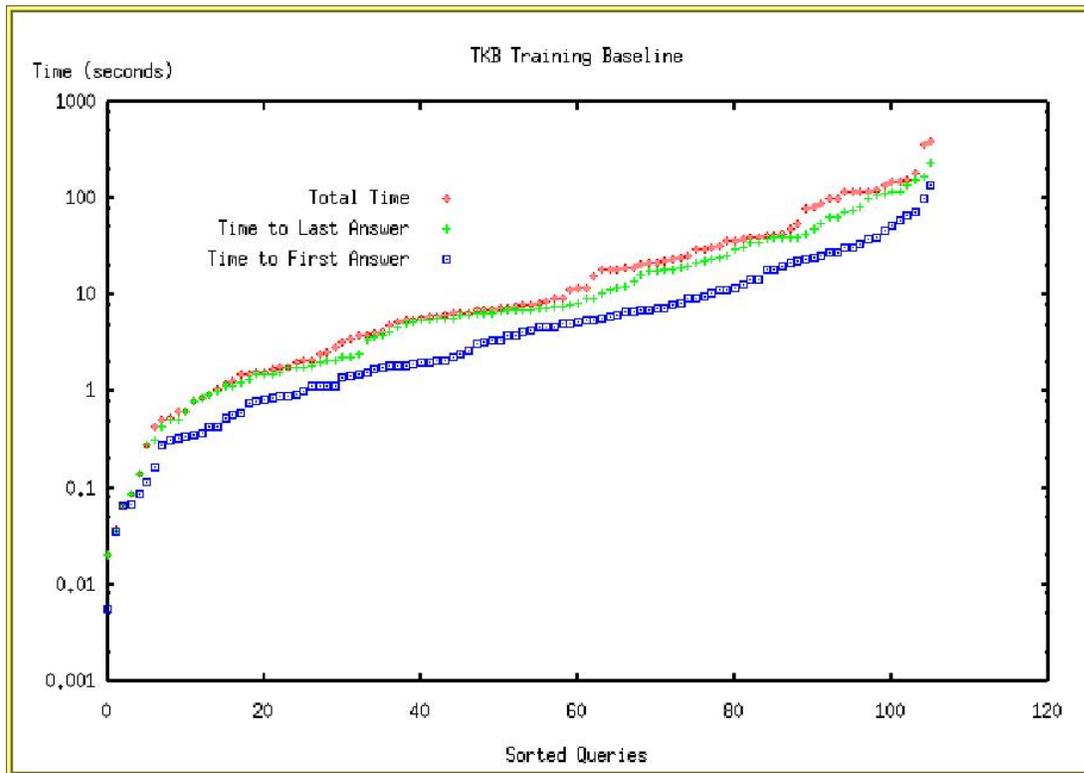
In summary, the TKB's temporal inference abilities in conjunction with its commonsense geospatial knowledge enable it to conclude the fact that Rafik was not in Los Angeles at T, from Rafik's presence in Tulsa at T.

## 3.2 Experiments

### 3.2.1 Project Arete.

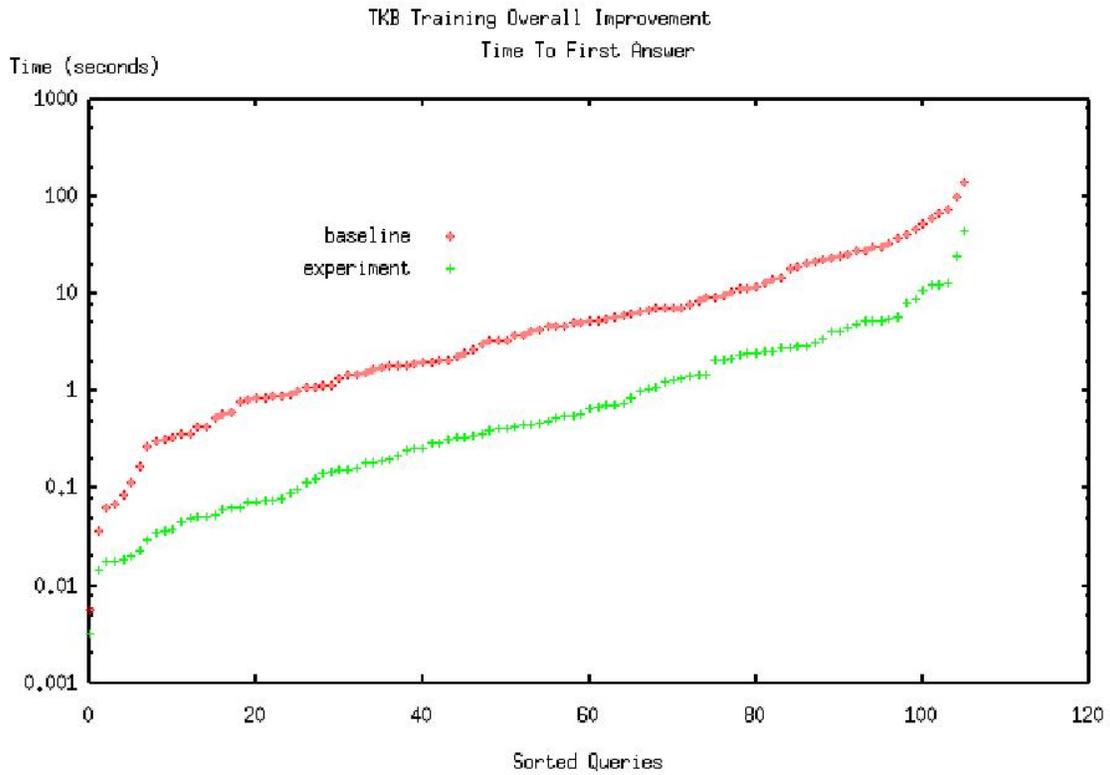
By mid-2004 we had already amassed a relatively large number of example queries in the terrorism domain. Many of these queries already produced results. Project Arete was a two-month-long effort with the goal of improving inference times for relatively shallow queries. The experiment was conducted over a corpora of terrorism domain queries. A training corpora that consisted of 106 of these queries was selected. Of those 106, ninety-five of them were multi-literal queries that required appeal to various inference modules that provided capabilities such as temporal subsumption, transitivity, etc, but did not require appeals to a general theorem-prover to solve.

The queries were each run and a baseline established. See the following figure:

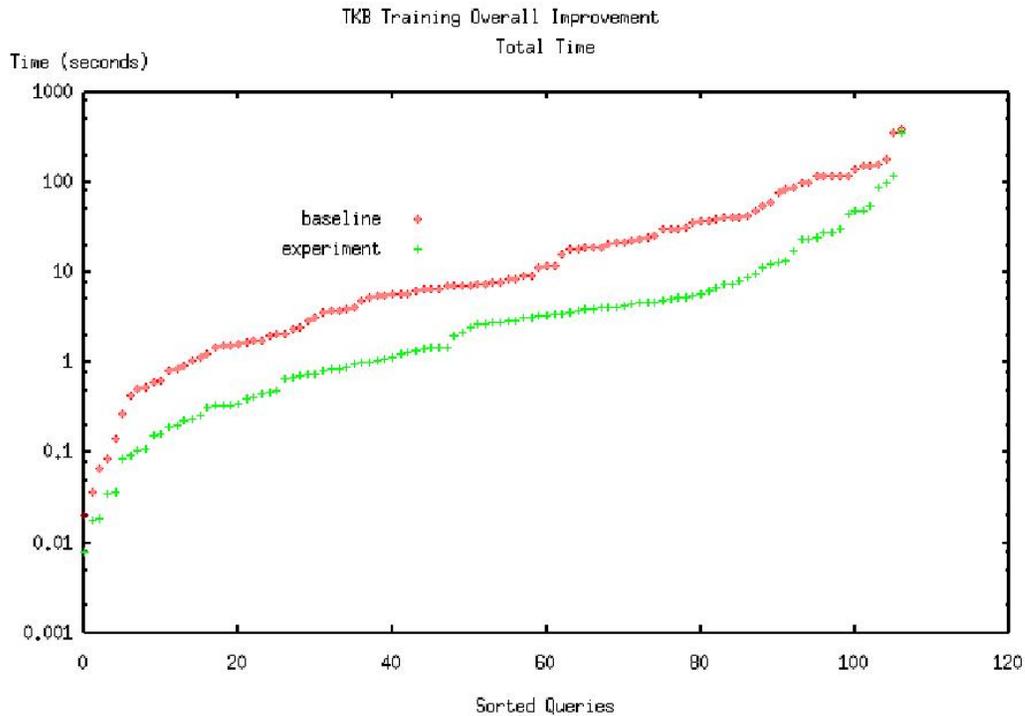


The time to first answer varied from the low of less than  $1/10^{\text{th}}$  second to just over a minute and a half with time to last answer following not long after. At this point a series of experiments were performed ranging from suspending or compiling out argument type checking to disallowing continuation of inferences that time out. The experiments resulted in improvements to the inference engine that resulted in an overall improvement in both time to first answer and total time.

The following graph show the time to first answer improvement after implementing the insights provided by the experiments:



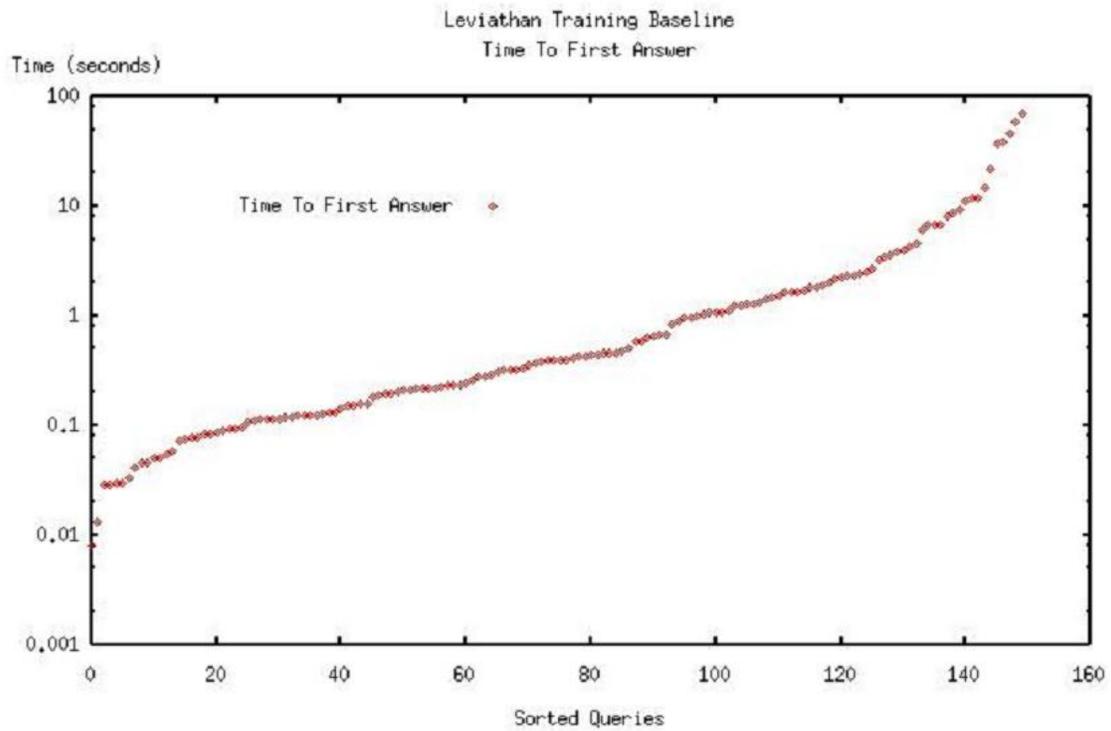
This next graph shows the improvement in total query time after the experiments:



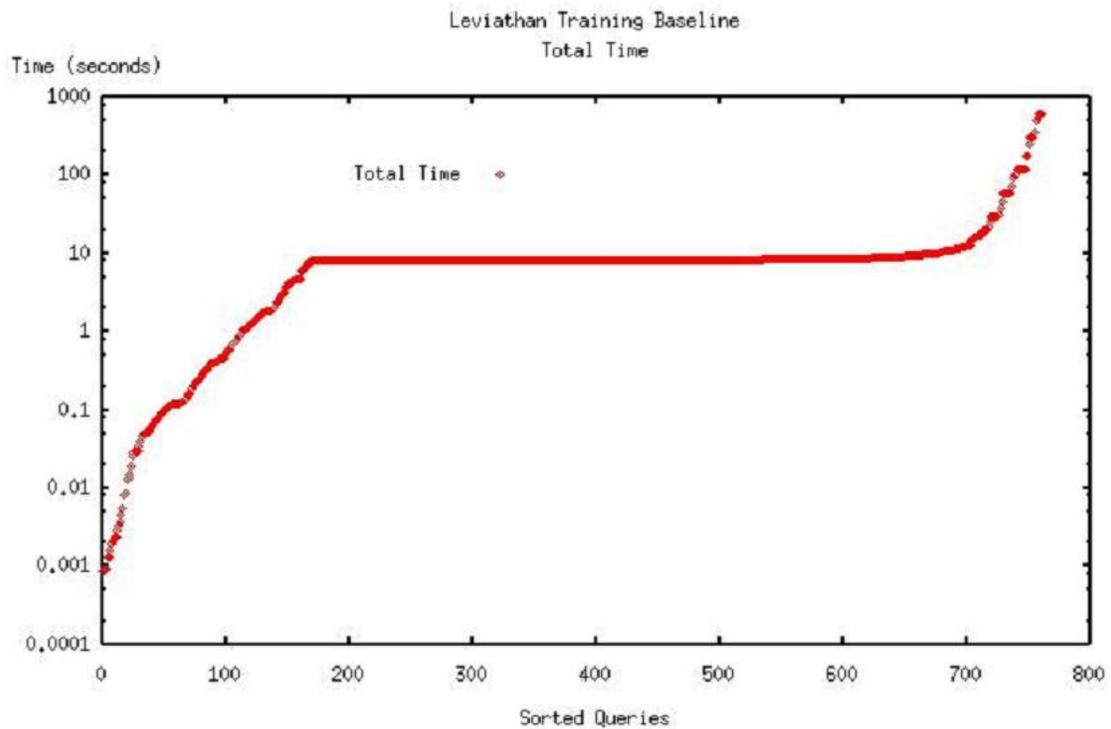
### 3.2.2 Project Leviathan

Project Leviathan was a follow-on inference experiment that aimed to optimize deeper queries that needed to appeal to rules asserted in the KB in addition to specialized inference modules. Contrast these queries with those of Project Arete. The Arete corpus consisted of queries that could be solved by appeal to specialty inference modules alone. Leviathan queries, on the other hand, all required some appeal to a general theorem-proving module.

The corpus for this experiment consisted of 411 already existing queries and 378 queries generated automatically for the experiment. Roughly half of the total corpus should, theoretically, result in answers being returned. The following graph gives the initial baseline analysis of the corpus:



This next graph gives the total time spent in inference for each of the queries in the corpus:



As with Project Arete, several different experiments were performed. Unlike Arete though, there was no single experiment that led to dramatic increases in performance. But there were several insights gained that ultimately led to increased performance for these queries. The most notable of these insights is that “remembering” successful rule combinations that led to progress in solving the query led to an overall increase in the number of queries that returned answers. This led to the development of experience-based rule sorting. That is, the system will now give higher motivation to performing transformations with rule combinations that in the past have led to further progress in proofs for previous queries. An experiment was conducted in which an experience collecting first run of the corpus was done and then the corpus was re-run using the data from the first run to sort which rules were appealed to during the proofs in the queries.

Rerun with experience gained from the first run

•TOTAL-ANSWERABLE	150 -> 222
•MEDIAN-TIME-TO-FIRST-ANSWER	0.39 -> 0.32
•MEDIAN-TIME-TO-LAST-ANSWER	0.43 -> 0.37
•MEDIAN-TOTAL-TIME	1.08 -> 1.05
•MEDIAN-TIME-PER-ANSWER	1.05 -> 0.85

The big difference is in the number of queries that were unanswerable that became answerable. Only 2 queries went from answerable to unanswerable. But we went from being able to answer only 150 of the queries to being able to answer 222 of the queries.

The results from this particular experiment led to the implementation of the generalized utility-based rule-pruning module that is in place in the inference engine today.

### *3.3 The development of the Fact Entry Tool (FET) and the Cyc Analytic Environment (CAE)*

In early 2003 we held a series of workshops with leading terrorism experts to determine the optimal scheme for the representation of terrorism information. It was determined that the knowledge base will represent a wide range of information about three kinds of terrorism-related entities.

Terrorist Organizations  
Individual Terrorists  
Terrorist Attacks and Operations.

See Appendix I for a detailed listing of the schema element that resulted from these meetings. At that point we turned to the development of two distinct user-interfaces that were eventually combined into a single application.

Utilizing some earlier work performed for the Rapid Knowledge Formation project we designed an assisted knowledge entry tool that allowed lightly trained users to enter structured information. The FET (described in detail in section 1) is a java application embedded within the CAE. Most all the information that governs the appearance and

content of the FET application is stored declaratively in the Cyc Knowledge base. That is, information about what knowledge entry templates should be presented to the user, their organization and order, and the assertion that would result from a user entering a value in the appropriate field are stored in the knowledge base using vocabulary designed for that purpose. At startup, the system polls the knowledge base to determine what knowledge entry templates it should present to the user and in what order.

The knowledge is organized into collections of specific template types. The top-level template type is called a “topic”. Associated with each topic is a particular entity type that an FET user can enter information about. The TKB’s FET has several topics including #TerroristAttack, #Terrorist, and #TerroristGroup. The immediate children (in CycL terms, specializations) of a topic correspond to particular tabs in the FET screen.

For example, #TKBTemplate-Attack-HumanTargets is the topic that contains the template for specifying the type of person injured/killed/captured/etc. during an event as well as the template for specifying particular named casualties. Each template is associated with a CycL formula. For example the templateTKBHumanTargetTemplate-Parsed is associated with a formula via the following assertion.

```
(formulaForFormulaTemplateTKBHumanTargetTemplate-Parsed
  (relationInstanceExistsCount
    (SomeExamplePredicateOfTypeFn AgentCasualtyPredicate)
    (SomeExampleFn TerroristAct)
    (SomeExampleFn
      (SpecsFn IntelligentAgent))
    (SomeExampleFn NonNegativeIntegerRange)))
```

The “SomeExample...” portions of the above formula are placeholders for terms that are either chosen by the user via a drop down menu or the result of parsing a user’s input string to a particular Cyc concept.

When a user fills in all the fields in this template either via parsing or menu selection, the following type of assertion is made to the knowledge base

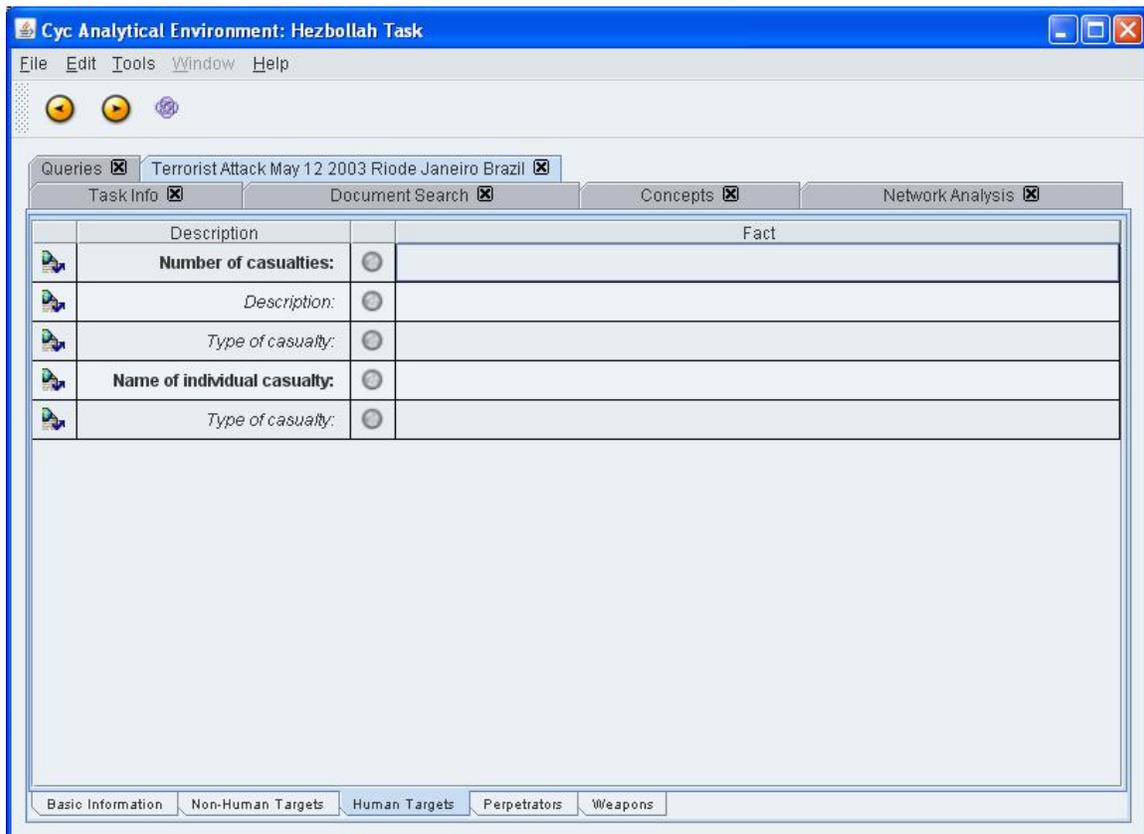
```
(relationInstanceExistsCount PREDICATE ATTACK TYPE-OF-AGENT NUM)
```

Adding additional topics, tabs, or templates consists of defining terms to represent the topic, tab, or template and asserting the relevant defining information into the knowledge base. This allows the FET to be dynamically updated with no additionally coding and recompilation required. For example, to add a new knowledge entry template to an existing FET tab, one simply defines a new template, associates it with a formula and asserts that it is an instance of the template type that corresponds to that tab. At this point, updating or adding to the knowledge entry templates requires users knowledgeable about CycL and the Cyc knowledge base, but we are investigating interfaces that would allow end users to define their own templates.

### 3.3.1 Improvements to natural language understanding

Since, ultimately, the FET and CAE systems being developed had to be able to be used by people with little to no experience in formal languages, we were required to continue our work in improving and expanding Cyc's lexicon of mappings from natural languages like English to concepts in the knowledge base as well as improve our ability to compositionally generate appropriate CycL formulas from natural language strings. Although we spent some time investigating the generation of a CycL interpretation of full natural language sentences, the state of the art of natural language understanding is not yet at the point where we can reliably generate a fully machine-understandable formal representation of full natural language sentences. But we have made great strides in understanding small to medium chunks of natural language text in the context of some overarching task like terrorism analysis. The CAE makes use of this functionality in various places.

One prominent place where the CAE makes extensive use of focused natural language understanding is in the FET. Examples of this are given in the overview of the FET earlier in the document, but it will be helpful to examine in detail what is going on in the system when a user enters a natural language phrase into a FET field. Imagine that the user is trying to state that 3 German tourists were killed during a particular terrorist attack. The user would navigate to the human targets tab in the FET. See below.



They would type “3” or “three” into the *Number of casualties* field and then type a short phrase that would describe the casualties into the *Description* field. In this case they would type “German tourists”. They may have been no prior explicit representation of “German tourist” in the knowledge base at that time. But there are explicit representations of “German person” and “tourist”. The system will then dynamically create the appropriate concept of “German tourist” automatically and use the newly generated term in the resulting assertion. In this case, the term the system creates is the logical intersection of the concepts #GermanPerson and #Tourist.

(CollectionIntersection2Fn GermanPerson Tourist)

This denotes or has as its extension the collection of German persons who are also tourists. The resulting assertion into the knowledge base for this particular example would be

(relationInstanceExistsCount organismKilled  
TerroristAttack-May-12-2003-Rio-de-Janeiro-Brazil  
(CollectionIntersection2Fn GermanPerson Tourist) 3)

This is a compact representation that is short for

(thereExistsExactly 3 ?X  
(and  
(isa ?X (CollectionIntersection2Fn GermanPerson Tourist))  
(organismKilled TerroristAttack-May-12-2003-Rio-de-Janeiro-Brazil ?X)))

In English this could be rendered as

“There were exactly three German tourists that were killed in TerroristAttack-May-12-2003-Rio-de-Janeiro-Brazil.”

There is a very large number of nouns, noun phrases, date expressions, etc. for which the TKB can understand and generate the proper interpretation. In cases where there is no single unique interpretation, the system will offer up a list of possible interpretations for the user to choose from. The user may also reject any interpretation given by the system and force the system to generate a completely new concept to represent the input string. In general, the TKB project resulted in about a 50% increase in the number of actual parsing rules in the system (from 2000 such rules to around 3000 such rules) used to generate interpretations for general classes of natural language phrases. Each rule added covers a large class of possible input strings. For example, the generation of

(CollectionIntersection2Fn GermanPerson Tourist)

from

“German tourist”

is enabled by the noun compound rule #IsraeliSoldier-NCR. This covers a broad class of possible inputs where one of the terms maps to an instance of #PersonTypeByActivity (#Tourist is an example of one) and the other term maps an instance of #PersonTypeByNationality (#GermanPerson is an example of one). This allows for interpretations of such expressions as “French archaeologist”, “Sudanese diplomat”, “British pilot”, “Israeli government official”, and so forth. In fact, since there are 2088 instances of PersonTypeByActivity and 395 instances of PersonTypeByNationality, this single rule covers over 800,000 possible input strings. Of course, not every possible input string is likely to be used in a FET session – there is no “Egyptian emperor” -- but we still understand the meaning of that term even if it doesn’t currently name anyone.

Various improvements to natural language generation from CycL assertions and meta information about what sorts of information is summary worthy have enabled the system to do a decent job of generating what we call “fact sheets” for the various terrorism related entities in the TKB. A fact sheet is a system generated html page that contains a summary, in English, of most important assertions involving the focal entity. For example, the following is an excerpt from the fact sheet on Hezbollah.

**Ideology**

- After 1979 and before 09:53:49, March 3, 2005, the Lebanese Hezbollah believed in [the teachings of Ayatollah Khomeini](#)  <sup>81</sup>
- At some point prior to February 2005, the Lebanese Hezbollah believed in [anti-American ideology](#)  <sup>70</sup>
- At some point prior to February 2005, the Lebanese Hezbollah believed in [Shiite Islam](#)  <sup>70</sup>
- At some point prior to February 2005, the Lebanese Hezbollah believed in [Islamic Society ideology](#)  <sup>70</sup>
- (2 similar sentences)

**Affiliations**

- From sometime in 2000 or earlier to at least sometime in November 2003, [Shadi Abu-Hasin](#) had leadership responsibilities in the Lebanese Hezbollah.  <sup>82</sup>
- From sometime in the 1980s or earlier to at least September 2004, [Secretary General Sheik Sobhi Tufeli](#) had leadership responsibilities in the Lebanese Hezbollah.  <sup>83</sup>
- From sometime in the 1980s or earlier to at least July 2004, [Hussayn al-Mussawi](#) had leadership responsibilities in the Lebanese Hezbollah.  <sup>84</sup>
- From sometime in 1984 or earlier to at least sometime in 1989, [Sheikh Abdel Karim Obeid](#) had leadership responsibilities in the Lebanese Hezbollah.  <sup>83</sup>
- (5 similar sentences)

The Lebanese Hezbollah has [al-Jihad al-Islami](#) as one of its military wings.  <sup>85</sup> The Lebanese Hezbollah has [Lebanese Platoon](#) as one of its military wings.  <sup>86</sup> The Lebanese Hezbollah has [Faithfulness to the Resistance Parliamentary Bloc](#) as one of its political wings.  <sup>87</sup> The Lebanese Hezbollah has [Loyalty to the Resistance Bloc](#) as one of its political wings.  <sup>88</sup> [Martyr](#)

At the end of each assertion in a fact sheet is a pointer to a footnote that indicates the source the subject matter expert used when entering that information. If the representation of the source is associated with a URL (e.g. if the source is an online article), then clicking on the URL in the footnote will open a web link to either that

online source or a locally cached copy of it. See the following screen shot for an excerpt from the footnote listings in the Hezbollah fact sheet.

82	 MEIB staff, "Hezbollah Increasing Terrorist Activity", <i>United States Committee for a Free Lebanon</i> , March 2004, <a href="http://www.meib.org/articles/0402_12.htm">http://www.meib.org/articles/0402_12.htm</a> .
81	 David Lewis, "Lebanon -- Party of God", <i>PBS Frontline World</i> , May 2, 2003, <a href="http://www.pbs.org/frontlineworld/stories/lebanon/thestory.html">http://www.pbs.org/frontlineworld/stories/lebanon/thestory.html</a> .
80	 Hezbollah hits Israeli post in Shabaa, <i>United Press International</i> , May 21, 2000 .
70	 Harvey W. Kushner, <i>Encyclopedia of Terrorism</i> , 2003, Sage Publications, Inc..
68	 Christopher Reuter, <i>My Life is a Weapon: Modern History of Suicide Bombing</i> , 2004, Princeton University Press.
55	 "Hezbollah - Terrorist Organization of Global Reach", <i>Intelligence and Terrorism Center at the Center for Special Studies</i> , March 11, 2003, <a href="http://www.intelligence.org.il/eng/bu/hizbullah/hezbollah.htm">http://www.intelligence.org.il/eng/bu/hizbullah/hezbollah.htm</a> .
52	 TKB Fact Entry Source Newspaper Agence France Presse October 8 2002
7	 MIPT Terrorist Attack Database 1968-1997, <i>Oklahoma City National Memorial Institute For The Prevention Of Terrorism</i> , <a href="http://db.mipt.org/rand_68_97.cfm">http://db.mipt.org/rand_68_97.cfm</a> .

### 3.3.2 The CAE and the Query Library

One of the biggest improvements to the TKB system involved improving our focused extraction of concepts from user search strings. This led to an order of magnitude improvement in how we provide query fragments for users to use in building their queries in the CAE. In the initial stages of the development of the TKB, the query interface accessed via the CAE allowed users to modify existing query templates, e.g. change the terrorist group, the dates or event types referenced in a pre-built query like the following

“What suicide bombings has Hezbollah performed between July 1995 and March 2000?”

The user can change “suicide bombing” to any one of a number of different event types. “Hezbollah” can be changed to reference any organization known by the TKB, and the above dates can be changed to specify any temporal interval whatsoever. See Appendix II for a complete listing of all the pre-built queries generated for this project. In general, if a term that appears in the query construction interface is hyperlinked, you can replace it with another concept either through a drop-down menu or via direct parsing to another concept. So one method of using the CAE query functionality would be for the user to look for a query that was structurally similar to the query he/she wants to ask and perform various substitutions to the query until they were satisfied it expressed what they wished to ask.

As the project progressed, additional methods of generating queries were introduced. One such method was utilizing query fragments in the query library and building up a query step by step by dragging and dropping snippets of English text into the query building section of the CAE. See the latter part of Appendix II for a complete listing of all the query fragments or “builder queries” included in the TKB. These were organized into builder query folders that mimicked the organization of the fields and tabs in the

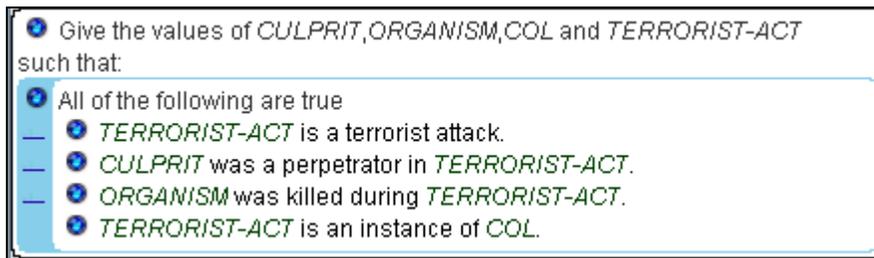
FET. This provided much needed functionality to the CAE. For many queries, it was quicker and easier to simply build the wanted query step by step than to search for structurally relevant pre-built queries and modify them as needed. For example, to build

“In what kidnapping events did Hezbollah kidnap soldiers”

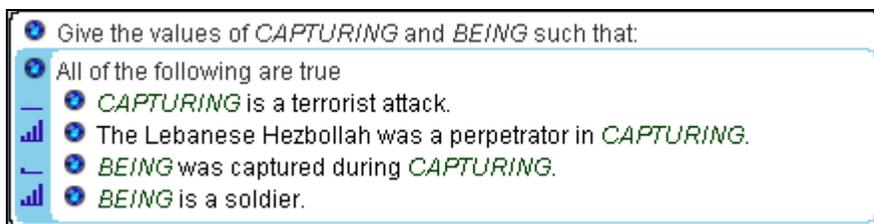
the user could drag and drop the following builder queries into the construction window

THING is a terrorist attack.  
ATTACK was perpetrated by AGENT.  
PERSON was killed in ATTACK.  
THING is an instance of TYPE.

The system will attempt to unify the various variables according to various constraints asserted in the knowledge base (argument constraints, disjointness assertions, etc.). This will result in the following appearing in the user’s query construction window:



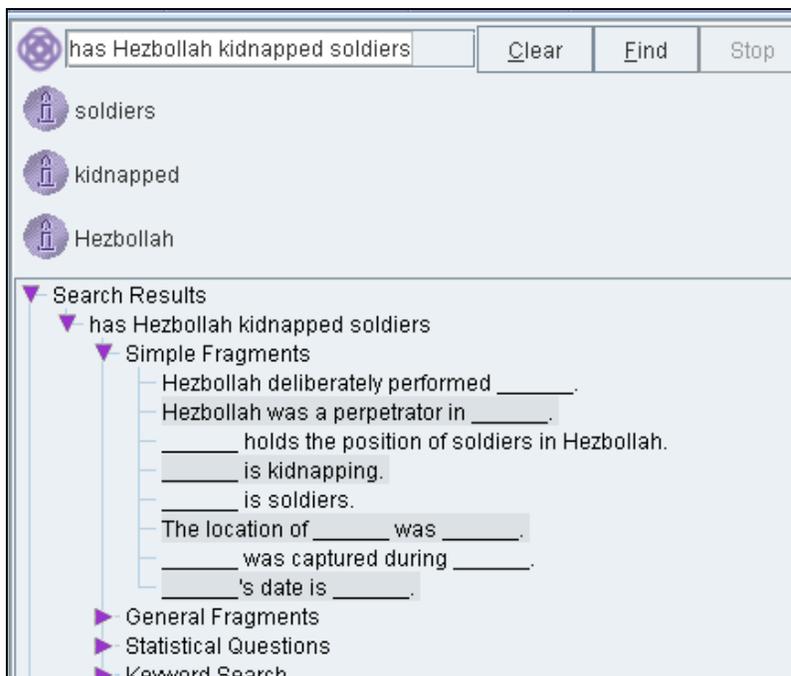
The user would then change “terrorist attack” to “kidnapping” and “killed” to “captured”, via a drop-down suggestions menus. Then, by dragging and dropping, the user could replace *TERRORIST-ACT* as it appears in the last clause with *ORGANISM* to indicate that the organism that was kidnapped during the event should be a member of a particular collection. The user would then click on *COL* and type “soldier” thereby specifying that the organism captured during the attack had to be a soldier. Similarly, replacing *CULPRIT* with “Hezbollah” would indicate that they wanted kidnappings that were perpetrated by Hezbollah. This would result in the following query:



This greatly increased the number of queries that users could ask of the system, but it still required the users to search through folders to find the relevant builder queries to use to construct their queries. That was relatively easy for users with experience using the FET since the organization of the builder queries was largely isomorphic to the structure of the various tabs and fields within the FET. But for persons unfamiliar with the FET, the process took longer. The substantial improvements to concept extraction from natural

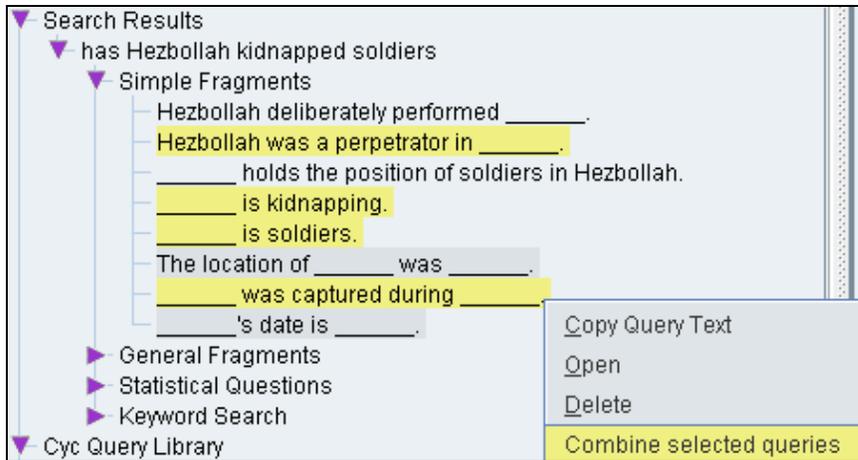
language text allowed us to take the next step and automatically generate relevant builder queries from a user search string.

Currently, the most efficient way to build a query within the CAE is to simply type a short English query into the search box. The system analyzes the input string, extracts relevant concepts, filters out irrelevant concepts (given the analysis task) and supplements those concepts with additional concepts and relations according to the context of the operative task the user is engaged in. Then, using rules in the system, the TKB will generate a set of plausible builder queries that, in most cases, are jointly sufficient to build the desired query. For example, to build the above query where the user wants events in which Hezbollah has kidnapped some soldier they could simply enter “has Hezbollah kidnapped soldiers” into the query search box resulting in the following.

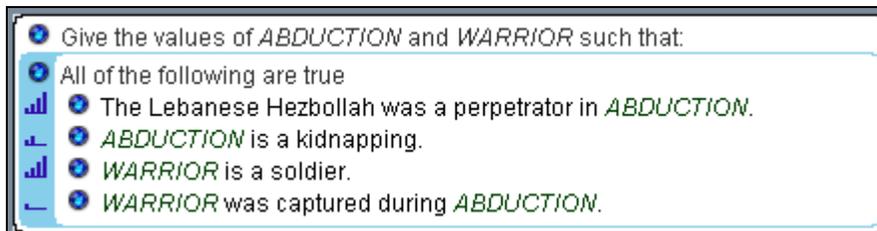


In addition to fragments that directly correspond to terms in the user’s query, there are various supplementary fragments that, based on the concepts involved in the user’s input text, the system believes may be relevant to the user’s query. For example, since “kidnapping” is a type of terrorist attack tactic and a type of event, the system suggests fragments that allow the user to specify or ask for the location of the event and its date.

Once the fragments are returned, the user can then select the fragments that they believe to be relevant to the query and ask the system to combine them into a single query. The following screen shots shows the selected query fragments and the menu option to combine them into a single query.



In some cases, the user may have to perform some additional modifications of the query to generate the exact query desired, but in many cases the above steps result in exactly the query the user wished to ask. The following shows the result of the user asking the system to combine the selected fragments.



### 3.3.3 Other Important CAE functionality

Since every piece of information entered into the TKB by the subject matter experts is required by be associated with its source, we give the users several different levels of interaction with the sources during their querying. When the user receives answers for a particular query, each row has a set of source icons that indicate the type of source used in the inference.

Answers (14)		
Abduction	Warrior	Sources
Hezbollah cross-border assault	some Israeli soldier	
Hezbollah cross-border assault	Eldad Regev	<b>JP</b>
Hezbollah cross-border assault	Ehud Goldwasser	<b>JP</b>
the terrorist attack in Har Dov on October 7, 2000	some soldier	

Hovering over a particular cell in the Sources column will give the name of the sources used in one or more of the proofs for that particular answer.

the terrorist attack in Har Dov on October 7, 2000	some soldier	 
the terrorist attack in Har Dov on October 7, 2000	some	 Media Watch International  Intelligence and Terrorism Center at the Center for Special Studies

Clicking on one of the icons will generate a complete reference to the source and if associated with a URL will be hyperlinked allowing the user to click the link to either go directly to the source (if it is still online) or to go to a locally cached copy of the source document.

Since the sources themselves are richly represented in the TKB – we represent the publisher, author, date of publication, type of source (e.g. newspaper article, database, book, etc.), edition or version, title, etc. – future functionality could include allowing the users to specify that their queries can only use sources with certain specified properties allowing the user to specify that the inference should appeal to only certain “trusted” sources or to examine differences in answers when only sources from foreign countries are used.

For each row in the answer section of the CAE, the user can also generate an English justification for why that answer was returned. The following is the justification screen for the query we explored earlier.

▼ **Query:** What values of *WARRIOR* and *ABDUCTION* are there such that

- *ABDUCTION* is a [kidnapping](#),
- *WARRIOR* is a [soldier](#),
- [the Lebanese Hezbollah](#) was a perpetrator in *ABDUCTION*,
- and *WARRIOR* was captured during *ABDUCTION*?

**Answer:**

*WARRIOR:* [Eldad Regev](#)

*ABDUCTION:* [Hezbollah cross-border assault](#)

▼ **Because:**

[The Lebanese Hezbollah](#) was a perpetrator in [Hezbollah cross-border assault](#).<sup>1</sup>

[Hezbollah cross-border assault](#) was a [kidnapping](#).<sup>1</sup>

[Eldad Regev](#) was captured during [Hezbollah cross-border assault](#).<sup>JP<sup>2</sup></sup>

In 2006, [Eldad Regev's](#) occupation was [soldier](#).

▼ **Detailed Justification:**

- ▶ [The Lebanese Hezbollah](#) was a perpetrator in [Hezbollah cross-border assault](#).<sup>1</sup>
- ▶ [Hezbollah cross-border assault](#) was a [kidnapping](#).<sup>1</sup>
- ▶ [Eldad Regev](#) was captured during [Hezbollah cross-border assault](#).<sup>JP<sup>2</sup></sup>
- ▼ [Eldad Regev](#) is a [soldier](#) in 2006.
  - ▼ In 2006, [Eldad Regev's](#) occupation was [soldier](#).  
Entered by [Matthew Kovner](#) on Dec 13, 2006.

▼ **External Sources:**

<sup>1</sup>  [TKB Fact Entry Source Sit Report IDF Spokesman Hizbullahattackonnorthernborderand IDFresponse](#)

<sup>2</sup> **JP** [the July 13, 2006 issue of The Jerusalem Post and The Jerusalem Post's newspaper article](#)

The user can drill down to see as much of the details as he/she wishes. Ultimately, the justification drill-down terminates in the date that fact was entered and the name of the person who entered it.

The CAE also includes a specialized interface for creating network analysis or “related-to-via” queries. These sorts of queries allow the users to explore via visualization complex association networks. The following screen shot shows the interface that can be used to quickly generate a related-to-via query.

How can  <click to enter term>

be connected to  <click to enter term>

by at most  links using:

**Possible relations:**

- actor
- affiliate
- leader
- member
- performer
- relative
- religious teacher
- staging
- teacher
- thing that is somehow linked to some

choose >

remove <

**Chosen relations:**

Add

The following screenshot shows the tool being used to create a query for Iranian persons connected to Hezbollah via a link of “leader” associations that is at most 3 links long.

How can

be connected to

by at most  links using:

**Possible relations:**

- actor
- affiliate
- leader
- member
- performer
- relative
- religious teacher
- staging
- teacher
- thing that is somehow linked to something else

choose >

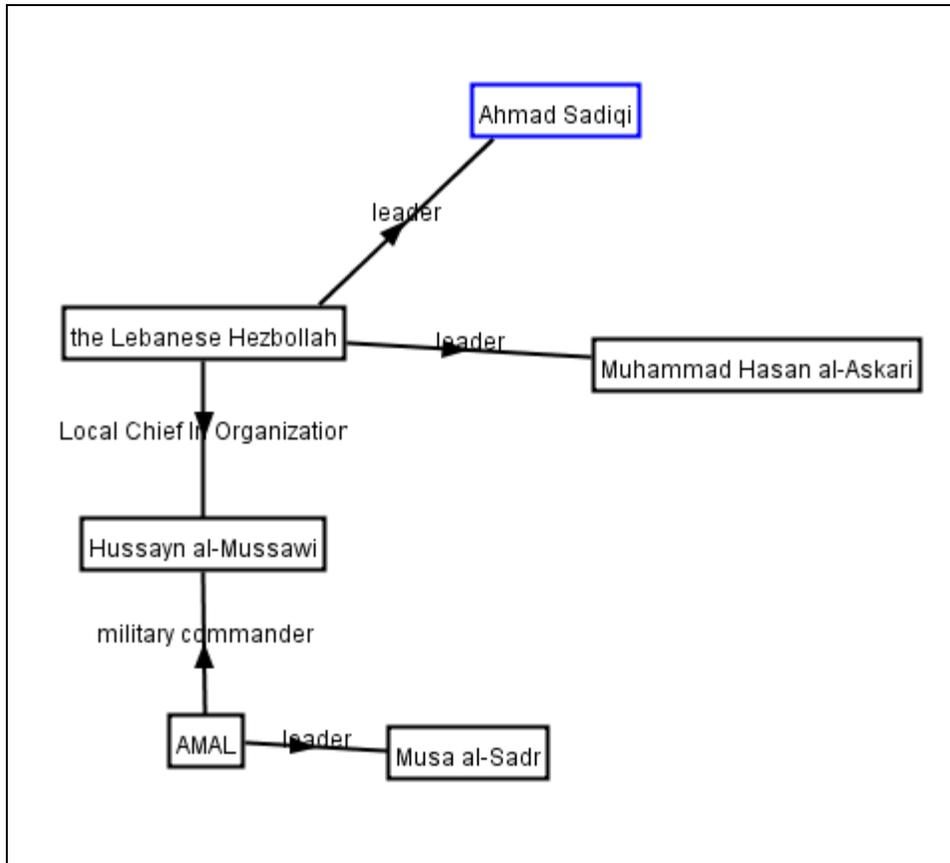
remove <

**Chosen relations:**

- leader

Add

The user can then import the query into the query construction panel and further edit it or ask it as is. The user can then generate a graph visualization of the answers returned. See the following screenshot for an example of this.



The system also has the capability to use Analyst’s Notebook™ to generate these social network visualizations.

In addition to graph visualizations of social networks, the system is also capable of generating timeline and bar chart visualizations as appropriate. The implementations of these additional visualizations are somewhat rough and included for “proof of concept” purposes only. Exporting the information from the TKB/CAE into a software product dedicated to producing visualizations as we did with the social network graphs is the preferred deployment solution.

## 4 Collaboration

During the course of the TKB project, Cycorp shared the content and functionality of the TKB with various other organizations engaged in research for the United States Government. This collaboration ranged from generating in database format the contents of an imported and mapped set body of information to more involved technology integration experiments (TIEs). We indirectly collaborated with 21<sup>st</sup> Century Technologies during the course of this project when a different project within Cycorp generated a database of the information imported into the TKB from our mapping of Marc Sageman's Excel spreadsheet of individuals involved in the 911 terror attacks. This information was then used in 21<sup>st</sup> Century's research into pattern matching in social network graphs.

More formally, Cycorp and the TKB were involved in an official technology integration experiment with Aptima, Inc. and Kathleen Carley, Ph.D., Carnegie Mellon University. Aptima, Inc. is developing NEMESIS, a counter terrorism tool integration and analyst collaboration environment that focuses on tools and data sources that look at terrorist organizations as networks of people, knowledge, resources, locations, events, tasks, and other organizations. Two tools that have been integrated into this environment are Organizational Risk Analysis (ORA) tool from Carnegie Mellon University and the Adaptive Safety and Monitoring (ASAM) tool from the University of Connecticut.

We first provided to Aptima an OWL export of TKB content that consisted of a small number of persons, terrorist events, and terrorist organizations and a sampling of the various relations and attributes involving those individuals. Initially restricting the size of the export allowed us to study which of the widely varied TKB data types was best suited for use by NEMESIS. Based on feedback from this initial study we took into consideration the sort of links that Nemesis's Organizational Risk Analysis (ORA) component could best utilize when we produced the next export.

For the second export, we started with a list of 373 individuals that were mentioned in Marc Sageman's Matrix database of social network data on the 911 conspirators. We had earlier mapped that database into the TKB, thus integrating it with any information on those individuals already present in the system. The SAIC subject matter experts (SMEs) already entered a significant number of assertions about most of these individuals. Each of the individuals in the Matrix database was involved in between 50 and several hundred assertions. Since links between individuals were of prime importance in doing dynamic network analysis, we decided to concentrate on various types of "personal association relations" that involved the individuals mentioned in the Matrix. Personal association relations are binary relations that relate instances of the class of persons. Examples (written in Cyc's native representation language CycL) include `#$religiousTeacherOf`, `#$businessPartners`, `#$subordinates`, etc. The assertions are tagged and a java extraction program is then run. The java extraction program gathered these identified assertions and grouped the contained terms by type: predicate, collection and individual. OWL is an XML syntax for describing and transmitting ontological information, given certain expressiveness limitations (e.g. not first order predicate calculus).

For each Sageman individual, the identified assertions were converted into OWL format. The extracted OWL version of the example assertion looks like this:

```
<AdultMaleHuman rdf:ID="Terrorist-Salim">
  <rdfs:label xml:lang="en">
Mamdouh Mahmud Salim
</rdfs:label>
  <guid>
dff74888-a901-41d8-9051-ea6e5432b01a
</guid>
  <boss rdf:resource="#OsamaBinLaden"/>
</AdultMaleHuman>
```

The resulting file contained 2,926 lines of OWL, which were then transformed to ODL for further processing and use by their social network analysis.

Further information regarding the experiment can be found in the following report (copies of which are available by request from the authors, Stacy Webb., Chris Deaton, and Kathleen Carley) written for the 2005 International Conference on Intelligence Analysis, McLean, Virginia: "Transforming a Terrorism Knowledge Base for Use by Network Organization Analysis Tools: A Case Study."

In addition to the collaborations indicated above, there are currently several requests pending for evaluation copies of the TKB and for exported data dumps of the TKB terrorism domain content.

## 5 Outside Expert Evaluation

The TKB and CAE underwent evaluation by Research and Development Experimental Collaboration (RDEC) during December of 2006. We reprint here the Executive Summary of the final version of the report written by David Fado titled “Capability Review for Cycorp Analyst Environment”

**“A team of RDEC analysts reviewed the Cycorp Analytic Environment tool (CAE) to assess capabilities related to rapid knowledge acquisition for intelligence analysts and automated reasoning about terrorist events, terrorist organizations, and individual terrorists. The capability assessment provides an opportunity for early review of tools to provide feedback that can have maximum influence over tool development. The assessment also provides an opportunity to review scenarios for the use of the tool in a classified environment.**

**The assessment included a review of a CAE scenario and script that provided insight into the organizational structure of Hezbollah up to April 2006. The metrics team prepared selected Factiva news feeds that provided events from May to August 1 2006 in Lebanon related to the Israeli/Lebanon battles. These new events would be used during the scenario. Three RDEC technology analysts participated in the two day scenario focused on Hezbollah structure and terrorist activities. These analysts initially used the script to guide their introduction to Hezbollah, showing capabilities related to rapid knowledge acquisition. This generated questions about Hezbollah strategy and tactics as the new events emerged during the Summer of 2006. For this scenario, the analysts found the CAE helped them answer many of the questions posed about Hezbollah and the new events of 2006. In terms of analyst feedback, the tool most successfully demonstrated the capability for rapid knowledge acquisition. The tool also showed impressive capabilities for answering basic questions related to terrorist events. However, when these events crossed over into international political or economic events, CAE could provide limited or no information. CAE needs more effective mechanisms for assessing the breadth and quality of the information delivered to the analyst.**

**This assessment finds CAE contains promising technology for rapid knowledge acquisition and analysis of events. Cycorp should continue to improve CAE for more effective interfaces with common analyst tools, such as Analyst’s Notebook. Cycorp will also need to provide better mechanisms for run-time assessment of data quality. CAE should remain on the Development Platform (DP) and continue to look for classified interest in an experimental scenario. With pull from a client, the RDEC DP can do additional work to help prepare for that scenario.”**

The full RDEC report is available by request.

In addition to the RDEC evaluation of the TKB, during the latter part of the contract, the terrorism domain experts at the Terrorism Research Center (TRC) used the full TKB/CAE system to enter and retrieve knowledge about Hezbollah. We reprint here a short note detailing their experiences with the TKB/CAE written by Vice President for Research James T. Kirkhope at the end of this project:

“TRC’s work on the project began in the initial meetings that helped formulate the composition of the profiles, determining the necessary categories and connectors for individual terrorist, terrorist group, and terrorist incident profiles. Subsequently, TRC analysts researched a variety of terrorist groups across the political spectrum (HAMAS, Al Qaeda, Hezbollah, Irish Republican Army, Salafist Group for Call and Combat, ETA, etc.) and performed data entry of relevant information in an effort to populate the system. After approximately one year, Cycorp directed TRC to focus their research solely on Hezbollah. The driving notion was that one heavily populated terrorist group would be ideal for demonstrations of TKB’s functionality and value. TRC performed exhaustive research and data entry for the Hezbollah organization, individual members, affiliates, and attacks. TKB is populated with information beginning with Hezbollah’s progenitors in the late 1970s up through its 2007 standoff with the Lebanese national government. In between, Hezbollah’s two wars with Israel, global terrorist activity, support for Palestinian terrorism, and involvement in the current Iraq conflict are all thoroughly documented.

During this process, TRC’s technical and analytical refinements submitted via weekly and later monthly logs have significantly helped shape the profile structure and data representation. These refinements include countless adjustments and additions in the individual to individual, individual to group, and group to group connections with the aim of improving the database’s viability and efficiency for an analyst. To that end, TRC also experimented with the user interface, running multiple tests with various terms and queries that tested the speed and ability of the database to locate and clearly represent information in a timely manner.

The current configuration of the user interface is far more efficient and user-friendly than its original version. The ability to enter relevant terms that return usable queries is more practical than the earlier, laborious process that required a more formulaic approach to searching the database. In addition to multiple internal tests of the system with different queries, TRC also utilized the system for a separate open-source research project for a U.S. government client that focused on IED makers. TKB was able to quickly locate over 40 individual terrorist profiles that had expertise with explosive devices or had served as an explosives expert for a terrorist organization. The process of having to copy and paste the profiles into a word doc was time consuming, and I’d suggest a print function ultimately be added to the TKB. However, considering that 5 months of TRC research found approximately 150 IED maker profiles, the ability of the TKB system to produce 40 additional profiles of terrorist with explosive expertise makers was impressive.

In sum, the TKB has great potential as a tool to assist an analyst or law enforcement/military official in an investigation of terrorist individuals, groups, or events. The searchability of the system still needs improvement, though it has come great distances in terms of ease for the user. Moreover, TKB’s relevancy as a tool requires that the data be constantly updated. Without new information, the system loses its value to the analyst.”

## 6 Conclusion

The TKB had as its objective to remedy a critical gap in US Intelligence: the absence of a comprehensive knowledge base about terrorist groups, individuals, and events. By “knowledge base here” were meant that the content be represented formally, in logic (and numbers), so that a machine can mechanically deduce (and arithmetically produce) the same entailments to a set of assertions as would a human being, given those same assertions. We considered it a serious gap, i.e., that there is no such knowledge base in existence, nor even such a *database* – i.e., a comprehensive terrorism database that “bottoms out” in only in structured fields (rather than being allowed to contain a large number of opaque English or other natural language sentences and paragraphs).

In one sense, this project was a success. The research and development effort produced a comprehensive knowledge base ontology and schema, formulated by and agreed upon by panels of internationally recognized terrorism experts; produced a methodology for non-logician subject matter experts to directly enter assertions into the KB (via the Fact Entry Tool) without having to learn anything about logic, AI, Cyc, or programming; and produced a corresponding terrorism KB and user interface (for analysts to formulate *ad hoc* queries) that received a very positive RDEC and TRC (Terrorism Research Center) evaluation.

But in another sense, due to premature curtailment of funding for this effort, it had to focus, in its final year, differentially on one entity (Hezbollah) and its members and activities. A dictionary or almanac or encyclopedia which is only half complete is not nearly as useful as -- one which is complete. So, in conclusion, the TKB was a technical success, but in its current form it is only half complete, and less than half as useful as it would be if it were truly comprehensive, as it was originally designed and scoped to be.

## 7 Appendix I: Initial Terrorism Representation Schema

What follows is the initial characterization of the types of information the TKB is capable of representing. Note that not all of the schema elements are populated in the TKB. The content entered was limited by the restriction to open source data. But the vocabulary is present in the system, so all of the following types of information could, in principle, be entered by users of the system or mapped from databases for which the TKB has a defined schema.

### Information about terrorist groups

#### (1) Classifying the overall terrorist group:

- Broad Ideological type
  - e.g., religious organization
  - political group or party
- Relationship to an organized government
- State-associated (intelligence corps, etc.)
- Rebel/Dissident group
- Criminal organization
- Global reach
- Global vs regional/guerilla

#### (2) Group leaders (current and former)

- Specific leadership role or responsibilities within group
  - e.g., head of training
- Dates joined and left (if latter)
- Information on leader as individual (see above)

#### (3) Group members (current and former)

- Position or responsibilities within group
- time periods during which held positions
- Dates joined and left (if latter)
- Information on member as individual (see above)

#### (4) Others with leading roles within group (may or may not be members)

- Group founders
- Date founded group
- Group mentors, e.g., spiritual leaders
- Group spokespersons

#### (5) Group Structure and Sub-Organization (Group Family Tree)

- Group Predecessor: a group from which current group developed or broke off

Overall form:

Spoke-and-hub -Used in this category is this term referencing structure (i.e. cells and subs) as well as links?

Cells

Sub-organizations

List all group features that apply -- leader, members, ideology, etc.

Group Offshoot: a group that broke off from current group

(6) Group Factions

List all general features of groups that apply to the faction -- leader, members, ideology, etc.

(7) Other organizations affiliated with group

Political organizations

Nature of relationship

Political wing of group?

Charitable organizations

Directly funds group?

Other terrorist organizations

Nature of relationship

Supports/Directs/Tasks group

Umbrella organization including group

Ad-hoc association with group

\*Note that some affiliations may be derived from information about group members' affiliations & contacts.

(8) Group sponsors and providers of support

Who they are

(typed as state, criminal, etc.)

What stage of operations they support

recruitment

training, etc.

What type of active support they provide

Political

Financial

Physical Resources

territory or safe haven

facilities

weapons, etc.

Training

Intelligence

Logistics

What state sponsor permits  
Permitting terrorist command & control to operate within state  
Permitting training within state  
Permitting smuggling of materials into/through state  
Permitting financial activity  
Giving sanctuary or asylum to group members

NOTE: may be derived from events

(9) Group ideology & goals

Religious Affiliations  
Anti-Western, Anti-American, Anti-Country X, Anti-Global, Anti-Christian  
Political or Militant  
Nationalist / Separatist  
Islamic Militant Fundamentalist  
Marxist  
Maoist  
Fascist

(10) Enemies and other agents towards which group is hostile

\*Note that some hostile relations may be (defeasibly) derived from group ideology and goals.

(11) Recruitment and Indoctrination

Source of recruits  
Legal / illegal immigration/ asylum/ immigrant communities  
Marginalized and disfranchised in society  
Unstable political regimes  
Prison  
Religion and Education (i.e. mosques, madrasses)  
Familial relations

Indoctrination Methods

Isolation, friendship, “brainwashing”, no outside media  
Motivation and persuasion for suicide tactics  
Group morale and maintenance  
Incitement of ideological fervor

(12) Internal Security/Discipline

Process for internal security  
Security officer identification  
Methods  
Investigations

Infractions  
Punishments

(13) Group locations

Geographical locations in which group is known to:

recruit  
train  
reside  
conduct attacks and other operations

\*Note that some locations may be derived from movements of group members and locations of attacks.

(14) Material resources of group

Possessions  
Type of resource  
Weapons  
Weapon-izable material  
Facilities  
Communication equipment  
Attempts to acquire  
Type of attempt  
R&D  
Procurement  
Type of resource  
Successful?

(15) Group Finances

Known assets, including businesses owned  
Sources of funding, including financial sponsors  
Movement of funds  
individuals involved  
institutions involved  
means of money transfer /hiding funds (i.e., cash to commodities)

(16) Group Behaviors & Tactics

History and tendencies in all stages of operations:  
Intelligence Operations  
Attacks  
Types of targets  
Types of weapons

- Preferred means of delivery
- Single or multiple targets
- Degree of coordination between attacks
- Choosing recruits for operations
- Training
  - Type of training
  - Location
  - Duration
- Other preparatory actions (entering country, setting up cells)

(17) Group capabilities

With respect to each stage of a terrorist operation, i.e. recruitment, training, planning, execution

Capability to mount different types of attacks, e.g.

- cyber-attacks
- dirty bombs
- attacks on communication networks, or other critical infrastructure systems

\*Note that some capabilities may be derived from known group actions

(18) Group intentions

(19) Information Operations/Media  
Perception management

**Information about individual terrorists**

- Name, including alternate spellings
- Aliases
- Birth date
- Death date, if any
- SSN, if any
- Nationality
- Ethnicity
- Citizenship
  - birth
  - naturalized
  - dual
- Country of residence
- Current location
- Recruitment history -- where, by whom
- Training history -- where, by whom
- Movements/Travel History
- Disappearances (time?)
- Expertise
- Occupation
- Education
- Fields

## Institutions

Affiliations with organizations other than terrorist groups  
(universities, religious groups, businesses)

Personal associations and contacts with members of other organizations, suspected members, and connected individuals (roommates, business acquaintances, prison friendships, influential marriages, unclear relationships)

Criminal Record: arrest record, currently in custody, warrants, sentences served and in absentia

Position and responsibilities in terrorist organizations (see above)

Roles in previous terror attacks (see below)

## Information about terrorist attacks and activities

Note that this category includes foiled or aborted attacks as well

### (1) Type of attack

Types of illegal acts committed, e.g.,

Killing

Hijacking

Kidnapping

Cyber-attack

Broad classification by weapon-type, e.g.,

Shooting

Bombing(car, truck, and boat bombs)

Bombing with hazardous materials

Sabotage of hazardous materials

Chemical weapon/Gassing

Biological weapon

Radiological device

New / alternate/ emerging weapons: Lasers, Radio Frequency Weapons

Suicide attack?

### (2) Who planned and directed the attack

Planner

Persons, groups, or both

Specify stage(s) of operation planned, if different

persons or groups handled different stages

Initiator of attack, if different from planner

Direct Action/ General call for action / self-motivated attack

Issuer of call

### (3) Who carried out the attack

Persons, groups, or both

Accomplices

Stage(s) of attack involved in

Overall roles in event: leaders/deputies/soldiers

(4) Who claimed responsibility for the attack

when  
via what medium

(5) Target

Intended target -- possibly a subset of, or even different from,  
actual victims and structures damaged (Collateral damage)

Victims

Identities

Type of injury or damage suffered, e.g.,

killed

wounded

taken hostage

Number suffering each type of injury/damage

Organizations they belong to:

(government, military, companies, civilian)

Position in above organization

Ethnicities

Inanimate objects damaged or destroyed

Owner or operator

Other agents w/ interests in objects damaged or destroyed

Monetary value of damage

Type of inanimate object: (cultural, symbolic, economic, military, religious,  
political, general civilian)

(6) Weapons Used

(7) Means of delivering weapon

(8) Location

Geographical location

Type of fixed structure or vehicle

(9) Date

(10) Preparatory actions

Sub-events of the operation:

Operational Planning

Intelligence Operations

Surveillance

Acquiring weapons and other resources

Movements of perpetrators, including entry  
into target country or restricted area

Delivery of weapons to target

\*Note: For each, specify the directors and performers, when Known.

- (11) Precipitating events
  - e.g., killings, arrests, military movements or actions
- (12) Other related activity, e.g.,
  - Threats by groups involved (before or after)
  - Actions which incite terrorist activity
    - fat was
    - calls to action
    - other leader statements
  - Arrests of suspected participants or plotters (including arrests which foil a planned attack)
  - Increased communication chatter
- (13) Links to other terrorist activity
  - Other attacks coordinated with
  - Larger campaign subsuming
- (14) Does this attack involve a repeat target for group?
- (15) Does this attack indicate a change in tactics, weapons, targets, or delivery?
- (16) Operational Tempo
  - Time between attacks of similar size and complexity
  - Time between small and large attacks
  - Number of small attacks preceding large attacks
  - Reconstitution / Regeneration after a major attack
  - Reconstitution / Regeneration after legal, financial, military, diplomatic, law enforcement, intelligence and covert action responses to a specific attack by particular countries
- (17) Adaptation
  - After responses to attack

### **Information Source**

#### (1) Classification of Source

- Overall Type, e.g.,
  - Person
    - Position or type (spokesman, government official, etc.)
  - Newspaper
  - Periodical
  - Web site

Relational database

(2) Source Author, where applicable

(3) Source Date, where applicable

Date the information was published or posted.

(4) Source Reliability, where known

Reliability of source overall

Reliability of author

## 8 Appendix II: Listing of English Glosses of represented terrorism domain queries.

The following is a listing of the English glosses for all represented TKB query templates. In the actual templates, most of the major concepts involved are replaceable with other concepts. For example, in the query described by the following gloss,

“Which terrorist groups have carried out more than half of their attacks in Israel?”

the concepts referred to by these words in the gloss “terrorist groups”, “carried out”, “more than half”, “attacks”, and “Israel” can all be replaced with different, but similar concepts resulting in a distinct query that asks a similar question about different entities. So, the query corresponding to the above gloss could be turned into a query described by the following sentence:

“Which state-sponsored terrorist groups have directed less than 40% of their car bombings in Lebanon?”

### Query Glosses

Does North Korea have capability and motive for a missile attack on Japan?

Who has capability and motive for an missile attack against Jordan that targets an aircraft?

Who had means and motive for car bombing Rafik Hariri?

Find agents with the capability and motive for a missile attack on Israel targeting an aircraft.

Find agents with capability and motive for carbombing Rafik Hariri.

What agents have capability and motive for launching an attack on Jordan?

Did any senior Hizballah leader travel to the southern cone of Latin America between Hizballah's 1992 and 1994 attacks on Israeli targets in Buenos Aires, Argentina?

Who was accused of the attack on a Christian church in Zuk, Lebanon in February, 1994?

What military commander for the Chechen rebels with links to al-Qaeda died in 2002?

Was Hizballah responsible for the attack on a Christian church in Zuk, Lebanon, in February, 1994?

What terrorist groups have used motorcycles in staging a terrorist attack?

Has Hizballah ever used motorcycles in staging a terrorist attack?

Give the year and target of every terrorist attack that Hizballah has staged in Thailand

How many terrorist attacks did Hizballah stage in Western Europe between 1985 and 1995?

Hizballah has had operatives living in what Canadian cities, during what time periods?

Hizballah has had cells in what Canadian cities, during what time periods?

Find everyone who was a terrorist who died in Ayn al-Hilwah refugee camp in 2003, and who could be connected to Al Qaeda through two simultaneous affiliations or fewer.

What targets were destroyed in the 1994 Hizballah attack in Buenos Aires, Argentina?

What Israeli targets were destroyed in the 1992 Hizballah attack in Buenos Aires, Argentina?

Find all terrorists who died in Ayn al-Hilwah refugee camp in 2003, and who can be connected to Al Qaeda in two affiliations or fewer.

What is the most specific organizational position Ibrahim Aqil has held in Lebanese Hizballah?

What organizational positions has Ibrahim Aqil held in Lebanese Hizballah?

What organizational leadership positions has Ibrahim Aqil held in Lebanese Hizballah?

Find all persons who were relatives of Hassan Nasrallah and who were members of Lebanese Hizballah, and who were killed during a conflict event between Hizballah and Israeli forces.

Was any relative of Hassan Nasrallah killed as part of a fighting event in which Israeli forces and an organization in which the relative was a member were in conflict?

What is the latest year at which the IRGC is known to operate in Lebanon?

What is the earliest date at which it is known to be true that Osama bin Laden resides in Saudi Arabia?

What is the name of the suicide bomber who attacked the Jewish Community Center in Buenos Aires, Argentina in 1994?"

What is the name of the suicide bomber who attacked the Israeli Embassy in Buenos Aires, Argentina in 1992?

Did Imad Mughniyah meet Osama bin Laden in Sudan in the 1990s?

List events in which Imad Mughniyah collaborated with an agent of the Iranian government.

Who is Imad Mugniyeh's brother-in-law?"

When and where was the most recent hijacking in which Imad Mughniyah was involved?

Find all humanitarian organizations that give some amount of support to components (suborganizations, citizens, leaders, etc.) of Iran, that have also given some measure of support to terrorist groups at some time in the past, together with the components, groups, and support levels involved.

Find all government organizations, terrorist groups, and charities, such that the government organization is able to affect the charity, and some part of the charity (member, suborganization, or the charity itself) is affiliated with some terrorist organization.

Find government organizations, charities, terrorist organizations, and support metrics such that the government organization is able to affect the charity, and some extension of the charity (agent or suborganization, or the charity itself) provides that measure of support to the terrorist organization.

How many InstanceNamedFn-Ternary NARTs are linked with at least one term via possiblyIdentical-NeedToReview in the PotentiallyIdenticalConceptsCleanupMt?

Find all Iranians to whom Ibrahim Aqil can be linked in 2 steps or fewer using the relation of affiliation and its specializations.

For every key member of Lebanese Hezbollah, find the number of Iranians to whom said member of Lebanese Hizballah can be linked in two steps or fewer, using the relation of affiliation and its specializations.

Find the key member of Al Qaeda who can be linked to the largest number of terrorists who are positive interests of the Iranian Government in two steps using the deliberate action relation and its specializations.

For every key member of Al Qaeda, give the number of terrorists in whom the Iranian government takes a positive interest, and to whom the Al Qaeda member can be linked, in two steps using the relationship of deliberate action and its specializations.

Find all terrorists in whom the Iranian government can be proved to have a positive interest and to whom Mohammed Ibrahim Makkawi can be linked in two steps using the deliberate action relation and its specializations.

Find the key member of LebaneseHizballah who, out of all key members of LebaneseHizballah can be linked to the highest number of Iranian persons in four steps or fewer using deliberateActors, #organizationHasKeyMembers, subOrganizations, and their specializations.

For each member of Lebanese Hezbollah, enumerate the number of Iranian persons to which that individual can be linked in 4 steps or fewer using deliberateActors, #organizationHasKeyMembers, subOrganizations, and their specializations.

Find all Iranian persons to whom Abu Mahadi Najafi can be linked in 4 steps or fewer using deliberate actors, organization key members, sub-organizations, or their inverses.

Find the key member of LebaneseHizballah who, out of all key members of LebaneseHizballah can be linked to the highest number of terrorist attacks in two steps or fewer using hasMembers, performedBy, and their specializations.

For each member of Lebanese Hizballah, enumerate the number of terrorist attacks to which that individual can be linked in 2 steps or fewer using hasMembers, deliberateActors, and their specializations.

Find all terrorist attacks to which Imad Mughniyah can be linked in 2 steps or fewer using the

relations hasMembers, deliberateActors, and their specializations.

What agents can be found such that there is evidence to support the hypothesis that they assisted in Al Qaida's 1998 bombing of the US Embassy in Tanzania?

What agents can be found such that there is evidence to support the hypothesis that they assisted in the 2000 attack on the USS Cole?

What agents can be found such that there is evidence to support the hypothesis that they assisted in Al Qaida's 1998 bombing of the US Embassy in Nairobi, Kenya?

What kind of support to terrorist organizations other than Al Qaeda is known to have been given by countries whose governments have members who have met with members of Al Qaeda?

What government organizations are known to have members who have been in meetings with known members of Al Qaida?

What public officials that Al Qaeda is not currently known to be allied with might be allies of Al Qaeda.

What paramilitary organizations that are not currently known to be allies of Al Qaeda might be allies of Al Qaeda?

What commercial organizations that are not currently known to be allies of Al Qaeda might be allies of Al Qaeda?

What government organizations that are not currently known to be allies of Al Qaeda might be allies of Al Qaeda?

Who has perpetrated what events in Israel in which an Israeli was killed?

Who has perpetrated what attacks in Israel in which an Israeli was killed?

What bombings between 1998 and 2000 targeted places of business?

What car bombings that took place in Sri Lanka had a person as a maleficiary?

What types of things have been damaged in Sendero Luminoso attacks?

What bombings damaged public transportation devices between February 2000 and July 2003?

What car bombings took place in Sri Lanka and had at least one person as a maleficiary?

List any terrorist attacks in which somebody was wounded in that attack and was the transportee in a medical movement event.

List the bombings that occurred on January 28th 2004.

List the ratio of suicide bombings to regular bombings by terrorist groups that operate in Afghanistan.

For each asserted instance of AttackType what ratio of Hamas's attacks are of that type and what ratio of those attacks are performed in Israel

Who was the perpetrator of the attack that wounded the most people in Israel?

Who is the probable perpetrator of the October 19th, 2000 terrorist attack in Colombo, Sri Lanka?

For each major attack type, what is the ratio of Hamas attacks that are of that type?

What do Al Qaida and Jihad Group have in common?

Who was a spokesman for Hizballah at 18 seconds after 6:05 PM, March 4, 2005?

In what European cities have key members of Hezbollah resided?

Which (past or present) members of Hizballah were terrorists in the year 1800?

What percentage of car bombings in Israel were carried out by Islamic nationalists?

Is the ratio of suicide bombings carried out in Israel by Hamas to all terrorist acts carried out by Hamas in the period starting with the year 2000 greater than the ratio in the period before the year 2000?

What percentage of suicide attacks in Israel after 2000 were carried out by national independence groups?

Which types of terrorist attacks have been carried out in Israel since 2000 (inclusive), and for each of those types of attack, what is the ratio of attacks of that type to overall attacks?

Which terrorist groups have carried out more than half of their attacks in Israel?

Who were the spokespersons for Hezbollah?

In the year 2000, was it the case that Osama Bin Laden was acquainted with Imad Fayeze Mugnyah?

In 2001, where did the Al Qaida Hamburg cell used to operate?

What terrorist attacks occurred one day before (the start of) what Jewish holidays?

What terrorist attacks occurred one day before the start of Passover?

What car bombings took place in India on what holidays?

What terrorist attacks occurred on (starting dates of) Jewish calendar holidays?

What terrorist attacks occurred on what Jewish holidays after 1999?

Which terrorist attacks occurred on Israeli Independence Day?

Which terrorist attacks occurred on the (starting) dates of which holidays

Which terrorist attacks temporally intersect which holidays?

What terrorist attacks occurred on Jewish holidays?

Which terrorist attacks happened on Islamic holidays?

Which terrorist attacks occurred on a date that was temporally subsumed by a known instance of Ramadan?

Who or what is related via three steps or fewer to Mustafa Kamel through the relations teacherOf, actors and affectedAgent?

Test2- list the terrorists related to Terrorist-Karroum via acquaintedWith and 3 steps.

List the terrorists that Karroum is related to via acquaintedWith and 3 steps.

The RTV query the includes the link between the crash in PA and the letter copies.

Are Zacarias Moussaoui and Osama Bin Laden linked?

Are Zacarias Moussaoui and OsamaBinLaden linked?

List the terrorists that are linked to Zacarias Moussaoui via the set {acquaintedWith deliberateActors affiliatedWith}.

List the things related to Ayman Al-Zawahiri via the relations of membership and perpetrator or their inverses to 2 steps or fewer.

List the people Mohamed Atta is related to via the predicates: containsInformation deliberateActors eventOccursAt hasOwnershipIn inRegion possesses.

List the individuals that can be linked to Bill Clinton via acquaintedWith and deliberateActors.

List all terrorists that can be connected to Osama Bin Laden through a chain length of no greater than three through affiliatedWith and acquaintedWith

Between what times did Aum Supreme Truth perform what types of acts and where?

list the locations where Hamas is based and when.

List all known members of Abu Sayyaf Group.

What is the political wing of Chukaku-Ha?

List the types of things that Ansar Al Islam possesses.

List the agents who have given supplies to Lebanese Hezbollah and what those supplies were.

List the agents that have supported Hamas and what type of support.

List the attacks in which ETA was the perpetrator.

List the events in which FARC was the directing agent.

List the types of acts that FARC performs.

List all leaders of FARC.

What is the current ideology/belief system of FARC?

How many members does FARC have?

List all known goals of FARC

List all claims FARC is known to have made.

List the types of weapons that FARC uses.

List the agents FARC considers an enemy.

List the interesting information about FARC.

List all information from binary relations with ETA in the first argument.

When was ETA founded?

List the collections of which FARC is asserted to be an instance.

List the comments on Revolutionary Armed Forces Of Colombia.

List the places where Ocalan has lived and the times he lived there.

List the schools at which Mohamed Atta was enrolled and when.

in what country is Ayman Al Zawahiri a citizen via birth?

Where did Al Banna die?

Where was Ayman Al Zawahiri born?

When was Al Banna born?

List the interval during which Al Banna is asserted to have died.

List all things that Nawaq Alhamzi is known to possess.

List the agents Osama Bin Laden has supported and the type of support.

List the occupations Faris has had and when.

What are the comments on Faris?

What organizations was Attah a member of and when?

What is the full name of Al-Zawahiri

List all belief systems or religions in which Al-Zawahiri is known to subscribe.

List the academic fields in which Al-Zawahiri has received formal education.

List the countries in which Al-Zawahiri has citizenship.

List all the things Al-Zawahiri is related to via a binary relation where Al-Zawahiri is the first argument of that relation.

What is the ethnicity of Al-Zawahiri?

List all known aliases of Al-Zawahiri.

List the immediate collections of which Al-Zawahiri is an instance.

How many of what type we wounded in TerroristAttack-Manila?

How many things and of what type were killed in TerroristAttack-669?

Who planned TerroristAttack-Manila?

Who was the directing agent of TerroristAttack-Manila?

List all macro relations, relations, and the things that stand in these relations to TerroristAttack-670.

List the collections of which Terrorist-669 is asserted to be an instance.

Who was killed in TerroristAttack-669?

Who was the perpetrator of TerroristAttack-669?

What was the intended target of TerroristAttack-669?

Where did TerroristAttack-669 occur?

When did TerroristAttack-669 occur?

List all information from binary relations where TerrorAttack-669 is in argument 1.

What hostage takings have occurred in Colombia?

What percentage of the bombings performed by Hamas are suicide bombings that occur in Israel?

What members of Lebanese Hezbollah are members of other groups as well?

list the likely perpetrators of the February 5th, 2004 terrorist attack in Gaza City.

Give the known times after 1995 that Titi was the leader of the Al Aqsa Martyrs Brigade.

What is the ratio of terrorist attacks performed by Hamas that are performed in Israel?

During what time periods is MEK-MKO known to have resided in Paris?

What terrorist groups have carried out attacks in eastern European countries?

Which terrorists speak two or more languages?

Which terrorists speak Arabic and English?

Which terrorist agents operate in English-speaking regions?

Which terrorist agents operate in Arabic-speaking regions?

Which terrorist groups operate in Spanish-speaking regions?

In what western European countries have terrorist attacks occurred?

What terrorist attacks occurred in Paris between 1993 and 1997 (exclusive)?

Are there any suborganizations of Al Qaida that operate in Egypt and have carried out terrorist attacks in Paris?

Does Al Qaida have any suborganizations that operate in Egypt?

List all known events in New York state that happen in September 2001

In what U.S. cities have terrorists resided?

To what ethnic groups do terrorists belong?

In what terrorist agents do both Iraq and Iran have a positive vested interest?

List all people who were leaders of The Al Aqsa Martyrs Brigade after 1995.

List the attacks and perpetrators such that the attack killed at least one Israeli and the attack occurred in Israel.

Who has performed kidnappings and bombings in Iraq?

What types of people have been killed in assassinations?

What assassinations have taken place in Portugal or Spain?

What percentage of Al Qaida bombings are suicide bombings?

What percentage of the attacks by state sponsored terrorist groups are bombings?

List the attacks perpetrated by Hamas in which Israeli persons were targeted.

List the islamic terrorist group that has wounded the most Israeli persons

What terrorists were born in Asia?

In what countries do Islamist terrorist groups reside?

What have terrorists attempted to do in Greece?

What terrorists have the string 'ahm' in (one of) their names?

List all kidnappings in which some sort of United States person is captured

List all bombings that occurred between 1998 and 2000 (inclusive) in which some place of business was the intended target of the attack.

List all bombings where it is known that a law enforcement officer was killed during the attack.

List all bombings that occurred after 1997 and occurred in countries that are monarchies.

List all relevant events that have occurred in countries not diplomatically recognized by the U.S.A. that occurred later than 1997.

List all terrorist groups who have perpetrated acts where there are at least two attacks such that they are less than 2 days apart.

List the Islamic Jihad organization with the highest total wounded in their attacks.

List the terrorist group that operates out of Pakistan that has the highest casualty count.

Who is affiliated with a terrorist group that has carried out attacks in India?

Give the total number of people wounded in attacks by the terrorist group People Against Gangsterism And Drugs.

What types of businesses have been the intended targets of terrorist attacks?

What terrorist agents have carried out assassinations in Europe?

What cities suffered Terrorist Attacks in 2003?

What groups have carried out suicide bombings since 9-11?

What schools have terrorists attended?

List when and where the first known terrorist attack by Sendero Luminoso occurred.

List all attacks on government organizations that were later than 1997 and perpetrated by a terrorist group with more than ten attacks since 1997.

List all terrorist organizations that perform both suicide bombings and kidnappings.

List all terrorist organizations that have perpetrated attacks in multiple middle eastern countries.

What percentage of the people killed in attacks perpetrated by the Abu Nidal Organization were

killed in Italy?

How many people are known to have been killed in attacks perpetrated by the Abu Nidal Organization

How many people have been killed in attacks perpetrated by Abu Nidal in the Middle East?

What percentage of kidnappings in Israel were perpetrated by Hamas?

List the terrorist group that perpetrated the most number of attacks in which someone was injured.

Who has been taken hostage in terrorist attacks in Germany?

What terrorist attacks happened in Canada during the 1980s?

List the terrorist group who has performed the most suicide bombings.

For each terrorist organization that is known to have perpetrated a terrorist suicide bombing list the number of such bombings they are known to have perpetrated.

What is the number of known deaths from Hamas attacks?

List the attacks in the middle east in which the perpetrator is unknown

List all attacks and their perpetrators such that the attack targeted a type of workplace.

List all attacks in which it is known that some type of tourist was captured.

For each asserted type of tourist list the number of attacks in which an agent of that type was captured.

List the terrorist attack in which the most people were captured and its perpetrator.

List the kidnapping attack and its perpetrator such that the attack had the highest number of wounded persons

List the events in which a diplomat was kidnapped in Yemen.

List the deadliest attack that took place in Israel and its perpetrator.

List the types of things that Hamas has targeted.

List the attacks such that Hamas targeted some type of building in the attack.

How many attacks are such that Hamas targeted a type of building?

What terrorist attacks happened on the same date as what diplomatic events?

Does Djamel Beghal believe that Muhammad is a prophet?

Does Djamel Beghal believe that Jesus is not God?

Who wants Osama bin Laden to approve of him/her?

What are the bombings that targeted some building?

Is Mullah Mohammed Omar Jewish?

Who is not known to be dead and is affiliated with a terrorist group that has performed suicide bombings?

Who is affiliated with a terrorist group that has performed a suicide bombing?

Is Djamel Beghal a Muslim?

Does Djamel Beghal believe in Islam?

Does Djamel Beghal believe In Sunni Islam?

In what cities have multiple suicide bombings occurred?

In what countries have multiple suicide bombings occurred?

What terrorist groups have performed multiple suicide bombings in the same country?

What belief systems do (some) terrorist agents have?

List all attacks whose intended target was a type of site holy to some religion.

List all attacks and their perpetrators such that the attack was a suicide bombing in the middle east and its target was a school bus.

Who was the perpetrator of the attack that killed the most people in Israel?

Does Osama bin Laden believe in pacifism?

List all terrorist attacks that have killed at least one person and it is known that the device used was some type of mailable object.

To what non-terrorist groups do (some) terrorists belong?

Does Abu Zubaydah like George W. Bush?

Whom does Abu Zubaydah hate?

List all terrorist groups that, according to the testimony of Jamal Ahmed Al-Fadl, are able to control a company which has an employee that is also an employee of an intelligence agency.

What non-weapon device types have been used in terrorist attacks?

Who is linked to Zacarias Moussaoui via three or fewer iterations of acquaintance and/or affiliation and are not known to be dead?

List all attacks and countries such that the attack is known to have killed at least one soldier of that country.

What types of things have been damaged in terrorist attacks carried out by Sendero Luminoso?

What percentage of bombings carried out by Al Qaida are suicide bombings? (old)

How many terrorist suicide bombings have taken place since September 11, 2001?

What percentage of terrorist attacks are suicide bombings?

What terrorists who are not known to be members of Al Qaida are affiliated with Osama bin Laden or are affiliated with some agent that is affiliated with bin Laden?

What terrorists who have been in Afghanistan are not known to be dead?

Who has studied medicine and is either a terrorist or is affiliated with a terrorist or terrorist group?

List the attack, organization, and type of things such that the attack damaged that type of thing and an organization whom the U.S.A. has a vested interest in owns the thing damaged.

What terrorists are known to be members of two or more terrorist groups?

What terrorists are members of two or more terrorist groups?

List the attacks, dates, and perpetrators of any bombings that are known to have damaged public transportation devices.

List the known persons who have studied law that are members of a terrorist group that has been helped at least once by a state sponsor of terrorism.

List all known individuals that have studied law that have been members of terrorist groups that have been known to use harmful chemicals in an attack.

List the bombings that targeted embassy buildings.

List the known bombings in which the performer has claimed responsibility.

List the terrorist suicide bombing perpetrated by Palestine Islamic Jihad in Israel.

List all bombings after 2000 and before September 2001 that used pipe bombs.

List all Terrorist Attacks that were after the year 2000 that targeted government buildings in capital cities of greater than one million residents

List all known attacks that occurred in Israel between January 2000 and September 2002 that damaged a transport facility.

List the terrorist groups that control companies that own deadly objects.

List the countries that are capable of being the directing agent of TerroristAttack-420.

Which groups have carried out multiple bombings on a single day?

Which groups have carried out multiple, lethal attacks on a single day?

List the types of targets that ETA has attacked.

List the countries and attacks such that country was at least partially responsible for the attack.

List the commercial organizations and terrorist sponsors such that the sponsor can control the organization.

List the terrorist attacks that used bombs that involved banks in 1999

List all attacks that occurred in the 1990s and targeted restaurants in which at least 1 person was killed.

List known sponsors of terrorist groups.

Between March and April 2002, in Colombia, what percentage of kidnapping attacks directed at public officials were perpetrated by FARC?

Between March and April 2002, in Colombia, what percentage of kidnapping attacks had public officials as their targets?

Between March and April 2002, in Colombia, how many kidnappings occurred that targeted

civilians?

Between March and April 2002, in Colombia, how many kidnappings occurred that targeted professionals other than public officials, and which types of professionals were targeted?

In What country has HAMAS carried out at least 50% of its attacks?

What individuals are affiliated with terrorist groups engaged in drug trafficking?

What percentage of bombings in Northern Ireland were committed by the IRA?

List all countries affiliated with anti-American terrorist groups.

In the period 1998-2002, which groups perpetrated attacks that caused a total of over 100 deaths?

How many deaths has HAMAS caused in attacks from 1998-2002?

List all terrorist groups in Northern Ireland that have used bombs and list the specific types of bombs that each has used.

What was the average number of days between suicide attacks by Palestine Islamic Jihad in 2001 and 2002?

Has GRAPO carried out at least 50% of its attacks in any one city?

Which groups have carried out multiple, lethal bombings on a single day?

Which groups have carried out multiple attacks on a single day?

List all groups who have perpetrated attacks that resulted in over 10 deaths.

How Many hostages were taken in terror attacks between March and July 2002.

List all individuals affiliated with anti-American terrorist groups.

List all attacks such that the attack is known to be an adversarial response to a military event performed by the United States.

How many events that occur in Colombia between March and April 2002 are known to be kidnappings/hostage-takings?

Between March and April 2002, in Colombia, how many kidnappings occurred that targeted public officials?

How many deaths have the Basque Fatherland and Liberty group caused in attacks from 1998-2002?

List all known attacks on public officials in Latin America during July 2002.

List all acts perpetrated by Palestine Islamic Jihad.

What events occurred in Israel?

## **Builder Query Glosses**

In addition to the “pre-formed” queries described above, the TKB’s “query library” contains almost 200 simple “builder queries” (queries that are designed to be combined with each other to form new, more complex, queries). All of the variable terms (the words in all capitals in the glosses) can be replaced by any entity in the system of the correct type, e.g. AGENT-1 can be replaced with “Hamas”, or “Hezbollah”, etc. Forming combinations of conjunctions and disjunctions of these builder queries allow users to build up new queries. Let’s go through an example of how such a query could be build up from the provided builder queries. Consider the following query:

“What terrorists with skill in bomb making were members of organizations who performed kidnappings in Israel”

The user could generate this by combining the following fragments

PERSON has SKILL (at some time)

THING is an instance of TYPE

PERSON was a member of ORGANIZATION  
AGENT performed instances of TYPE in PLACE

By a combination of the system combining and unifying the fragments and the user resolving any ambiguities in the unification process and the replacement of certain terms with others, the user would build up their query in a step-by-step fashion. After the initial unification step the query would look as follows.

PERSON has SKILL  
PERSON was a member of ORGANIZATION  
ORGANIZATION performed instances of TYPE-1 in PLACE  
THING is an instance of TYPE-2

Since “THING is an instance of TYPE-2” is such a general fragment applicable to a large number of different types of entities, the user will have to manually select the entity whose type they wish to restrict. In the example we are formulating, they would unify PERSON with THING resulting in

PERSON has SKILL  
PERSON was a member of ORGANIZATION  
ORGANIZATION performed instances of TYPE-1 in PLACE  
PERSON is an instance of TYPE-2

At this point, the user would replace certain variable terms with concepts to finish the query. In this case, SKILL would be replaced with “bomb making”, TYPE-1 would be replaced with “kidnapping”, PLACE would be replaced with “Israel,” and TYPE-2 would be replaced with “terrorist” resulting in

PERSON is skilled in bomb making  
PERSON was a member of ORGANIZATION  
ORGANIZATION performed instances of kidnapping in Israel  
PERSON is a terrorist

At this point, the query could be asked and any values for PERSON and ORGANIZATION that satisfied the constraints would be displayed for the user.

### **Listing of Builder Query glosses**

AGENT-1 supplied AGENT-2 with TYPE (at some time).  
AGENT-1 supplied AGENT-2 with TYPE throughout TIME.  
AGENT-1 provided safe haven to AGENT-2.  
AGENT-1 provided training to AGENT-2.  
AGENT-1 gave support of TYPE to AGENT-2 throughout TIME.  
AGENT-1 gave support of TYPE to AGENT-2 (at some time).  
ORGANIZATION operates NUMBER facilities of TYPE in PLACE.  
ORGANIZATION-1 was merged into ORGANIZATION-2.  
ORGANIZATION-1 is a successor organization of ORGANIZATION-2.  
ORGANIZATION was founded at PLACE.

AGENT committed crimes of TYPE in PLACE (at some time).  
AGENT committed crimes of TYPE in PLACE throughout TIME.  
AGENT performed instances of TYPE in PLACE (at some time).  
AGENT performed instances of TYPE in PLACE throughout TIME.  
AGENT operated in PLACE (at some time).  
AGENT operated in PLACE throughout TIME.  
ORGANIZATION resided in PLACE (at some time).  
ORGANIZATION resided in PLACE throughout TIME.  
The force of the explosion in ATTACK (in units of TNT) was MASS.  
EXPLOSIVE-TYPE was used in ATTACK.  
AGENT-1 denies that AGENT-2 performed ATTACK.  
Someone has claimed responsibility for ATTACK.  
Responsibility for ATTACK was claimed on DATE.  
AGENT claims responsibility for ATTACK.  
AGENT claimed responsibility for ATTACK on DATE.  
AGENT-1 claimed responsibility for ATTACK to AGENT-2.  
AGENT-1 claimed responsibility for ATTACK to AGENT-2 on DATE.  
PERSON-1 worked with PERSON-2 in ORGANIZATION (at some time).  
PERSON-1 worked with PERSON-2 in ORGANIZATION throughout TIME.  
PERSON has STATUS at SCHOOL (at some time).  
PERSON is unskilled at SKILL throughout TIME.  
PERSON is a novice at SKILL throughout TIME.  
PERSON has SKILL throughout TIME.  
PERSON is an expert at SKILL throughout TIME.  
PERSON is competent at SKILL throughout TIME.  
PERSON is unskilled at SKILL (at some time).  
PERSON is a novice at SKILL (at some time).  
PERSON has SKILL (at some time).  
PERSON is an expert at SKILL (at some time).  
PERSON is competent at SKILL (at some time).  
PERSON has a degree of TYPE in FIELD (at some time).  
PERSON has a degree of TYPE in FIELD throughout TIME.  
PERSON's highest education level is LEVEL (at some time).  
PERSON's highest education level is LEVEL throughout TIME.  
PERSON was in hiding throughout TIME.  
PERSON was in hiding after EVENT (at some time).  
PERSON was in hiding after EVENT throughout TIME.  
PERSON was in hiding in PLACE (at some time).  
PERSON was in hiding in PLACE throughout TIME.  
PERSON was in hiding after EVENT in PLACE (at some time).  
PERSON was in hiding after EVENT in PLACE throughout TIME.  
PERSON was in hiding (at some time).  
PERSON was imprisoned in PLACE throughout TIME.  
PERSON was imprisoned in PLACE (at some time).  
PERSON was imprisoned throughout TIME.  
PERSON was imprisoned (at some time).  
PERSON is a naturalized citizen of COUNTRY (at some time).  
PERSON is a birth citizen of COUNTRY (at some time).

PERSON is a citizen of GEOPOLITICAL-ENTITY (at some time).  
PERSON is a birth citizen of COUNTRY throughout TIME.  
PERSON is a naturalized citizen of COUNTRY throughout TIME.  
PERSON is a citizen of GEOPOLITICAL-ENTITY throughout TIME.  
PERSON-1 was acquainted with PERSON-2 (at some time).  
PERSON-1 was acquainted with PERSON-2 throughout TIME.  
PERSON-1 and PERSON-2 were relatives (at some time).  
PERSON-1 and PERSON-2 were relatives throughout TIME.  
AGENT-1 was affiliated with AGENT-2 (at some time).  
AGENT-1 was affiliated with AGENT-2 throughout TIME.  
PERSON was a spokesperson for ORGANIZATION (at some time).  
PERSON was a spokesperson for ORGANIZATION throughout TIME.  
PERSON was a leader of ORGANIZATION (at some time).  
PERSON was a leader of ORGANIZATION throughout TIME.  
PERSON was a member of ORGANIZATION (at some time).  
PERSON was a member of ORGANIZATION throughout TIME.  
AGENT-1 bore the personal association relation RELATION to AGENT-2 (at some time).  
PERSON had ROLE in ORGANIZATION (at some time).  
ORGANIZATION employed PERSON (at some time).  
ORGANIZATION employed PERSON throughout TIME.  
PERSON had OCCUPATION (at some time).  
PERSON had OCCUPATION throughout TIME.  
PERSON has STATUS at SCHOOL throughout TIME.  
PERSON was in PLACE (at some time).  
PERSON was in PLACE throughout TIME.  
PERSON resided in PLACE (at some time).  
PERSON resided in PLACE throughout TIME.  
PERSON has AGE at TIME.  
PERSON is of GENDER.  
PERSON is of NATIONALITY.  
What types of persons were killed in ATTACK?  
Which full years occur between the start of World War II and the end of the Vietnam War?  
Which years occur between the start of World War II and the end of the Vietnam War?  
AGENT possesses some instance of TYPE.  
PERSON is a founder of ORGANIZATION.  
ORGANIZATION was founded on DATE.  
PERSON was exposed to a harmful substance in ATTACK.  
PERSON was injured in ATTACK.  
PERSON was tortured in ATTACK.  
PERSON was assassinated in ATTACK.  
TRANSPORTER was hijacked in ATTACK.  
THING was a possible intended target of ATTACK.  
THING was contaminated in ATTACK.  
TEXT is a description of EVENT.  
ACT is an unsuccessful attempt to perform an instance of ACT-TYPE.  
ATTACK is an instance of ATTACK-TYPE.  
THING-1 is linked to THING-2.  
PERSON's weight is WEIGHT.

PERSON's height is HEIGHT.  
PERSON's hair color is COLOR.  
PERSON's eye color is COLOR.  
AGENT believes in RELIGION.  
How many people have ever been a member of ORGANIZATION?  
ORGANIZATION-1 is a political wing of ORGANIZATION-2.  
ORGANIZATION has members of TYPE.  
Some initials designating THING are INITIALS.  
AGENT is named NAME.  
NOTE is a cyclist note about THING.  
AGENT1 works offsite for AGENT2.  
AGENT1 is a leader of AGENT2.  
PERSON is of ETHNICITY.  
PERSON speaks LANGUAGE.  
PERSON is an expert regarding TOPIC.  
AGENT has DEGREE in FIELD.  
PERSON has identification of type IDTYPE with description/number STRING.  
AGENT believes in BELIEFSYSTEM.  
AGENT was a birth citizen of COUNTRY.  
PERSON died at PLACE.  
PERSON died on DATE.  
AGENT's age is AGE.  
PERSON was born in PLACE.  
PERSON was born on DATE.  
PERSON's preferred name is NAME.  
ATTRIBUTING attributed responsibility for EVENT.  
DEVICE was a device used in EVENT.  
ATTACK was perpetrated by some TYPE.  
PERSON's alias is NAME.  
PERSON's former name is NAME.  
PERSON's middle name is NAME.  
PERSON's given name is NAME.  
PERSON's family name is NAME.  
NUM DEVICE-TYPEs were used in ATTACK.  
AGENT was a key participant in EVENT.  
EVENT was planned by AGENT.  
AGENT was a directing agent of EVENT.  
AGENT was a deliberate social participant in EVENT.  
AGENT was an assisting agent in EVENT.  
ATTACK was perpetrated by NUMBER terrorists.  
NUMBER persons were intended victims of ATTACK.  
NUMBER persons were captured in ATTACK.  
NUMBER instances of TYPE were damaged in ATTACK.  
ANIMAL was wounded in EVENT.  
PERSON was killed in ATTACK.  
PERSON was an intended victim of ATTACK.  
PERSON was captured in ATTACK.  
ATTACK damaged THING.

ATTACK destroyed THING.  
NUMBER of TYPE were destroyed in ATTACK.  
AGT has the background knowledge required to learn to perform acts of ACT-TYPE  
ATTACK damaged some TYPE  
ORG was founded on DATE.  
NUMBER of TYPE were targeted in ATTACK.  
ATTACK targeted THING.  
ATTACK wounded NUMBER of TYPE.  
ORG resides in GEOGRAPHICAL-AGENT  
THING-1 is the same thing as THING-2.  
THING-1 is different from THING-2.  
DATE-1 is later than DATE-2  
ATTACK used a device of TYPE  
ATTACK killed NUMBER of TYPE.  
ATTACK was perpetrated by AGENT.  
EVENT occurred on DATE  
EVENT occurred at PLACE.  
THING is a car bombing.  
THING is a bombing.  
THING is a terrorist suicide bombing that utilizes a portable nuclear device.  
THING is a terrorist suicide bombing.  
THING is a terrorist attack.  
THING is an instance of TYPE.  
Date1 starts later than Date2  
City is the capital city of Country  
TYPE-1 is a more specific kind of TYPE-2.  
Smaller geo-entity is a part of the territory of Larger geo-entity  
In Attack, the number of Type-Damaged damaged is Number  
What is the death toll of persons in ?ATTACK1?  
Attack is an adversarial response to Event  
In which event is the United States government a deliberate actor?  
What type of thing was the intended target of Attack?  
Shorter time is subsumed temporally in Longer time  
NUMBER persons were killed in ATTACK.

## 9 Appendix III: Characterization of TKB content

In general, the content of the TKB is focused on 3 types of entities:

### **terrorist attacks, terrorist groups, individual terrorists**

Almost all (99.99%) of the represented attacks in the system have date information (down to the granularity of a particular day), location information (typically at the level of city or village), and information pertaining to the type of attack it was (bombing, kidnapping, etc.).

### **Locations of the represented attacks**

We have more information about Middle East terrorist attacks than any other region and this information is more recent than our information about other areas. We know of about 1000 in Israel, 650 in Lebanon, 558 in the areas of the West Bank and Gaza strip, 316 in Iraq, and 156 in Egypt. Although not strictly in the Middle East, we also know of about 500 attacks in Turkey. We also know of about 243 in Pakistan, 763 in India, and around 100 in Afghanistan. The next most populous region is South America. We know of about 787 attacks in Colombia, 359 in Peru, and some in Argentina and Brazil. We also know about a significant number of European attacks. We know about a substantial number of IRA and other Irish nationalist terrorist group attacks in Great Britain. We also know a significant number of Basque Fatherland and Liberty (ETA) attacks in Spain, Italy, and other places. Much of our information about European, and South American terrorist attacks is more historical in nature – reaching as far back as the 70s but with little depth other than the basics. The exceptions are our information about the Revolutionary Armed Forces of Colombia (FARC) and ETA attacks.

### **Type of attack information**

Roughly ¼ of the attacks are known to have killed some people.

Slightly more than a quarter of them are bombings

Slightly more than 10 percent of them are known to be kidnappings or hijackings

### **Information present on a significant percentage of attacks in the system:**

Just under half of the terrorist attacks represented in the system have specific #PerformedBy or #Perpetrator

We have #DirectingAgent information for about 4.3% of the represented attacks (~600)

We have human casualty information -- what types of persons were killed, wounded, kidnapped, targeted, and how many for just over half of the terrorist attacks represented in the system. To drill down a bit, we have information on what types of persons were killed on about 35% of the attacks. We have information about what sorts of persons were wounded or injured in about 27% of the attacks. Of the 1373 kidnappings represented in the system, over 50% have some sort of information about the type or number of persons captured.

By far, the most assertions reference Person as the casualty type, but we often have more specific information regarding the nationality of the casualties especially with respect to UnitedStatePerson and IsraeliPerson. We also have a good amount of information regarding the occupation type of the casualties of the attack, e.g. whether they are a politician, soldier, police officer, etc. But it is not known if this information is complete with respect to attacks in any given area.

We have less information about particular persons who were killed, wounded, or captured in an attack. Around 25% of the kidnappings have information about the particular persons captured (as opposed to just the information that some number of a certain type were captured). This amounts to roughly 500 persons explicitly represented as being captured in some attack. Of these, just over half have just minimal information – their name and the fact that they were captured in that attack.

In about 6% of the attacks, we have specifically represented the person who was killed. This amounts to roughly 1650 persons represented in the system as being killed in some attack or other. Of these 1650 representations, roughly 1/3 have minimal information regarding the victim – date of death, name, and the fact that they were killed or assassinated in the attack. We only have specific persons who were injured or wounded in attacks represented in around 2% of the attacks. It is believed that the ratios of explicit representation of individuals to merely representing the information at the type level (e.g. 7 tourists were killed) correspond to the open source data used by the TKB SMEs. For example, if we only explicitly represent the person who was killed in an attack in 1/5<sup>th</sup> of the attacks in which we have that information at the type level, this is because in only about 1/5<sup>th</sup> of the news reports and open source documents the SMEs use do they provide specific information about the victim such as name and so forth.

There is a wide variety of target types that these attacks are said to have targeted. There are over 300 different types that are represented as being the target type of some attack or other. A good number of these representations involve moderately general CycL concepts like GovernmentalBuilding, CommercialBuilding, Village, TransportationDevice, ModernShelterConstruction, RealEstate, etc. But quite a few are spread out among the 300 or so specializations of these general concepts.

In about 19% of the represented attacks we have information about the sort of thing targeted. About 7% of the attacks have information about the sort of thing destroyed, and in roughly 17% of the represented attacks we have information about the sort of thing damaged.

As with the human casualty information we have less information about the specific objects targeted, damaged or destroyed. In about 7% of the attacks we have a representation of the specific object damaged. This amounts to about 1100 specific representations of particular objects. Unfortunately, 99.9% of these representations have minimal additional information. Most of these are generally represented as generic instances of HumanlyOccupiedSpatialObject that have a certain name. This is likely due to the fact that these individuals are often described and not named. For instance one attack is asserted to have damaged an instance of SpatialObject named “Karni (Qarni) border crossing”. We did not have a prior representation of that particular object and, in

this case, failed to understand that string as referring to a border crossing. In other cases, we were able to extract more explicit type information. In one case, we understood a use of “Iraqi police convoy” to refer to some instance of Convoy. Our ability to partially understand strings that refer to particular unrepresented individuals has only recently been improved to the point where it is possible to determine type information from partially understood strings so much of the information that is now trapped in strings like those described is, in fact, convertible to a more meaningful (Cyc) representation.

We have specific information about the thing targeted for about 9.5% of the represented attacks. This amounts to around 1000 particular representations of individual targets. These representations are, for the most part, minimal with much information currently trapped in strings in terms like (InstanceNamedFn-Ternary “Bayt Hanun crossing” PartiallyTangible GUID\_STRING).

In about 1% of the attack, we have a representation of the specific thing destroyed (roughly 14% of the cases where we have some type-level information). This amounts to around 160 particular representations of destroyed objects. Again, since these objects are usually described and not named, these representations generally have minimal information attached to them.

In a smaller percentage, we have information about what sort of weapons were used in the attack. In roughly 23% of the attacks we have information about the sort of device used during the attack. Much of this information is at the granularity of “handguns”, “IED”, “Rockets” and not at the granularity of “M1911 Colt 45”. This information is distributed across over 400 types of devices and weapons. About 75% of those types are underrepresented in that there is type information present in the string used to create the concept (strings that denote concepts that we fail to parse to an appropriate collection are used as an argument to # \$ProperSubcollectionNamedFn-Ternary to reify the concept at assert time). As with terms created with InstanceNamedFn-Ternary, we can attempt to re-parse the strings used to create these terms and improve the representation as our ability to extract useful information from strings that the system doesn’t fully understand improves.

### **Information present on little to no attacks in the system:**

We have no information about the particular time of day in which an attack occurred nor any information about its location at a granularity finer than city or village. But in 90% of the attacks we do know the specific city or village in which it occurred.

We have no information about the internal structure of most of these terrorist attacks -- e.g. we don’t have any information about pre-attack preparations or plans, or about any surveillance performed. We also don’t have much information about the number of perpetrators that were present during a given attack.

We know little about individuals who played some role in an attack. For example, for some given attack X, we may know it was said to be perpetrated by Hezbollah and we may know that 7 terrorists were involved, but we are not likely to know who those terrorists were except in high publicity cases like 911 and other attacks that are widely

discussed in the open source media. To be precise, of the attacks for which we have some perpetrator or directing agent information (roughly around 7000) only 166 of the attacks have this information about particular persons. It is not known whether the lack of this information in the TKB is due to its absence in open source documents or the knowledge enterers' reticence to attribute those sort of information to individuals when the source itself doesn't represent it at some acceptable level of certainty (e.g. instead of "officials say the Mughniyah was involved" it is reported at a lower level of certainty as in "Some reports suggest that Mughniyah was involved." Even though we have little information about individual terrorists who play roles in attacks, at least with respect to structured information, in this aspect we have greater coverage than the MIPT database which doesn't even have a field for individuals' roles in attacks. What information they do have on this subject is in unstructured text fields as comments on particular terrorist incidents.

### **Completeness of terrorist incident coverage:**

The following figures are based on a comparison with the MIPT terrorist KB.

The TKB has 74% of the attacks represented by MIPT that occurred during the 1970s – a total of about 1600 (Note the MIPT data doesn't take into account domestic attacks between 1968 and 1997), for the 1980s the TKB has almost 100% of the attacks represented by MIPT – around 3400, and the TKB has around 88% of MIPT represented attacks that occurred during the 1990s – around 4150. The TKB has around 24% of the 2000-2006 attacks that are present in the MIPT database – around 4686. Note, it is fairly clear that MIPT is most complete in the years between 1998 and now. So the fact that we represent around 100% of the attacks that MIPT represents for the 1980s should not be taken to mean we are complete with respect to actual attacks that occurred in the 1980s since only major attacks were reported in western open source media prior to the mid 1990s.

In the Middle East/Persian Gulf region, MIPT has roughly 9000 attacks represented for the time period later than 1999. We have roughly 1400 represented for that same time period -- roughly 30% of the attacks we have represented that occurred after 1999 occurred in the Middle East. This is less than the 24% coverage that we have with respect to attacks in all regions after 1999. But this is likely due to the tremendous number of attacks in Iraq they have represented – around 9000 to our 240. If we remove the attacks in Iraq from consideration, then we have around 36% coverage in the Middle East/Persian Gulf region (not including Iraq, but including Turkey) from 2000 through early 2006 (the date when the SMEs last entered knowledge into the system).

Drilling down some, we have around 70% coverage of attacks in Israel during this time period and close to 80% coverage of attacks in Lebanon. We have more attacks that occurred in Syria during this time period represented than the MIPT database has. Our representation of attacks that occur in the West Bank and Gaza during this time is about 25% complete with respect to MIPT. These ratios hold, in general, for randomly selected time periods subsumed by the time period 2000 through now.

## **Information present on Terrorist Groups**

In general, we represent what attacks the groups are said to have perpetrated. We represent their membership and generally have some notion (some precise, some vague) of when those people were members. We have information about what sort of weapons they are armed with. Unfortunately, that information is not very precise and probably not very useful. We generally have information about when they were founded, what areas they operate in, what sort of attacks they perform, and where their headquarters are located.

Hizballah is the group we have the most information about, ~5k substantive assertions. Hamas is next with about 2.5k assertions with about the same amount of information about al Qaida. We have a good amount of information about the Palestinian Islamic Jihad, Lashkar e Tyyiba, Jemaah Islamiyah, and the Abu Nidal Group. We also know a good number of things about fairly recent activities (~ 2000 to 2004) of al Fatah.

Comparing our data on Hezbollah to MIPT, we find that we have almost twice as many attacks attributed to Hezbollah than MIPT does (298 to 179). We have around 380 members of Hezbollah explicitly represented. MIPT provides no easy way to generate a list of all past and present members. They do represent 14 of their leaders and 26 of their members who have been indicted.

In general, our information about membership is fairly incomplete. But, for those organizations listed, we do have a significant amount of information about their leaders and often about their founders and other senior members. If MIPT is approaching completeness in this area with respect to unclassified/open sources, then the TKB is complete in this respect as well. In general, we have much more information about a groups leaders, members and affiliations than MIPT provides.

TKB knows nothing about their recruiting habits, but does have some information about persons who are the recruiters for a particular organization.

## **Information present on individual terrorists and other persons of interest**

There are a significant number (just over 100) of individuals for which we have over 100 assertions about. Of these people, we often have information about the groups they belong to and when (though the “when” is often vaguely stated) and what sort of role they play in the organization – leader, intelligence officer, explosives expert, etc. TKB also has a large number of affiliation assertions between individuals and between individuals and groups. To a much lesser extent, we have specific information about the type of affiliation (other than membership and leadership information).

We almost always have information about the individual’s birth date, birth place, ethnicity, and nationality. We generally have some information about their education level and expertise, but little information about when and where they went to school. TKB sometimes knows a little about their family relations, but not as a rule. It also knows a smattering of facts about the roles these individuals played in particular attacks.