

Natural Language Processing (NLP)

10/30/02

NLP Definition

- a range of computational techniques

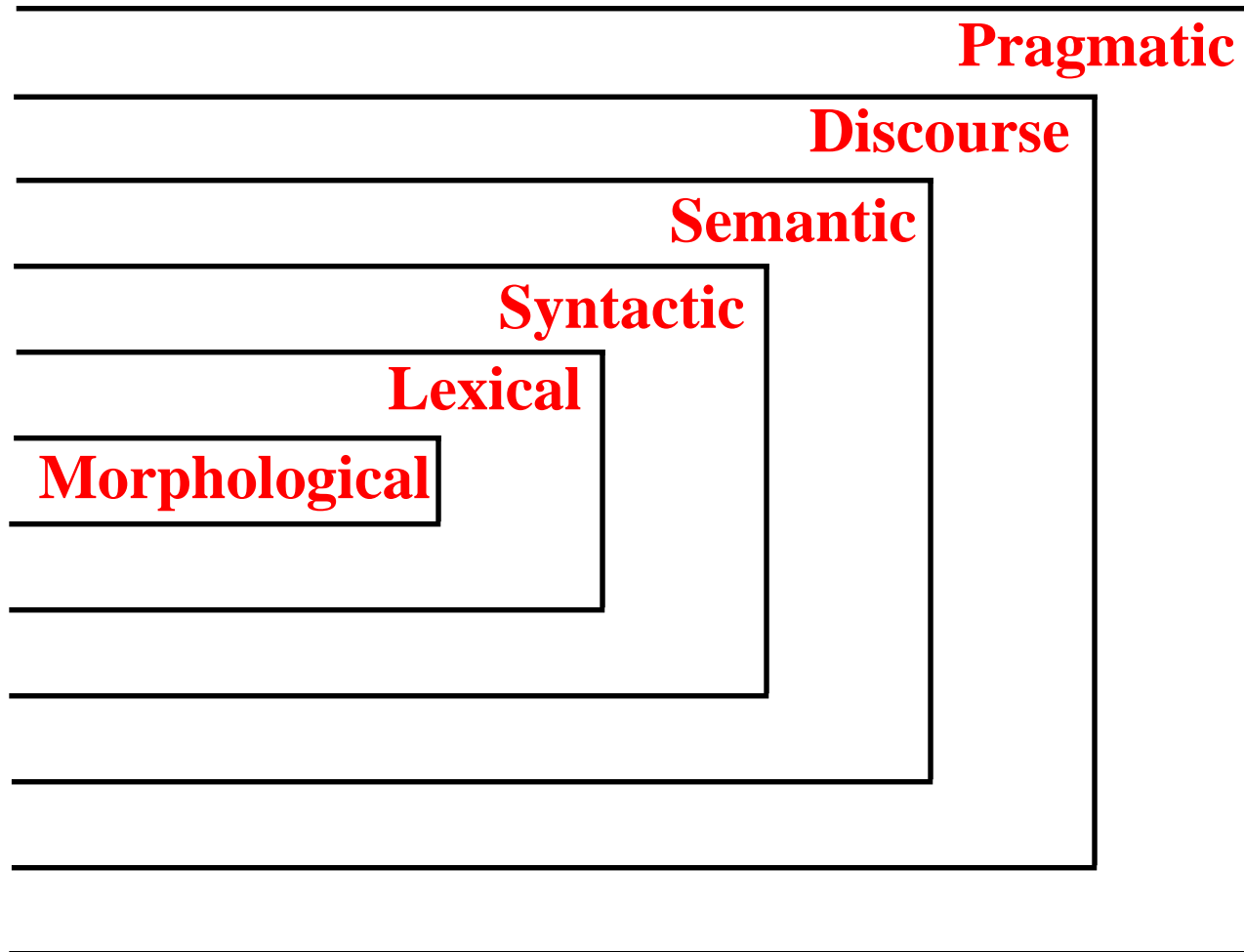
NLP Definition (cont'd)

- a range of computational techniques
- for analyzing and representing naturally occurring texts

NLP Definition (cont'd)

- a range of computational techniques
- for analyzing and representing naturally occurring texts
- at one or more levels of linguistic analysis

Levels of Language Understanding



NLP Definition (cont'd)

- a range of computational techniques
- for analyzing and representing naturally occurring texts
- at one or more levels of linguistic analysis
- for the purpose of achieving human-like language processing

NLP Definition (cont'd)

- a range of computational techniques
- for analyzing and representing naturally occurring texts
- at one or more levels of linguistic analysis
- for the purpose of achieving human-like language processing
- for knowledge intensive applications

Goals of Information Extraction

- **A robust information extraction system**

Goals of Information Extraction

- **A robust information extraction system**
 - Recognize concepts and the implicit relations amongst them

Goals of Information Extraction

- **A robust information extraction system**
 - Recognize concepts and the implicit relations amongst them
 - Convert **vast** amounts of textual data into a semantic representation

Goals of Information Extraction

- **A robust information extraction system**
 - Recognize concepts and the implicit relations amongst them
 - Convert **vast** amounts of textual data into a semantic representation
 - **Provide knowledge discovery tools for multiple analyst activities**

Goals of Information Extraction

- **A robust information extraction system**
 - Recognize concepts and the implicit relations amongst them
 - Convert **vast** amounts of textual data into a semantic representation
 - Provide knowledge discovery tools for multiple analyst activities
 - visual exploration
 - data-mining via NLP queries
 - link analysis

High Level Task Description

- Evaluate the application of automatic knowledge extraction to link analysis

High Level Task Description

- Evaluate the application of automatic knowledge extraction to link analysis
 - Specialization of generic relations
 - Prototype IE to Link Analysis tool

High Level Task Description

- Evaluate the application of automatic knowledge extraction to link analysis
 - Specialization of generic relations
 - Prototype IE to Link Analysis tool
- Identify current technological barriers

High Level Task Description

- Evaluate the application of automatic knowledge extraction to link analysis
 - Specialization of generic relations
 - Prototype IE to Link Analysis tool
- Identify current technological barriers
- Establish high-payoff research directions

High Level Task Description

- Evaluate the application of automatic knowledge extraction to link analysis
 - Specialization of generic relations
 - Prototype IE to Link Analysis tool
- Identify current technological barriers
- Establish high-payoff research directions
- Produce substantive report on current state-of-the-art

KNOW-IT Overview

- Automatically identifies and extracts concepts and relations involving people, events, places, and organizations, etc from massive volumes of digital textual data
- For purpose of building / adding to Knowledge Bases for use by human & automated reasoners
- General technology capability
 - currently used for various text types & domains
 - can be specialized for specific applications

KNOW-IT's Building Blocks:

Natural Language Processing

+

Knowledge Extraction

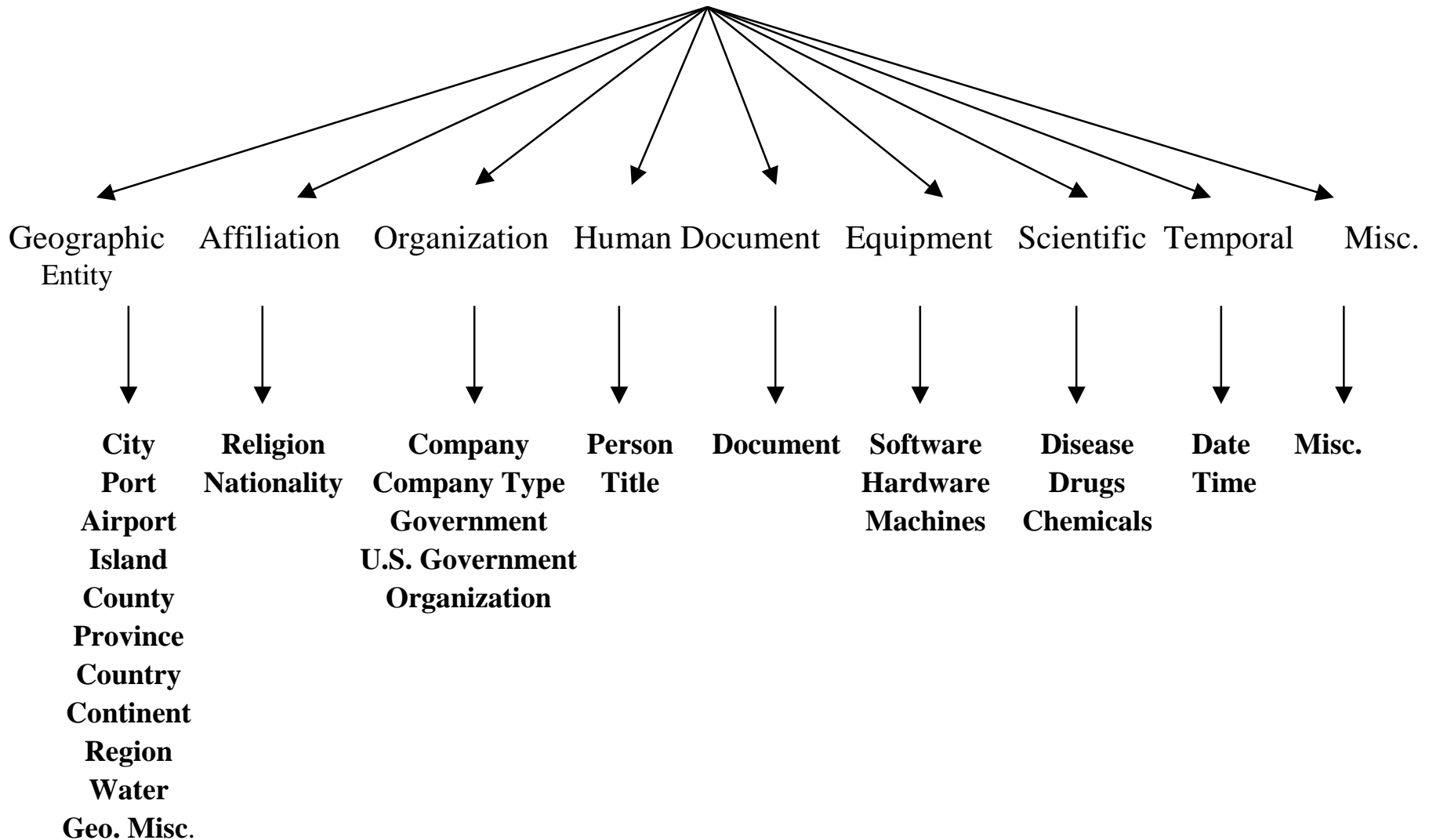
+

Graphical Visualization

KNOW-IT components

- Concepts
 - 60 + Proper Noun Categories

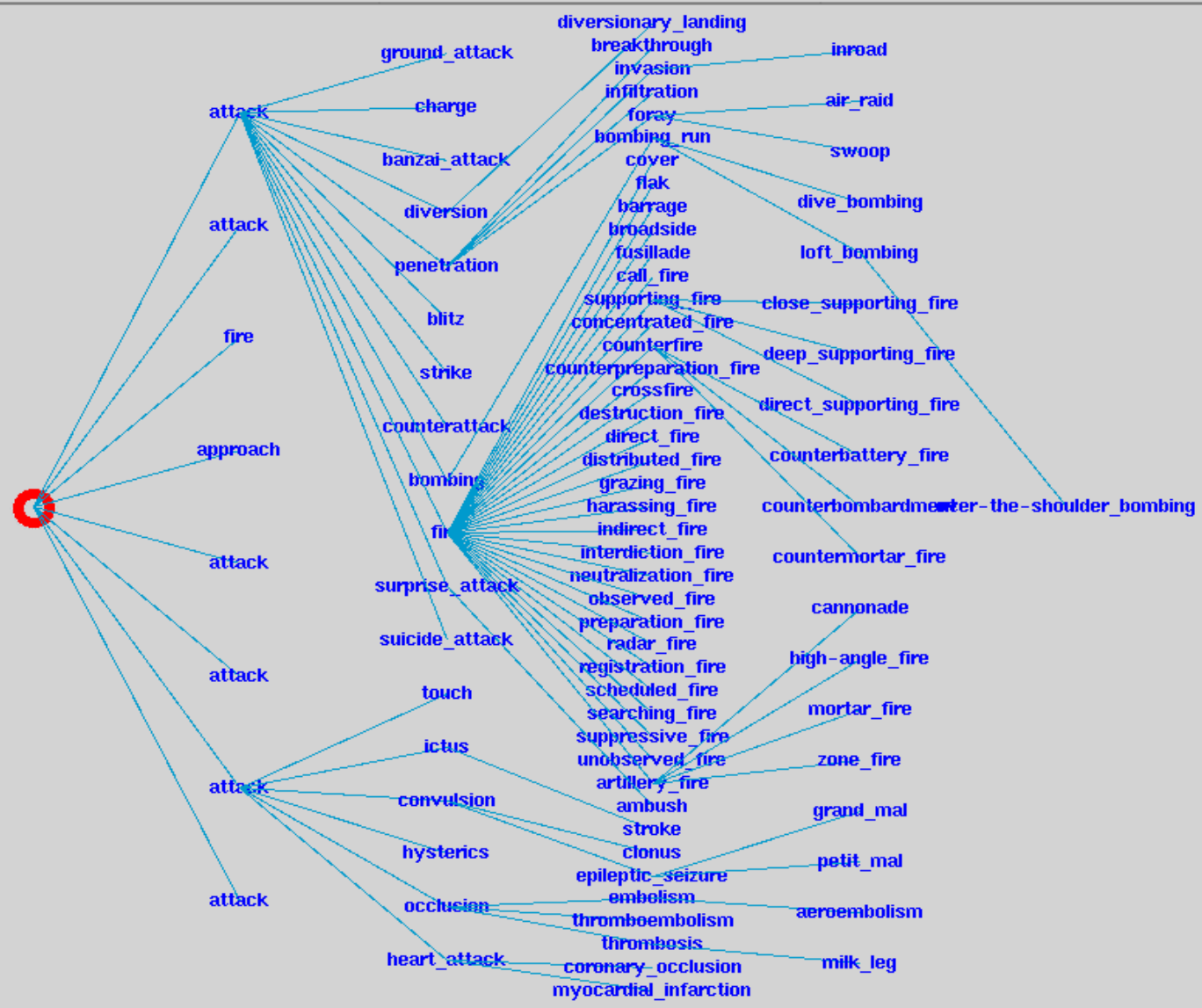
Proper Noun Categorization Scheme



KNOW-IT components

- Concepts
 - 60 + Proper Noun Categories
 - WordNet Synsets

Target Word: Relation: Part of Speech: Class:



KNOW-IT components

- Concepts
 - 60 + Proper Noun Categories
 - WordNet Synsets

- Relations
 - 40 + generic semantic relations

Semantic Relations

Relations

AGNT (act, animate)

PART (entity-x, entity-y)

PTIM (T, time)

CAUS (state-x, state-y)

PURP (act-x, act-y)
(state/entity, act-y)

Definition

*animate is performer
(agent) of action*

entity-x has part entity-y

T occurred at specific time

x has a cause y

*act-x has purpose act-y
state has purpose act-y*

KNOW-IT components

- Concepts
 - 60 + Proper Noun Categories
 - WordNet Synsets

- Relations
 - 40 + generic semantic relations
 - Concept-Relation-Concept

Concept-Relation Extraction

HEADLINE: Albanian suspected to have links to bin Laden arrested

SOURCE: Agence France Presse, 01/10/99

Maksim Ciciku was arrested by the Albanian police in Tirana. ... Ciciku met Osama bin Laden in April 1994.

Concept-Relation Extraction

HEADLINE: Albanian suspected to have links to bin Laden arrested

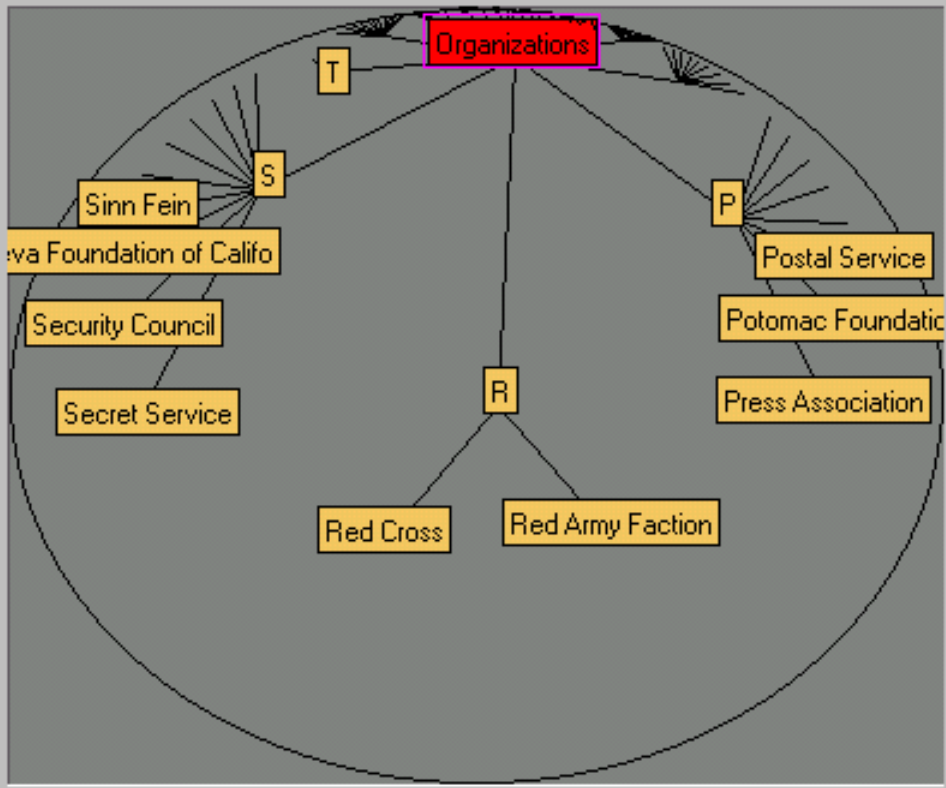
SOURCE: Agence France Presse, 01/10/99

Maksim Ciciku was arrested by the Albanian police in Tirana. ...
Ciciku met Osama bin Laden in April 1994.

CG_1 OBJ (arrest, Maksim Ciciku|*person*)
 AGNT (arrest, Albanian police)
 CHRC (police, Albanian|*nationality*)
 LOC (arrest, Tirana|*city*)

CG_2 AGNT (meet, Maksim Ciciku|*person*)
 OBJ (meet, Osama bin Laden|*person*)
 PTIM (meet, April 1994)

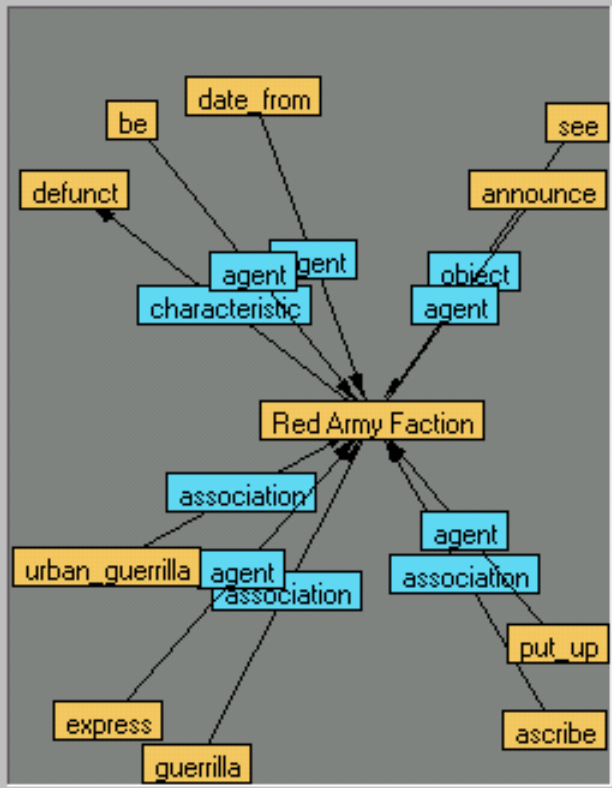
Expand Hyperbolic Browser



Red Army Faction Show this Category

Node Spacing

Expand Concept Visualizer



Viewing: 11 concepts

Node Spacing

Red Army Faction
 announce
 ascribe
 be
 date_from
 defunct
 express
 guerrilla
 put_up
 see
 urban_guerrilla

- Find Concept
- View Selected Concepts
- View More Information
- Freeze Organization
- Hide Relations

View Full Text View Sentences Set Date Range Clear Date Range

Title ▲	Author	Source	Date	DocId
<input type="checkbox"/> RED ARMY FACTION SAYS IT HAS DISBANDED	ALAN COWELL	N.Y. Times News Se	04/22/98	65f
<input type="checkbox"/> U.S. and Britain Guard Old Soviet Nuclear Material	unknown	N.Y. Times News Se	04/25/98	39c

Adapting KNOW-IT for Link Analysis

- Extraction in KNOW-IT is **broad** and **shallow**
 - based on linguistic regularities
 - not domain-dependent rules
- But the technology can be extended to **narrow** and **deep** applications for Link Analysis
 - ‘terrorism’ domain for HPKB

Specialization Methodology

- Map 2 or more *general* C-R-C extraction rules into a more specific link rule, e.g. for SUPPORT:

Specialization Methodology

- Map 2 or more *general* C-R-C extraction rules into a more specific link rule, e.g. for SUPPORT:

$$\begin{array}{c} C1 - R - C2 \\ + \\ C2 - R - C3 \end{array}$$

Specialization Methodology:

- Map 2 or more *general* C-R-C extraction rules into a more specific link rule, e.g. for SUPPORT:

C1 - AGNT - C2
+
C2 - OBJ - C3

Specialization Methodology

- Map 2 or more *general* C-R-C extraction rules into a more specific link rule, e.g. for SUPPORT:

C1 - AGNT - C2
+
C2 - OBJ - C3



<international agent*> AGNT <support verb*>
+
<support verb*> OBJ <X|54>

Where,

International agent = **any Proper Noun whose category is an element of the set {7, 40, 411, 412, 413, 414, 415, 416, 417, 50, 501, 51, 52, 53, 54}**

AND

Support verb = **any element of the synsets containing verbs such as: {*fund, back, support, aid, help, assist, sponsor, subsidize, patronize, cosponsor, bankroll, champion, defend*}**

Then,

.... extremist Harkatul Jihad group,
reportedly backed by Saudi dissident
Osama bin Laden ...

Then,

.... extremist Harkatul Jihad group,
reportedly backed by Saudi dissident
Osama bin Laden ...



AGENT (back , Osama bin Laden|*person*)

OBJECT (back , Hartakul
Jihad|*terrorist_group* group)

Then,

.... extremist Harkatul Jihad group, reportedly backed by Saudi dissident Osama bin Laden ...



AGENT (back , Osama bin Laden|*person*)

OBJECT (back , Hartakul Jihad|*terrorist_group* group)



SUPPORT (Osama bin Laden|*person* , Hartakul Jihad|*terrorist_group* group)

03/14/1999 (AFP) Bangladesh bomb blast toll 10, opposition wants judicial probe
... the extremist Harkatul Jihad group, reportedly backed by Saudi dissident Osama bin Laden ...

03/14/1999 (AFP) Bangladesh bomb blast toll 10, opposition wants judicial probe
... the extremist Harkatul Jihad group, reportedly backed by Saudi dissident Osama bin Laden ...

... the|DT extremist|JJ Harkatul_Jihad|NP|1 group|NN ,|, reportedly|RB backed|VBD
by|IN Saudi|NP|2 dissident|IN Osama_bin_LadeNP|3 ...

03/14/1999 (AFP) Bangladesh bomb blast toll 10, opposition wants judicial probe
... the extremist Harkatul Jihad group, reportedly backed by Saudi dissident Osama bin Laden ...

... the|DT extremist|JJ Harkatul_Jihad|NP|1 group|NN ,|, reportedly|RB backed|VBD
by|IN Saudi|NP|2 dissident|IN Osama_bin_LadeNP|3 ...

<PN>

1	54	Harkatul Jihad
2	17	Saudi
3	30	Osama bin Laden

</PN>

03/14/1999 (AFP) Bangladesh bomb blast toll 10, opposition wants judicial probe
... the extremist Harkatul Jihad group, reportedly backed by Saudi dissident Osama bin Laden ...

... the|DT extremist|JJ Harkatul_Jihad|NP|1 group|NN ,|, reportedly|RB backed|VBD
by|IN Saudi|NP|2 dissident|IN Osama_bin_LadeNP|3 ...

<PN>

1	54	Harkatul Jihad
2	17	Saudi
3	30	Osama bin Laden

</PN>

CG0: AGNT (back , Osama bin Laden|*person*)
OBJECT (back , Harkatul Jihad|*terrorist_group* group)
CHRC (Harkatul Jihad|*terrorist_group* group , extremist)
MANR (back , reportedly)
ISA (Osama bin Laden|*person* , Saudi|*nationality* dissident)

03/14/1999 (AFP) Bangladesh bomb blast toll 10, opposition wants judicial probe
... the extremist Harkatul Jihad group, reportedly backed by Saudi dissident Osama bin Laden ...

... the|DT extremist|JJ Harkatul_Jihad|NP|1 group|NN |, reportedly|RB backed|VBD
by|IN Saudi|NP|2 dissident|IN Osama_bin_LadeNP|3 ...

<PN>

1	54	Harkatul Jihad
2	17	Saudi
3	30	Osama bin Laden

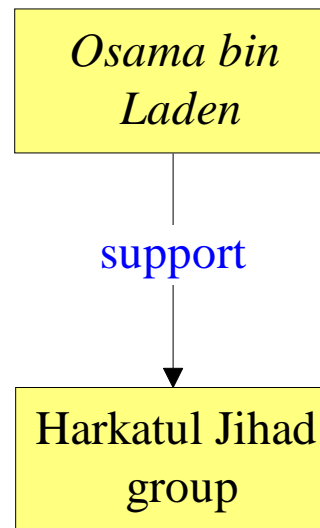
</PN>

CG0: AGNT (back , Osama bin Laden|*person*)
OBJECT (back , Harkatul Jihad|*terrorist_group* group)
CHRC (Harkatul Jihad|*terrorist_group* group , extremist)
MANR (back , reportedly)
ISA (Osama bin Laden|*person* , Saudi|*nationality* dissident)

CG0: *AGNT* (back , Osama bin Laden|*person*)
OBJECT (back , Harkatul Jihad|*terrorist_group* group)
...
SUPPORT(Osama bin Laden|*person* , Harkatul Jihad|*terrorist_group* group)

03/14/1999 (AFP) Bangladesh bomb blast toll 10, opposition wants judicial probe
... the extremist Harkatul Jihad group, reportedly backed by Saudi dissident Osama bin Laden ...

support (Osama bin Laden|*person* , Hartakul Jihad|*terrorist_group* group)



03/12/1999 (AFP) Bangladesh arrest Afghan war veteran over bomb attack
... from Monirul Hassan, a member of the Harkatul Jihad group who was reportedly trained by the Taliban militia in Afghanistan ...

03/12/1999 (AFP) Bangladesh arrest Afghan war veteran over bomb attack
... from Monirul Hassan, a member of the Harkatul Jihad group who was reportedly
trained by the Taliban militia in Afghanistan ...

CG0

ISA (Monirul Hassan|*person* , member)
AFFL (member , Harkatul Jihad|*terrorist_group* group)
AGNT (train , Taliban militia|*organization*)
OBJ (train , Monirul Hassan|*person*)
LOC (train , Afghanistan|*country*)

03/12/1999 (AFP) Bangladesh arrest Afghan war veteran over bomb attack
... from Monirul Hassan, a member of the Harkatul Jihad group who was reportedly
trained by the Taliban militia in Afghanistan ...

CG0

ISA (Monirul Hassan|*person* , member)

AFFL (member , Harkatul Jihad|*terrorist_group* group)

...

affiliate (Monirul Hassan|*person* , Harkatul Jihad|*terrorist_group* group)

AGNT (train , Taliban militia|*organization*)

OBJ (train , Monirul Hassan|*person*)

...

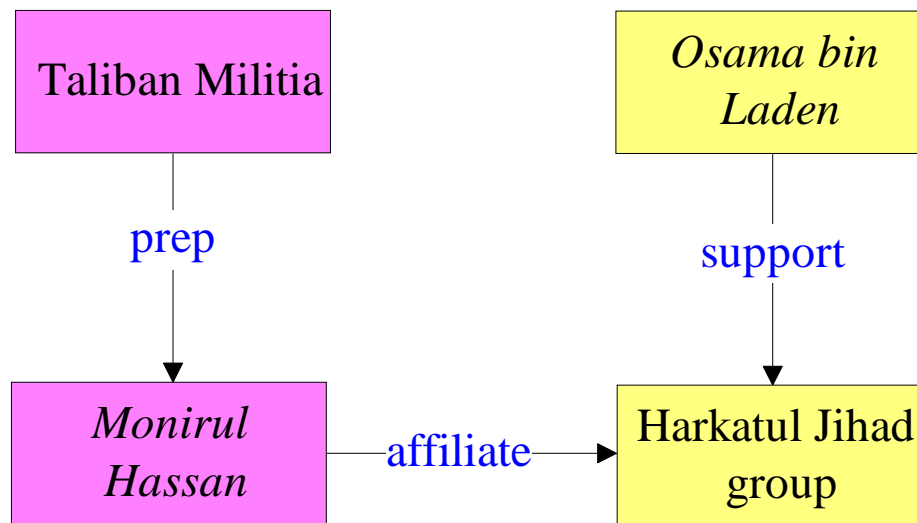
prep (Taliban militia|*organization* , Monirul Hassan|*person*)

03/12/1999 (AFP) Bangladesh arrest Afghan war veteran over bomb attack
... from Monirul Hassan, a member of the Harkatul Jihad group who was reportedly trained by the Taliban militia in Afghanistan ...

affiliate (Monirul Hassan|*person* , Harkatul Jihad|*terrorist_group* group)
prep (Taliban militia|*organization* , Monirul Hassan|*person*)

03/12/1999 (AFP) Bangladesh arrest Afghan war veteran over bomb attack
... from Monirul Hassan, a member of the Harkatul Jihad group who was reportedly trained by the Taliban militia in Afghanistan ...

affiliate (Monirul Hassan|*person* , Harkatul Jihad|*terrorist_group* group)
prep (Taliban militia|*organization* , Monirul Hassan|*person*)



03/08/1999 (AFP) 16 soldiers killed, 21 wounded in Algerian ambush

The Salafist Group for Preaching and Combat (GSPC), led by Hassan Hattab, recently distributed ...

Created at the instigation of bin Laden, the group is especially active ...

03/08/1999 (AFP) 16 soldiers killed, 21 wounded in Algerian ambush

The Salafist Group for Preaching and Combat (GSPC), led by Hassan Hattab, recently distributed ...

Created at the instigation of bin Laden, the group is especially active ...

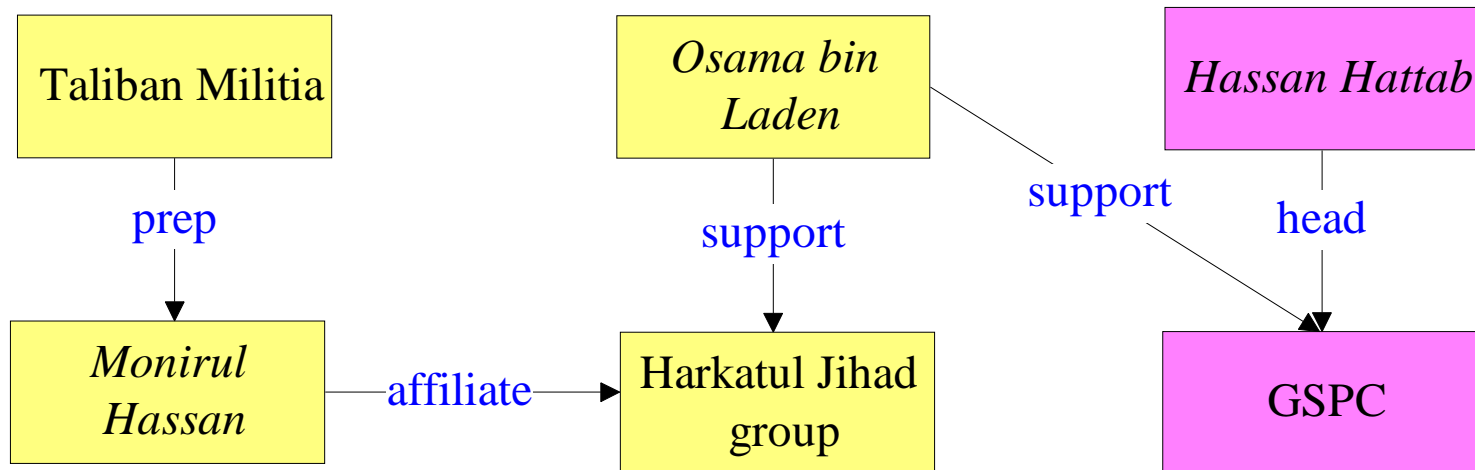
head (Hassan Hattab|*person* , GSPC|*terrorist_group*)

support (Osama bin Laden|*person* , GSPC|*terrorist_group*)

03/08/1999 (AFP) 16 soldiers killed, 21 wounded in Algerian ambush
The Salafist Group for Preaching and Combat (GSPC), led by Hassan Hattab,
recently distributed ...

Created at the instigation of bin Laden, the group is especially active ...

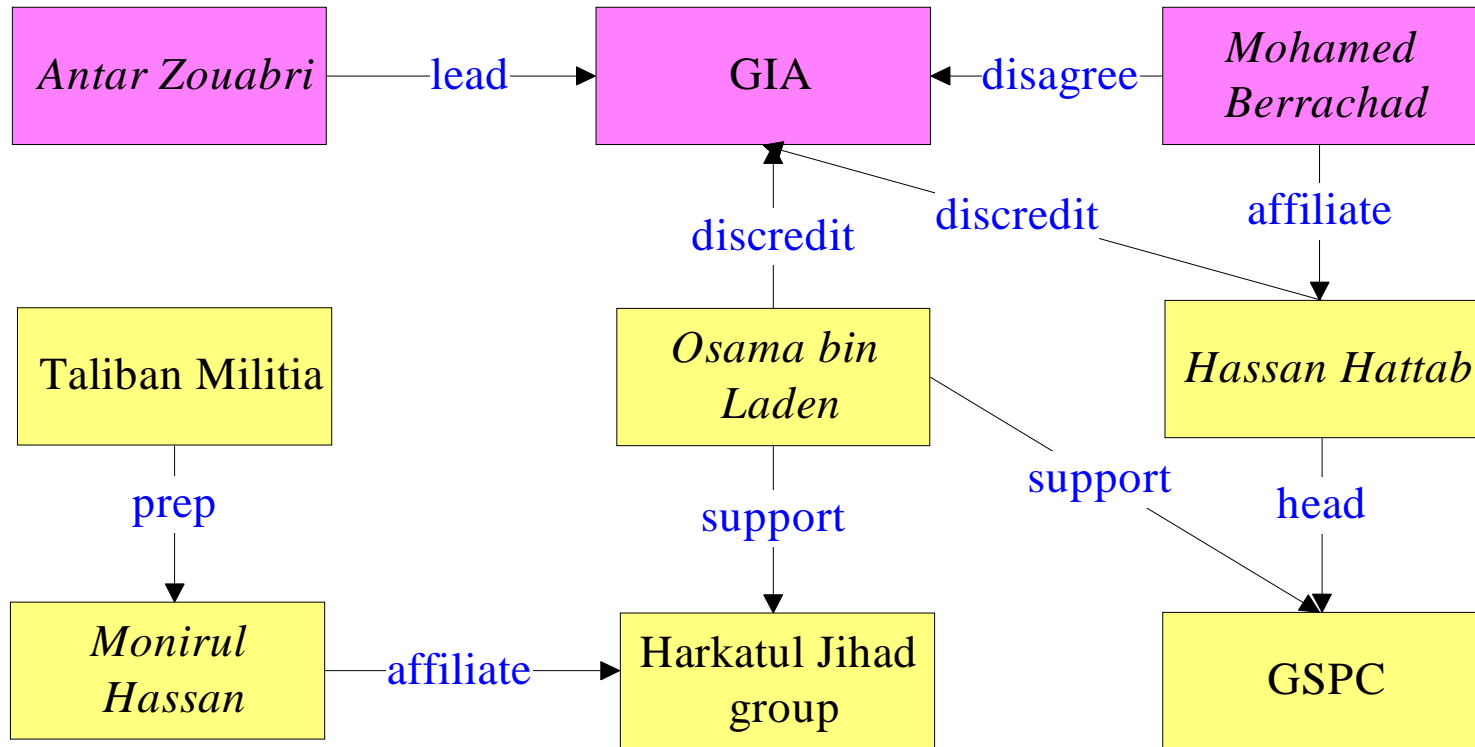
head (Hassan Hattab|*person* , GSPC|*terrorist_group*)
support (Osama bin Laden|*person* , GSPC|*terrorist_group*)

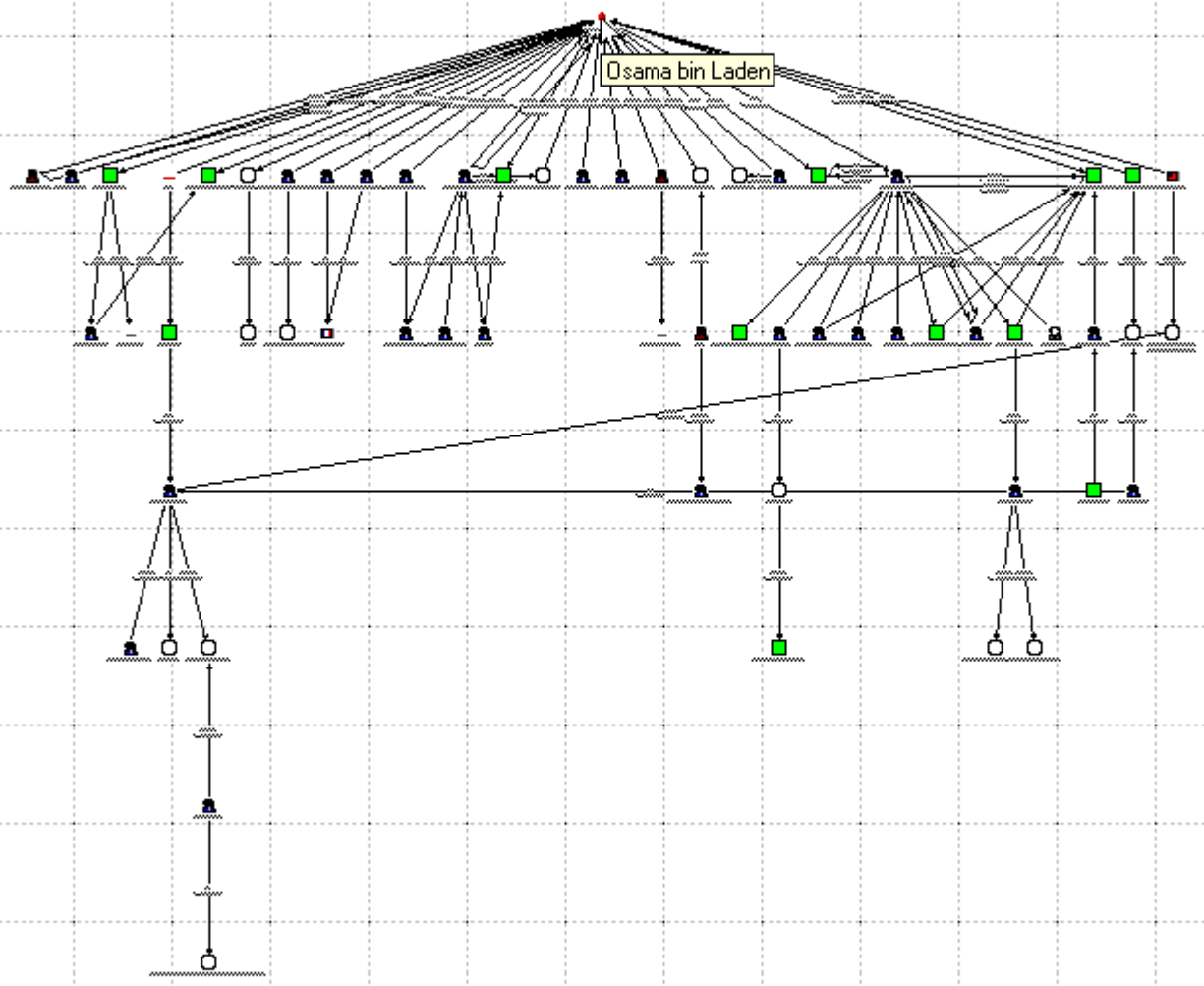


02/15/1999 (AFP) Bin Laden held to be behind an armed Algerian Islamic movement

Mohamed Berrachad had worked for Hattab, who is himself a dissident from the **Armed Islamic Group (GIA)** ...

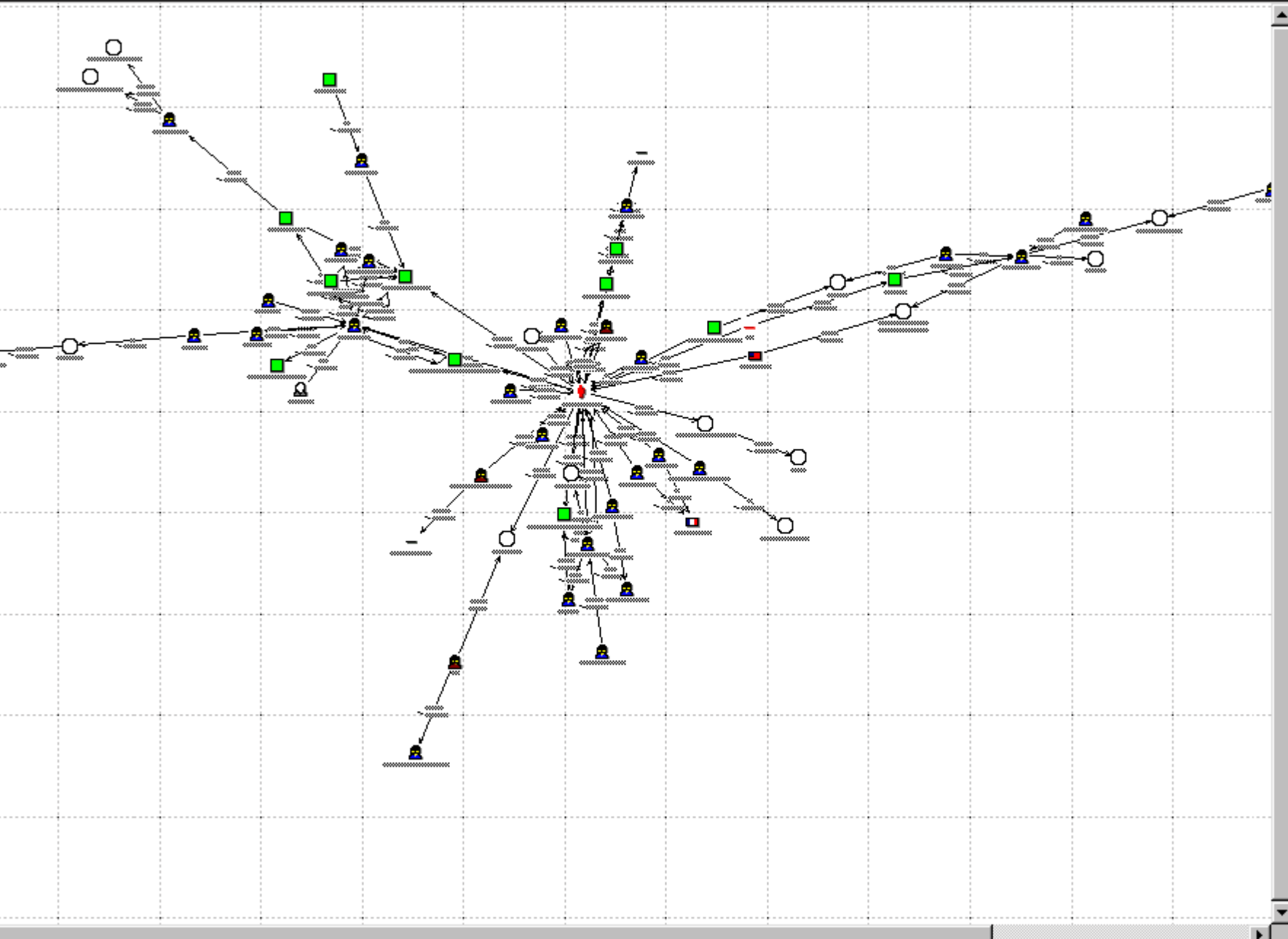
In his testimony, Berrachad said Bin Laden and Hattab communicated by satellite telephone and that he had heard their conversations, said to hinge on the discrediting of **Antar Zouabri's** GIA by its savage massacres of civilians ...





Standard

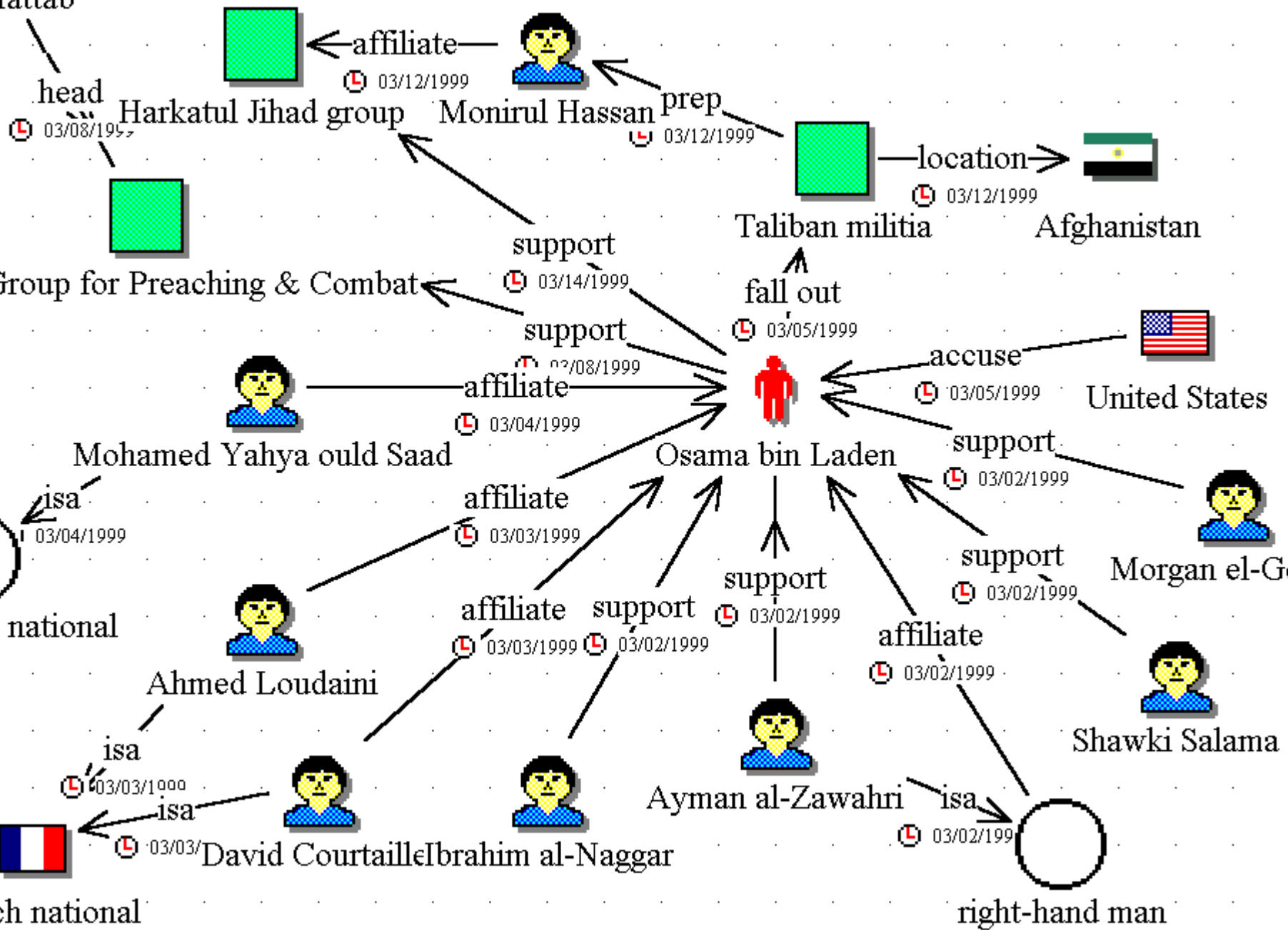
- Entity
- Organization
- Telephone
- Female
- Male



Standard

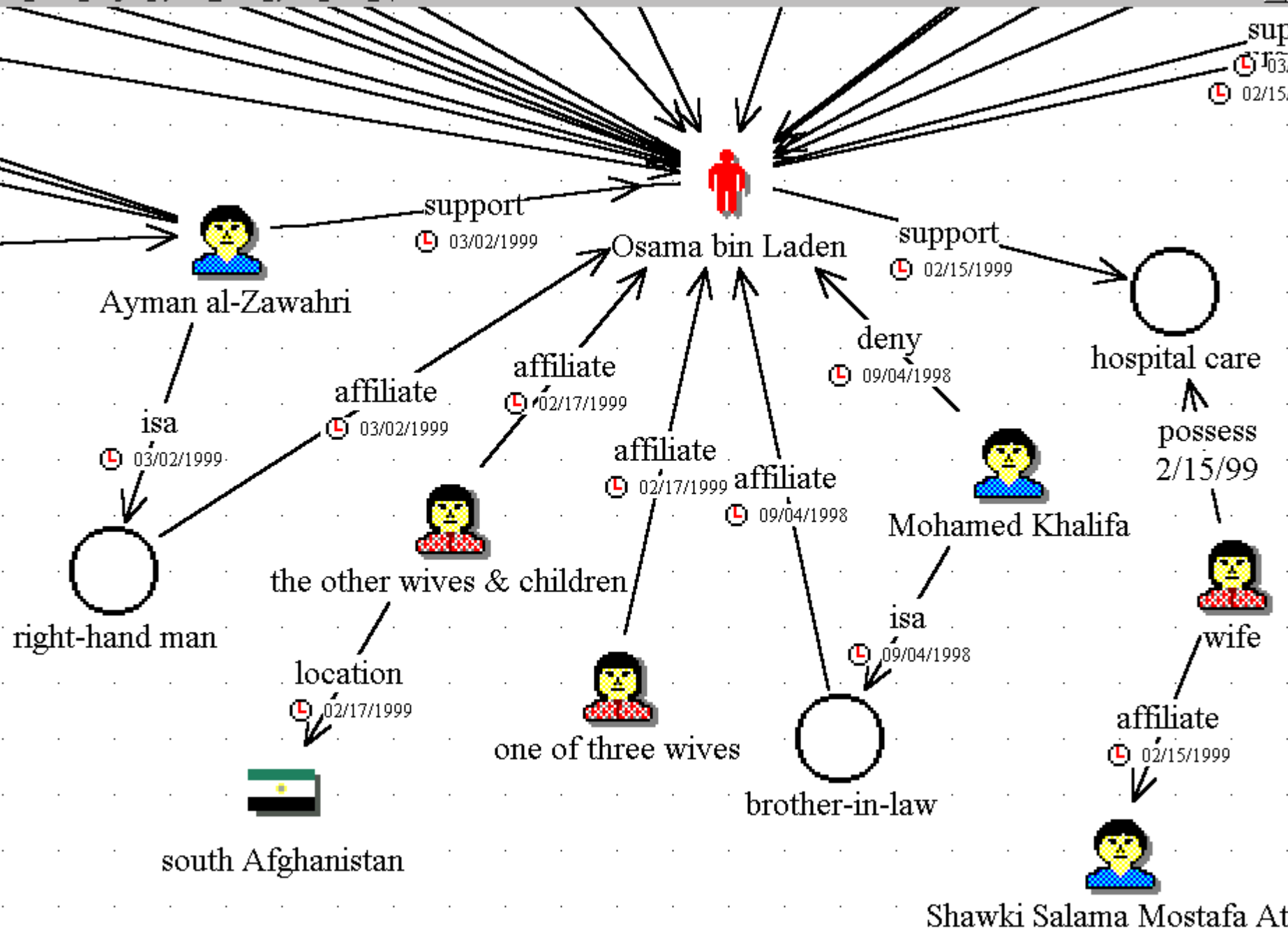
- Entity (White circle icon)
- Organization (Green square icon)
- Telephone (Telephone handset icon)
- Female (Female human icon)
- Male (Male human icon)

san Hattab



French national

right-hand man



As a link analyzer, KNOW-IT

- Assists analysts in appraising a potential crisis situation by determining the key players and the nature of their relations to one another
- Automatically filters, extracts, organizes, and analyzes textual intelligence data
- Generates and visualizes networks from relevant, unstructured text
- Allows analysts to specialize the links by easy-to-write specification & generalization rules
- Provides rich output to visualization tools