

Network Based Hard/Soft Information Fusion

Tractor: Soft Data Understanding

Stuart C. Shapiro, Daniel Schlegel







- Extract state estimates from English messages.
- Preserve as much information as possible.
- Enhance with contextually relevant background information.

propositional graph soft data attributed graph graph with aligned uncertainty Tractor

Scientific/Technical Approach

- Use state-of-the-art techniques from Computational Linguistics.
- Extract semantic relations from syntactic relations.
- Represent state estimates as propositional graphs.

Main Accomplishments

- Improved stemmer, co-reference editor.
- Syntax-semantics mapping.
- NL entity description facility.
- Network-Based Services.

Challenges

- Task implies full NLU.
- Limited personnel.



Motivation/Role in Architecture Tractor: Soft Data Understanding





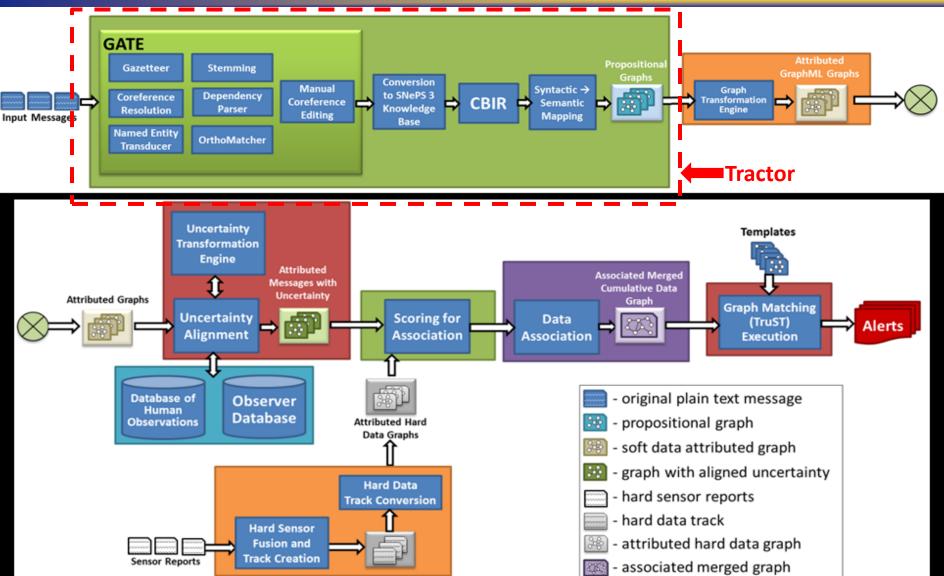
- Soft Data are a set of short English messages.
- Each must be converted into expressions of a formal representation language
 - for fusion with other messages
 - and with hard data.
- Fusion depends on semantic descriptions
 - of entities and events.
- Tractor constructs these semantic descriptions.



The MURI Architecture









Main Scientific/Technical Accomplishments Tractor: Soft Data Understanding

center for multisource informationfusion
A CUBRC and University at Buffalo Research Center





Designed framework, chose major components.

Main Year 2 Accomplishments

Initial implementation of framework, GUI.

Produced tagged dependency graphs.

Main Year 3 Accomplishments

Rewrote Stemmer, Co-reference Editor.

Rewrote all Post-GATE Components.

Implemented 65 Syntax-Semantics Mapping Rules.

Semantic Propositional Graph almost entirely Semantic.

Designed/Implemented NL Entity Description Facility.

Tractor processes messages at apx. 0.03 secs/word.



Project Statistics and Summary Year 3

Tractor: Soft Data Understanding





Personnel supported

- James Dobler, undergraduate student (Spring, Summer)
- Daniel R. Schlegel, graduate student
- Stuart C. Shapiro, faculty member/senior researcher

Students with leveraged (non-ARO) support

Paul Bunter: volunteer graduate student (Fall)

Publications

- 1 Book chapter
- 1 Conference paper
- 1 Manual

Transitions

- Jim Hendler (for Army Network Science CTA project)
- Mark A. Thomas, USA CIV (US)
- Gabor (Gabe) Schmera, SPAWAR



Project Statistics and Summary Year 3

Tractor: Soft Data Understanding





Year 3 Publications

- Daniel R. Schlegel and Stuart C. Shapiro, Visually Interacting with a Knowledge Base
 Using Frames, Logic, and Propositional Graphs. In Madalina Croitoru, Sebastian Rudolph,
 Nic Wilson, John Howse and Olivier Corby, Eds., Graph Structures for Knowledge
 Representation and Reasoning, Lecture Notes in Artificial Intelligence 7205, SpringerVerlag, Berlin, 2012, 188-207.
- Geoff A. Gross, Rakesh Nagi, Kedar Sambhoos, Daniel R. Schlegel, Stuart C. Shapiro, and Gregory Tauer, Towards Hard+Soft Data Fusion: Processing Architecture and Implementation for the Joint Fusion and Analysis of Hard and Soft Intelligence Data, Proceedings of the 15th International Conference on Information Fusion (Fusion 2012), IFIP, 2012, 955-962.
- Stuart C. Shapiro, Michael Prentice, and Daniel R. Schlegel, Tractor Manual, University at Buffalo, July 18, 2012.

Year 3 Presentations

- Stuart C. Shapiro, Tractor: Toward Deep Understanding of Short Intelligence Messages
 - Presented to the UB Center for Cognitive Science, March 28, 2012.
 - Presented to Milan Patel, 2WD/A2SF, June 19, 2012.



Technical Approach

Tractor: Soft Data Understanding





- Syntactic processing by GATE PRs and human co-reference editor.
 - → Typed "annotations" > 1 per text substring.
- Conversion to syntactic propositional graph
 - Annotation merging; type correction;
 canonicalize dates/times; structured headers.
- Add relevant geographical information; ontology; mereology.
- Syntax-Semantics mapping.
 - → Semantic propositional graph
- See: Gross, et al., Fusion 2012; Year 3 Interim Progress Report



Main Accomplishments Tractor: Soft Data Understanding Stemmer





- GATE used Porter Stemmer
- Designed for Information Retrieval
- Discards suffixes
 - "Courthouse" → "courthous"
 - "residents" → "resid"
- Improved version by Bunter
 - "Courthouse" → "courthouse"
 - "residents" → "resident"
- Why important:
 - Connect to ontology
 - "residents" → Group of people, each a resident
- Original mis-stemmed 9.6 % of word tokens in BBS messages;
 Improved version got all correct.



Main Accomplishments

Tractor: Soft Data Understanding Co-reference Editor





- GATE Co-reference Editor wasn't working.
- Schlegel & Dobler fixed and improved it.
- Allows human to
 - Correct incorrect coreference chains
 - Add to coreference chains
 - Add coreference chains
- Optional: Tractor can be run without it.
- Illustrated in Demo below.



Main Accomplishments Tractor: Soft Data Understanding Propositionalizer





- Converts
 - Set of annotations
 - to Syntactic Propositional Graph
- Combines annotations that cover same text
- Corrects minor errors in POS tags
- Converts dates/times to ISO8601
- Processes structured headers
- Expresses syntactic relations as formal expressions for Propositional Graph



Main Accomplishments Tractor: Soft Data Understanding CBIR





- Adds geographic data from NGA GeoNet Names Server database
 - Place categories
 - Latitude
 - Longitude
 - MGRS
 - MGRS Radius
- Adds ontology above category mentions
 - From ResearchCyc, OpenCyc, local ontology
- Adds relevant
 - mereology, dimension values, substance list



Main Accomplishments Tractor: Soft Data Understanding Syntax-Semantics Mapping





Converts

- Syntactic propositional graph
- To semantic propositional graph
- Reverse of
 - If I want to convey some belief to you, how might I express that belief in English.
- Handcrafted rules
 - Sensitive to COIN domain
 - Sensitive to SynCOIN data
- Currently 65 rules
- Example run in Demo below
- For more details: Year 3 Interim Progress Report



Main Accomplishments Tractor: Soft Data Understanding Description Facility





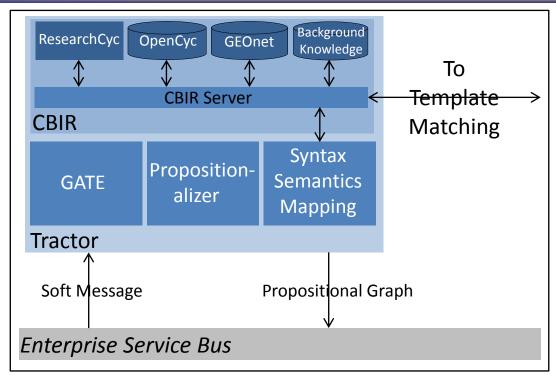
- Operates on Semantic Propositional Graph
- For any set of coreferring terms
 - Computes "best class"
 - A least general class the terms are instances of
 - Computes "best name"
 - Longest name of entity in domain
 - Or constructed from best class
- For any proposition
 - Computes English sentence
 - From sentence frame
 - And best names
- See algorithm in Year 3 Interim Progress Report
- See examples in Demo below



Main Accomplishments Tractor: Soft Data Understanding Network-Based Services







- New messages retrieved and processed automatically
- CBIR Server allows connections to many concurrent Tractor or Template Matching processes
- CBIR data sources can be local or distributed
- Will be discussed during Demo session.



Option Year Plans







Capability Goal:

- Expanded set of syntax-semantics mapping rules
- Research Goals:
 - Design and implement Concurrent SNePS
 - For parallel matching of LHSs of mapping rules
 - For parallel inference using quantified rules
- Test & Evaluation Goals
 - TAC evaluation
 - Compare with GOTS tools, e.g., ATEA



Demonstration 1 Example Message BBS-1





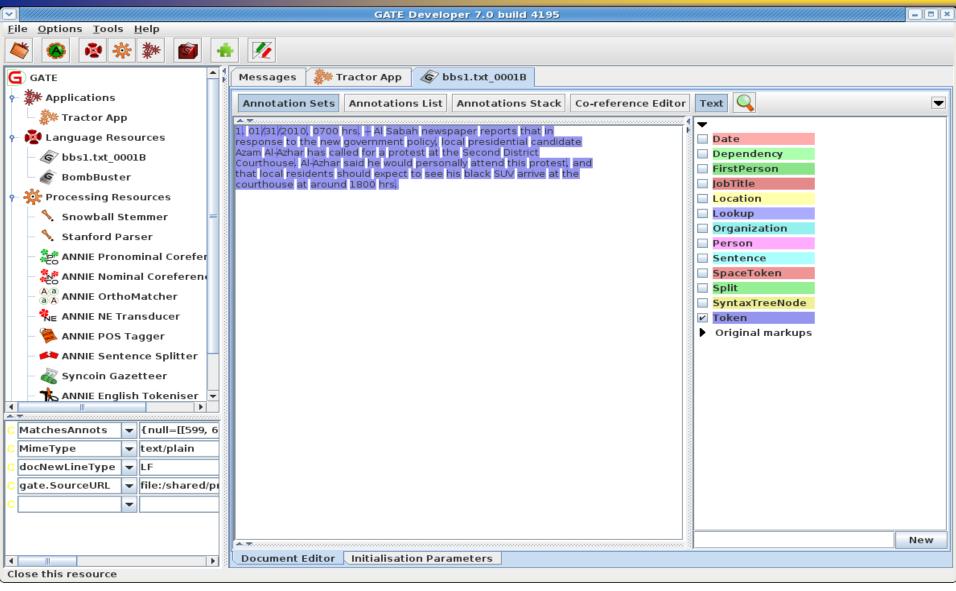
1. 01/31/2010, 0700 hrs. -- Al Sabah newspaper reports that in response to the new government policy, local presidential candidate Azam Al-Azhar has called for a protest at the Second District Courthouse. Al-Azhar said he would personally attend this protest, and that local residents should expect to see his black SUV arrive at the courthouse at around 1800 hrs.



Results of Tokenizer





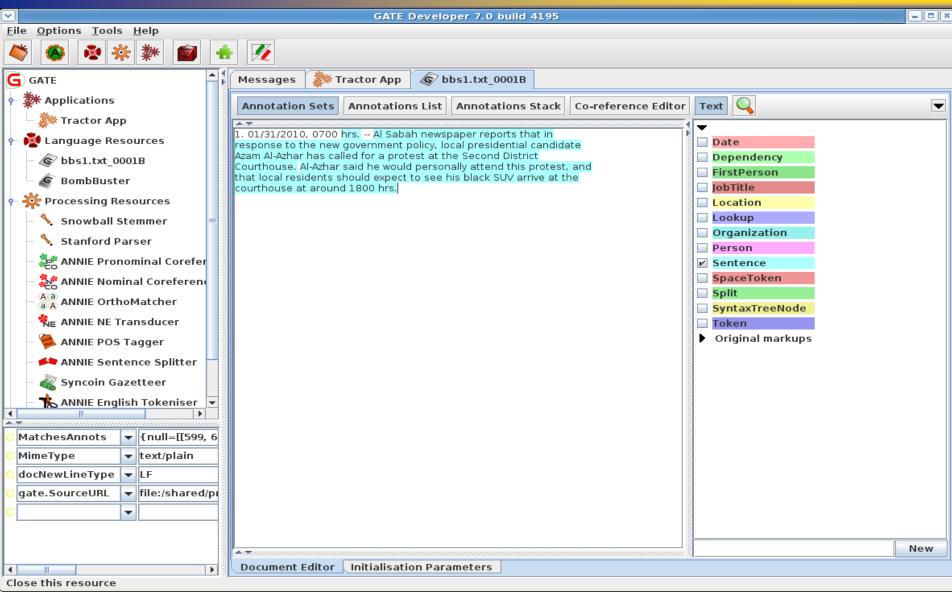




Results of Sentence Splitter







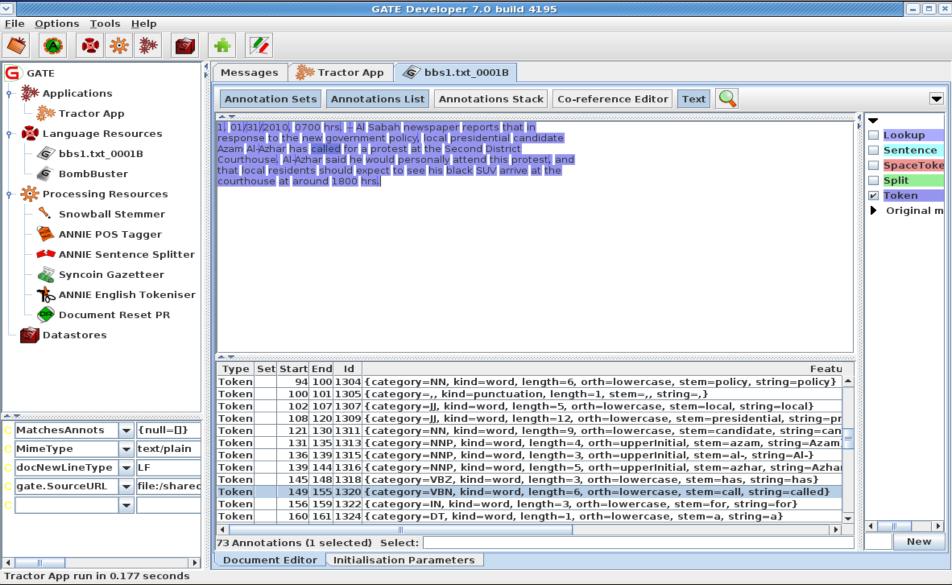


Results of POS Tagger & Stemmer







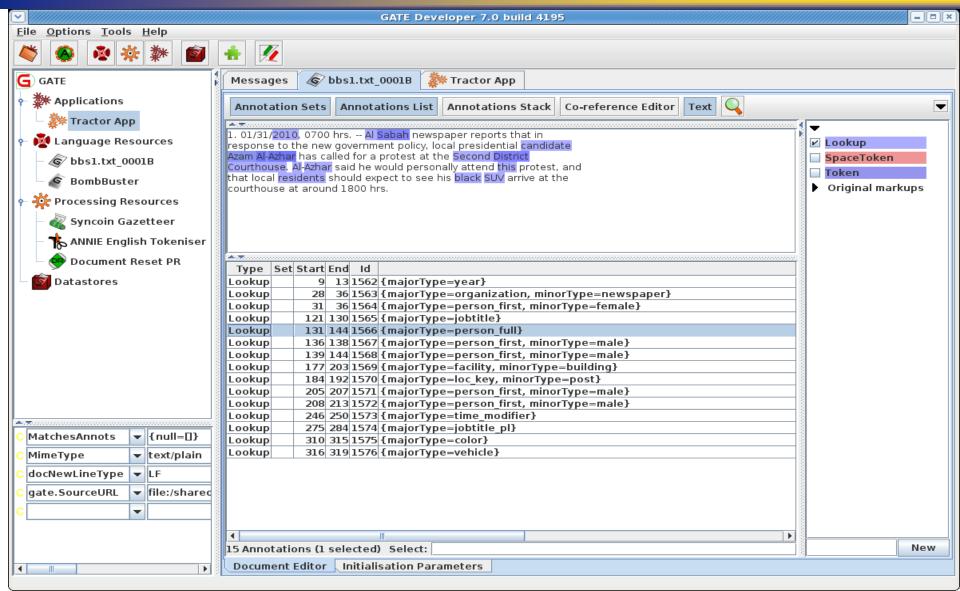






Results of "Gazetteer-Based" Named Entity Recognizer



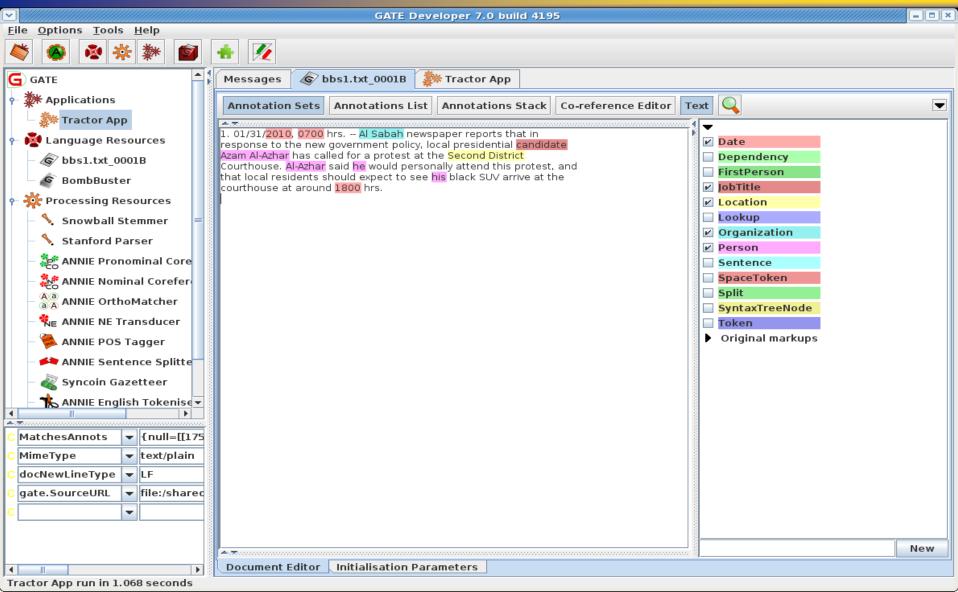






Results of Rule-Based Named Entity Recognition



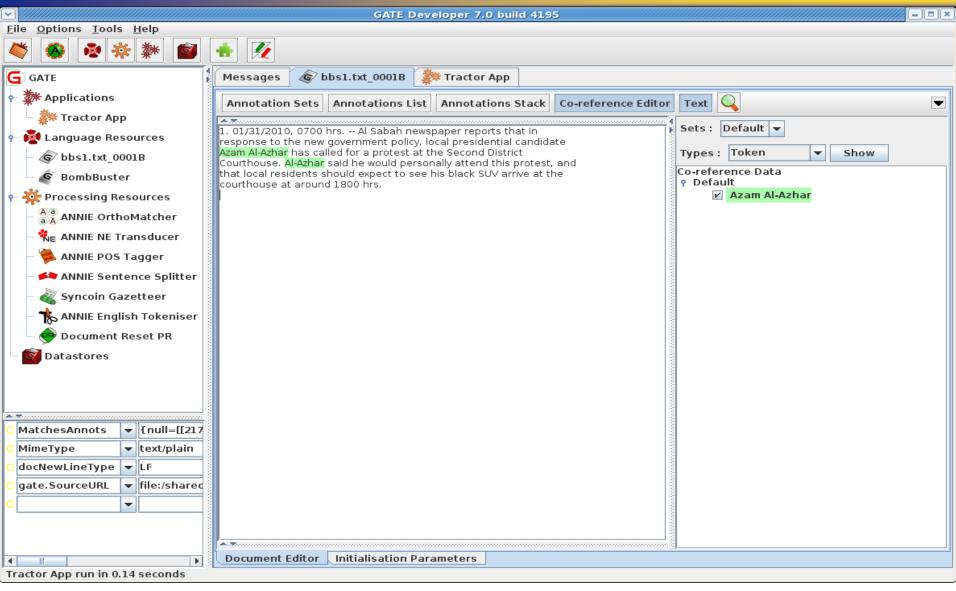




Results of OrthoMatcher







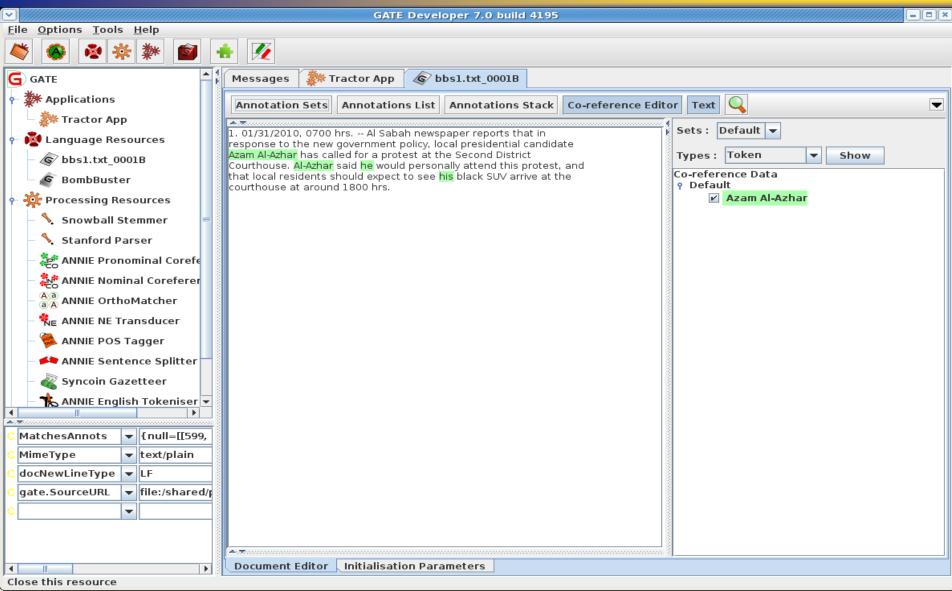


Results of Pronominal Coreferencer







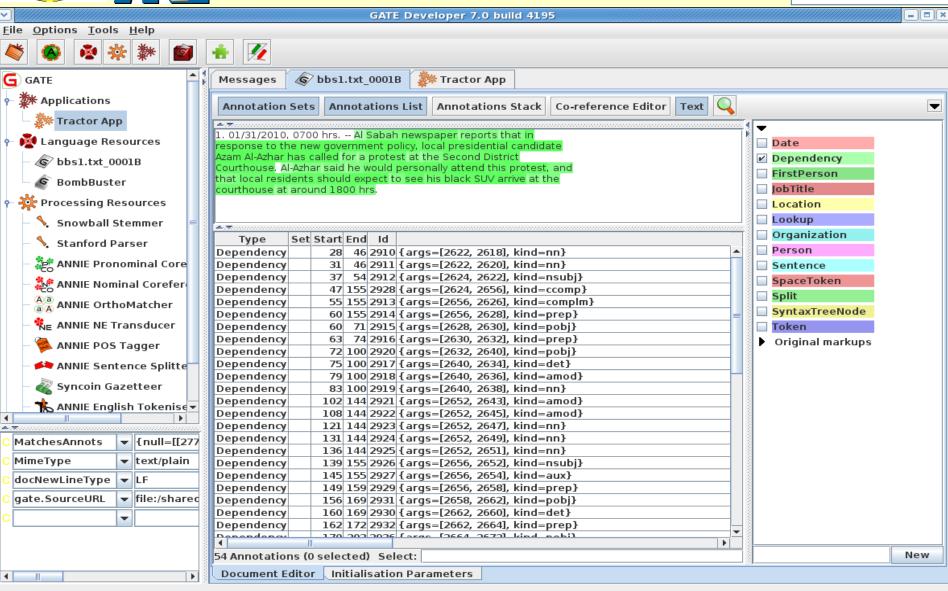




Results of Dependency Parse



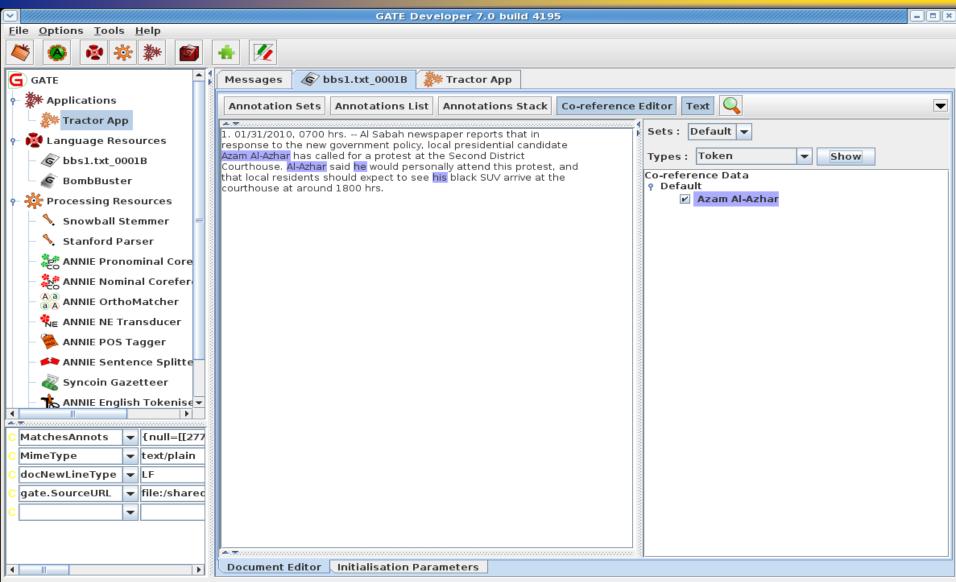






Co-reference Editor Initial Chain

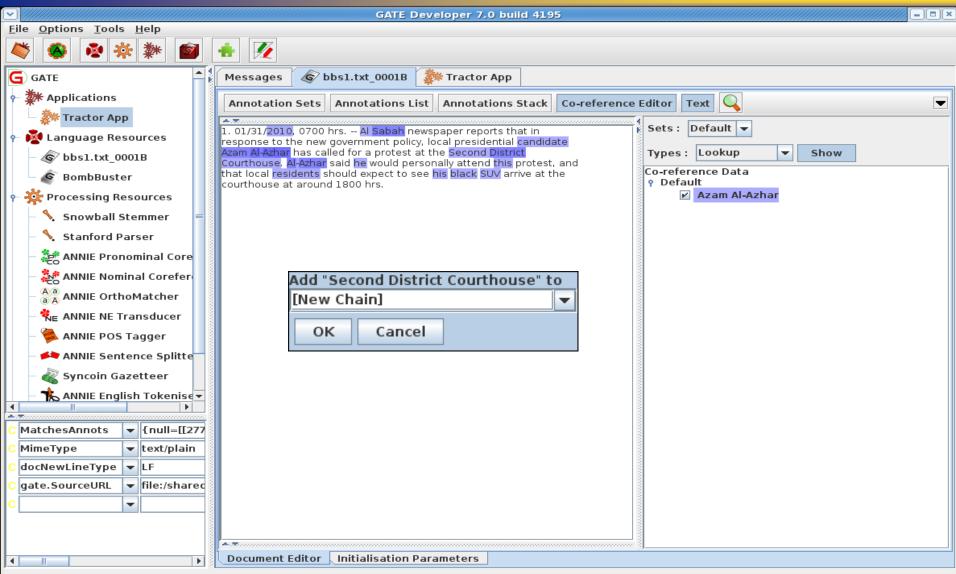






Co-reference Editor Adding a Chain

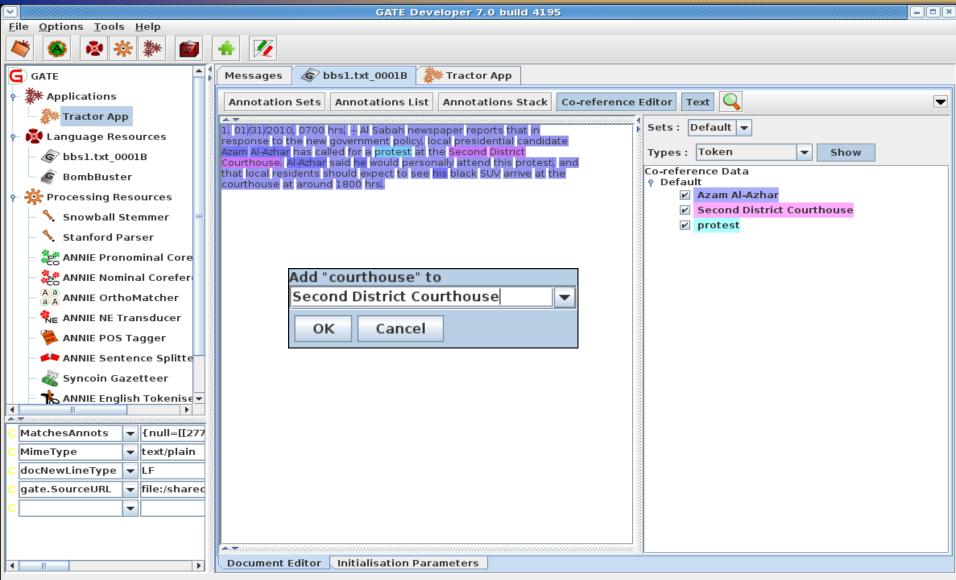






Co-reference Editor Adding to a Chain

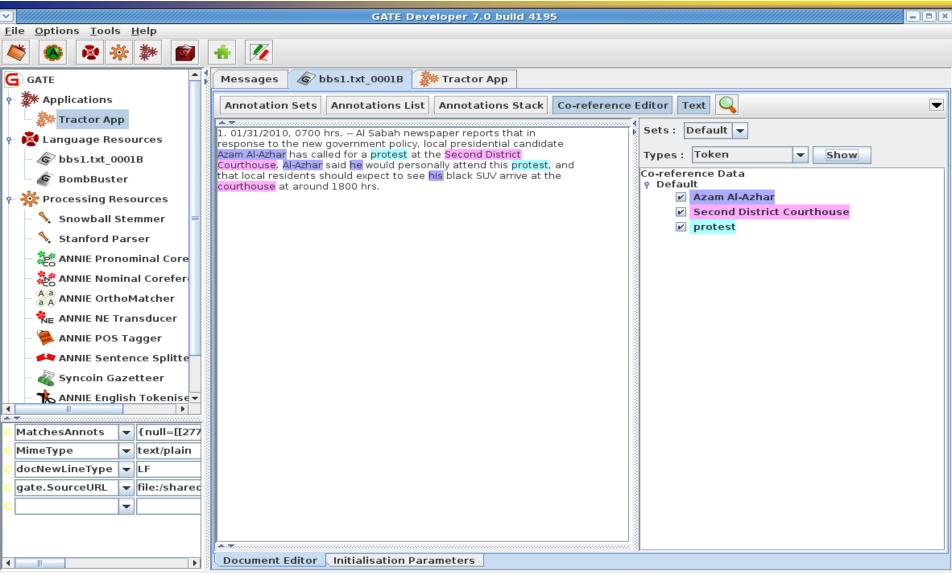






Co-reference Editor Final Chains







Demonstration 2 Example Message syn064





64. 01/27/10 - BCT forces detained a Sunni munitions trafficker after a search of his car netted IED trigger devices. Ahmad Mahmud was placed in custody after his arrest along the Dour'a Expressway, //MGRSCOORD: 38S MB 47959 80868//, in East Dora.

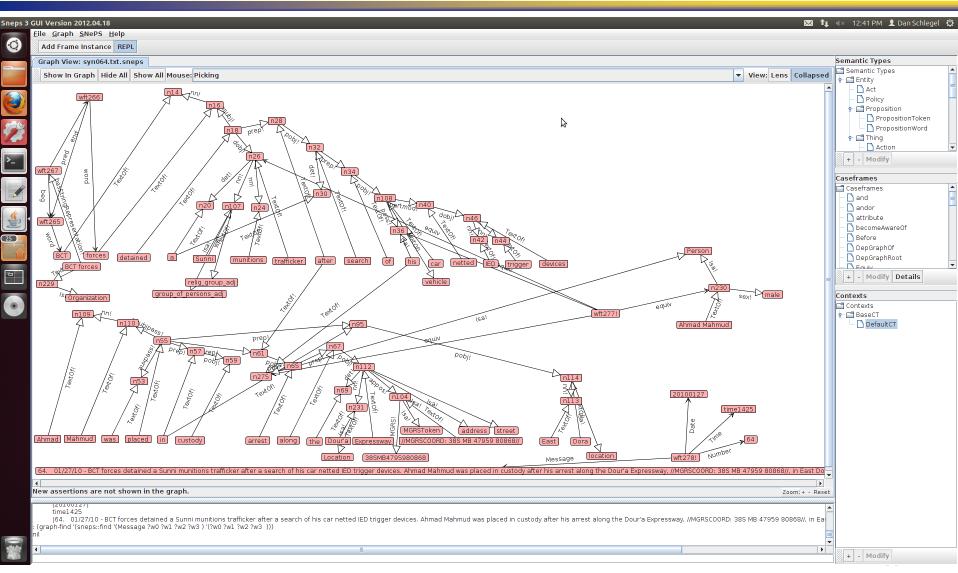


Syntactic Propositional Graph







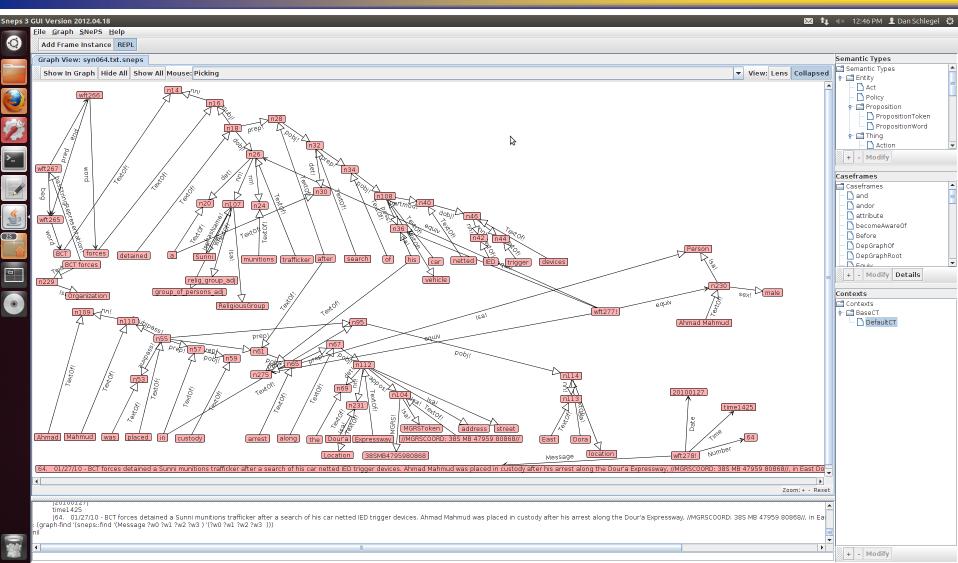




After regularizeStems







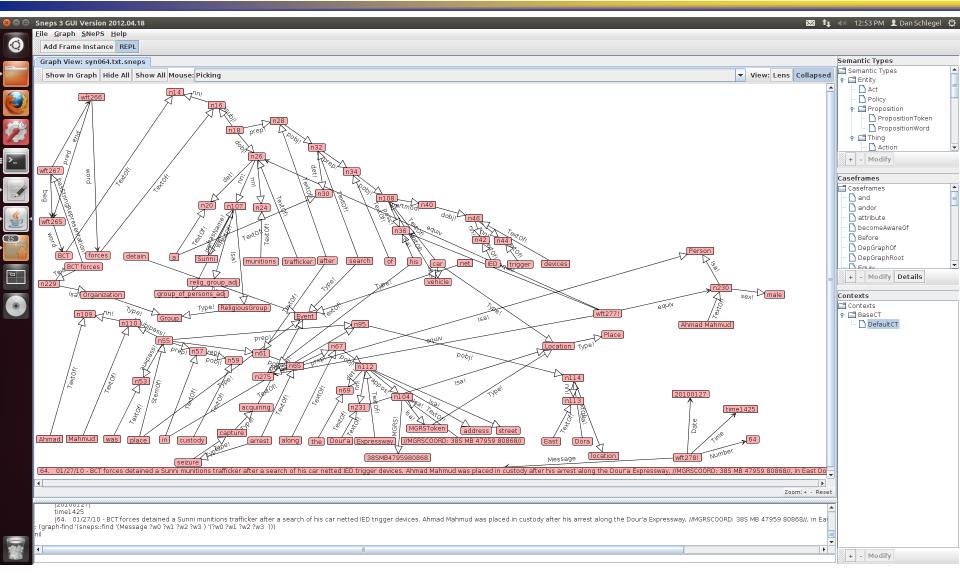


After addRelevantOntology









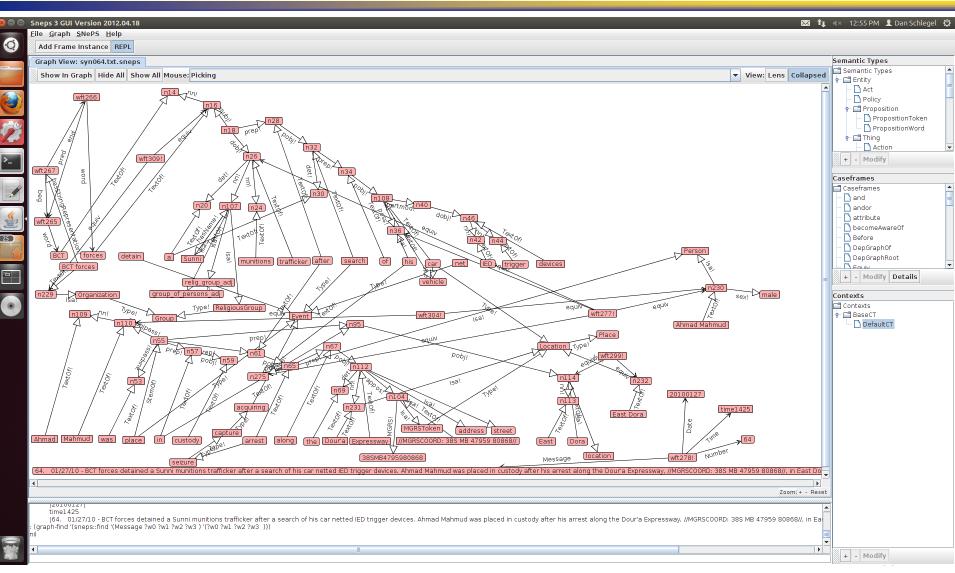


After substringCoreference







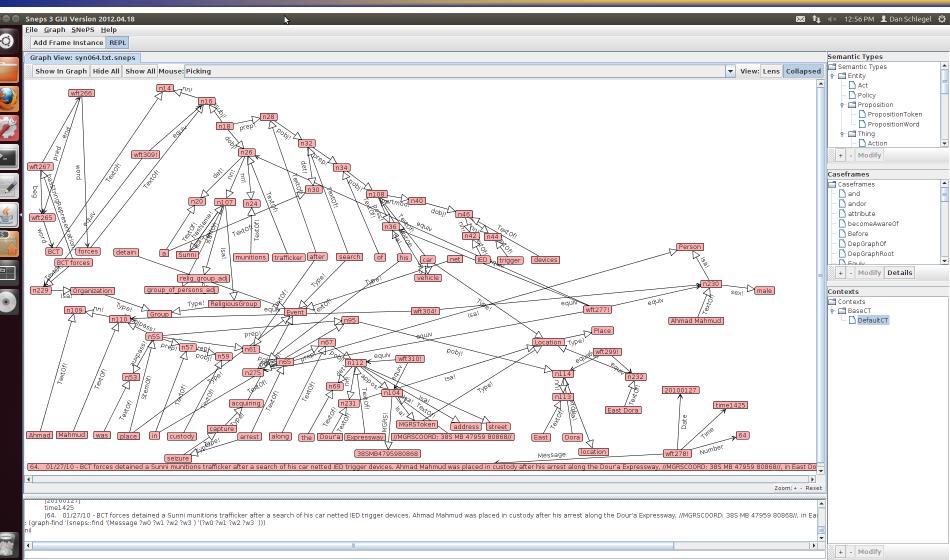




After apposCoreference





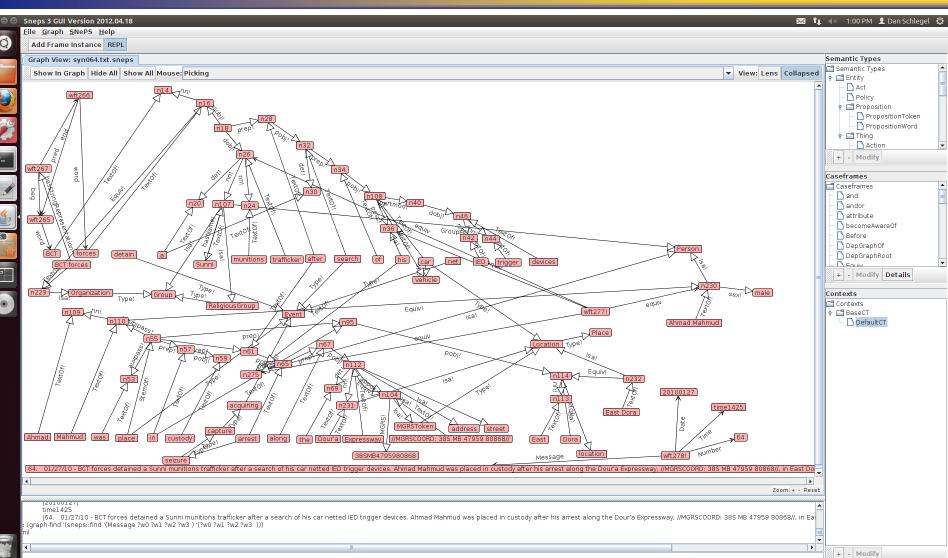




After synsemGroup





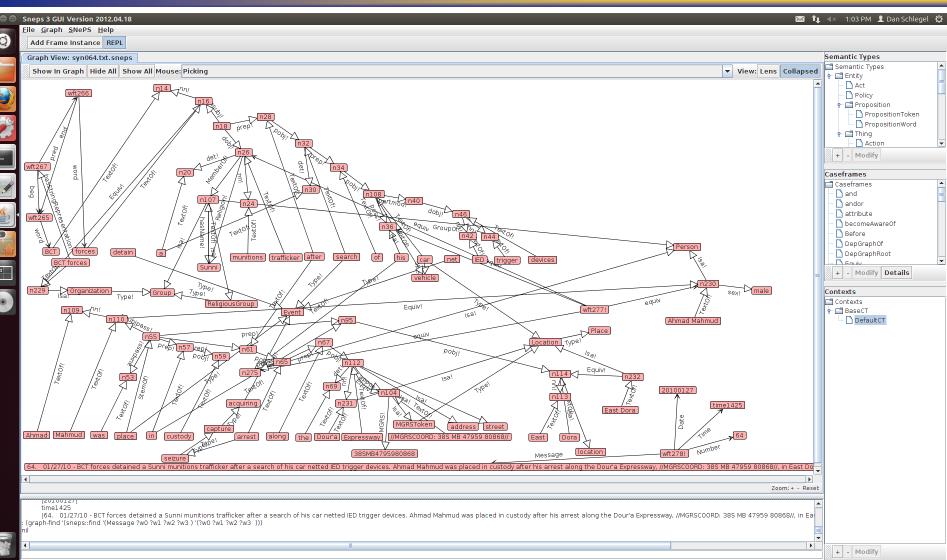




After hasReligion





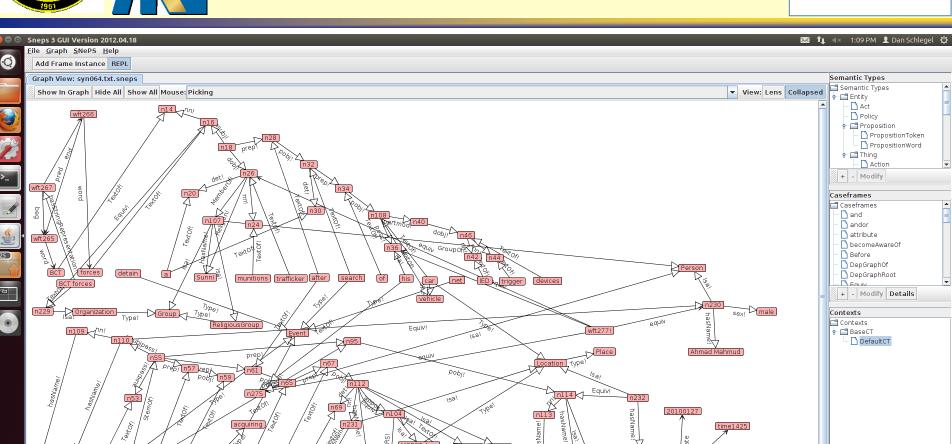




After properNounToName







164. 01/27/10 - BCT forces detained a Sunni munitions trafficker after a search of his car netted IED trigger devices. Ahmad Mahmud was placed in custody after his arrest along the Dour'a Expressway, //MGRSCOORD: 38S MB 47959 80868//, in Ea (graph-find '(sneps::find '(Message ?w0 ?w1 ?w2 ?w3) '(?w0 ?w1 ?w2 ?w3)))

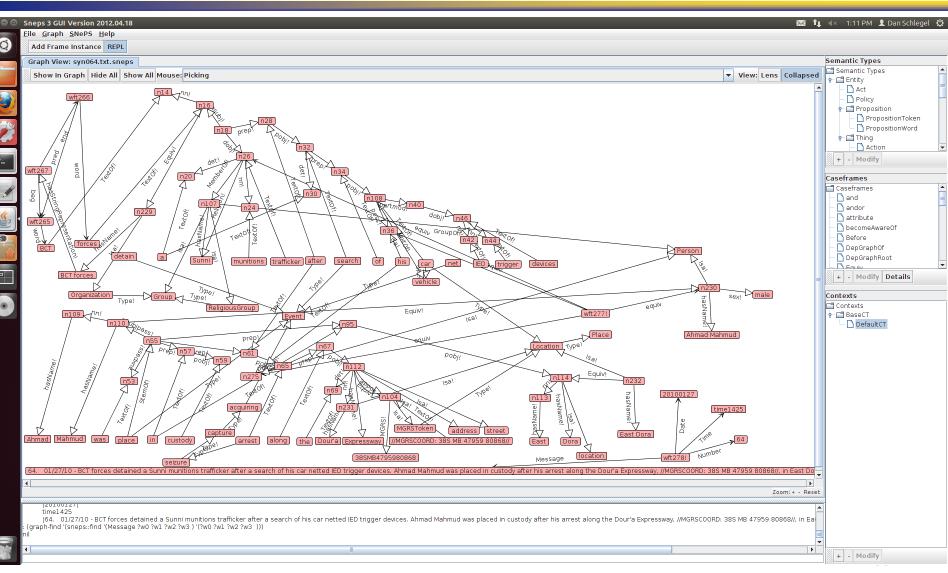


After organizationHasName









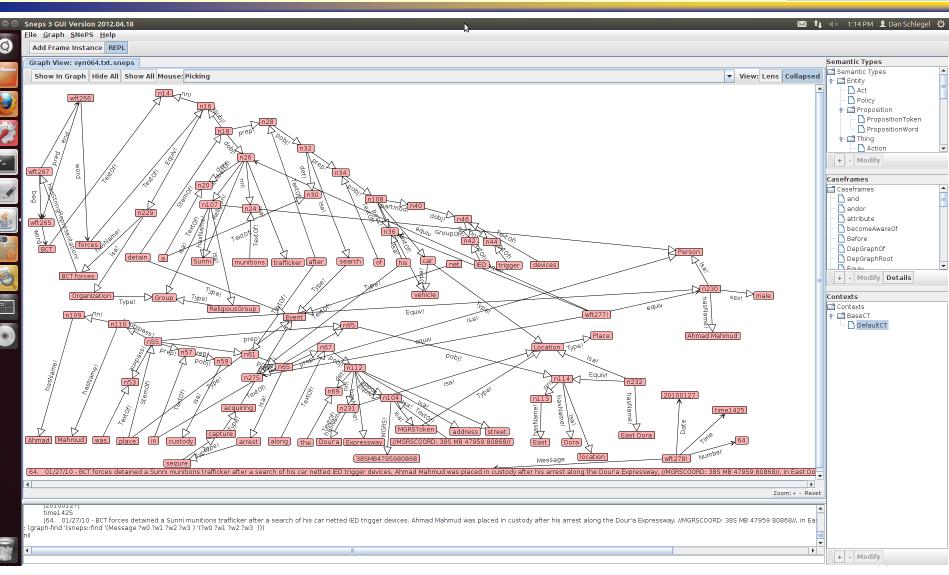


After nounPhraseToInstance







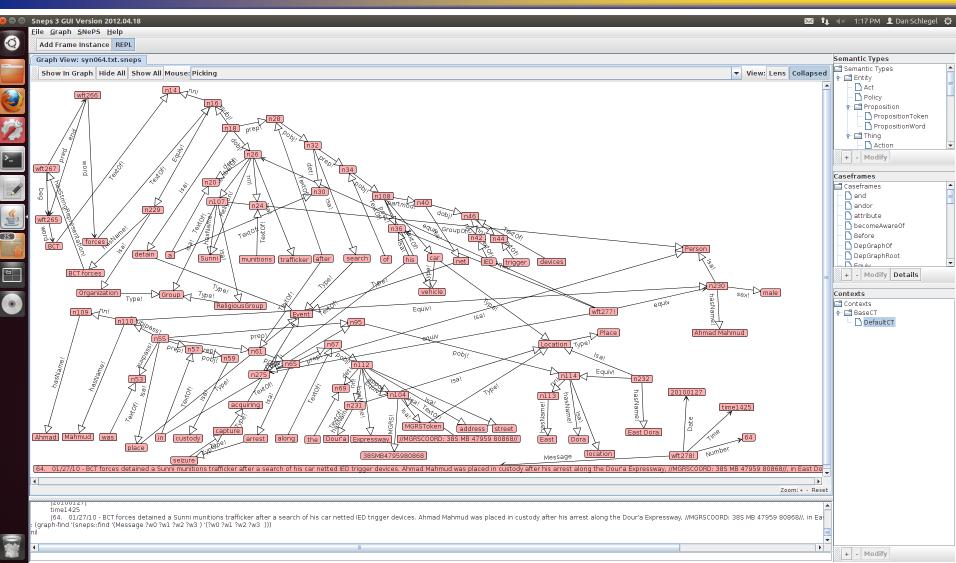




After eventToInstance







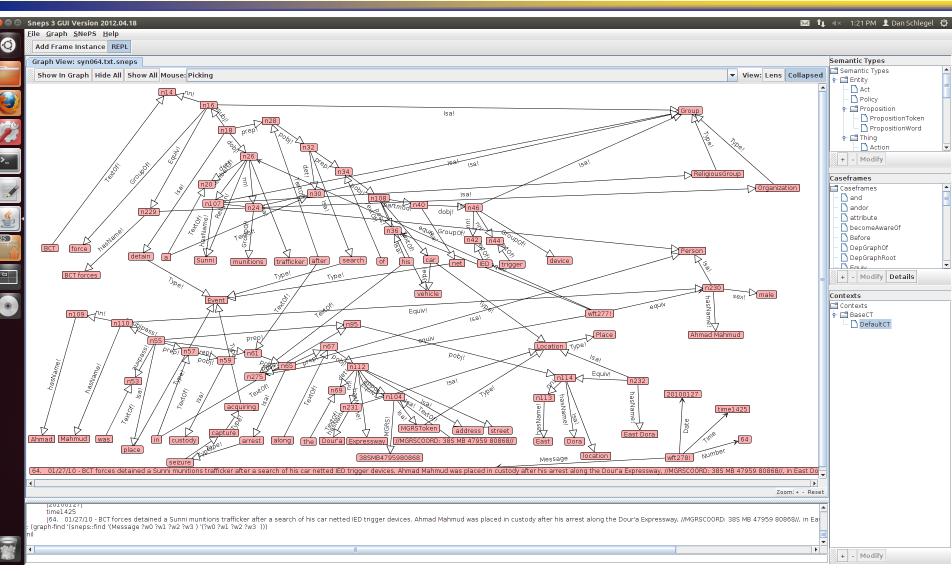


After pluralNounToGroup







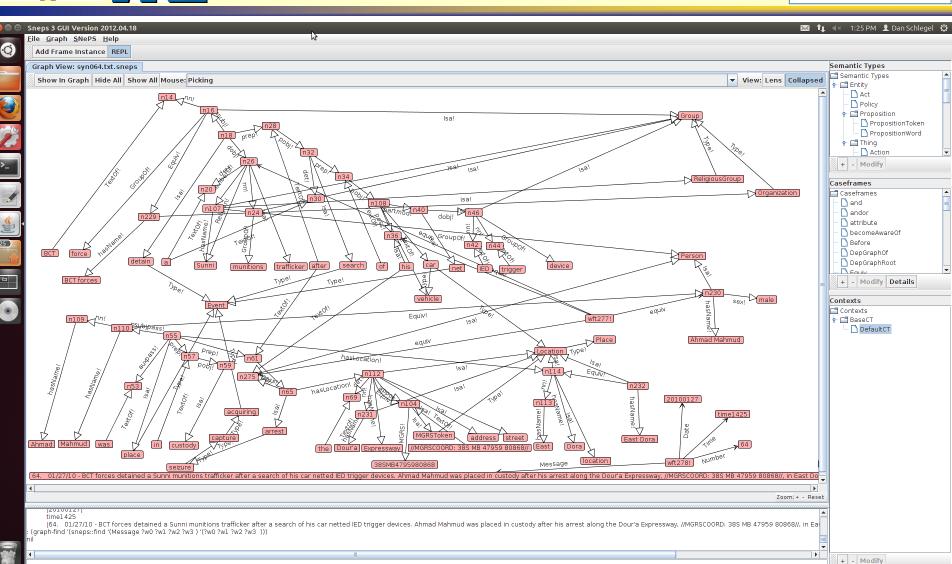




After prepLocation





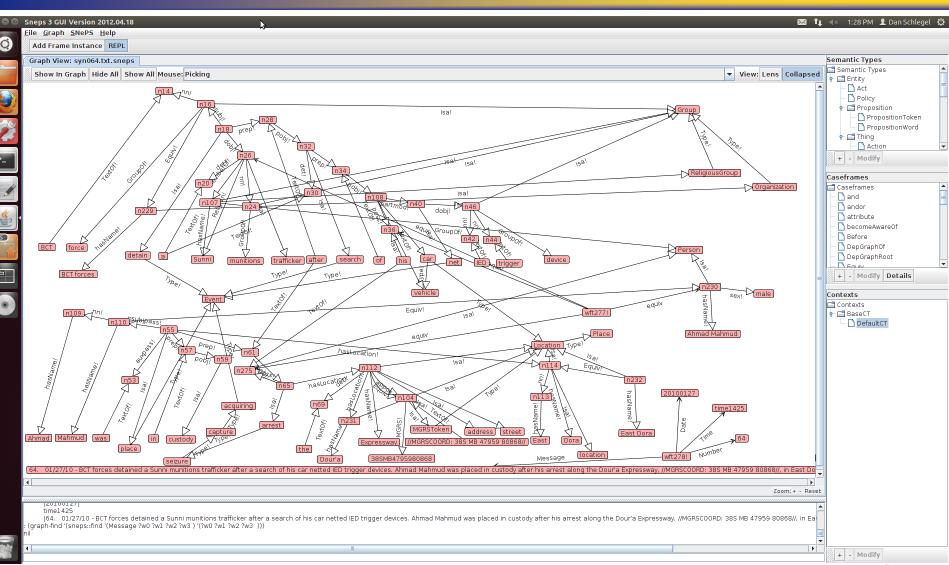




After modLocation





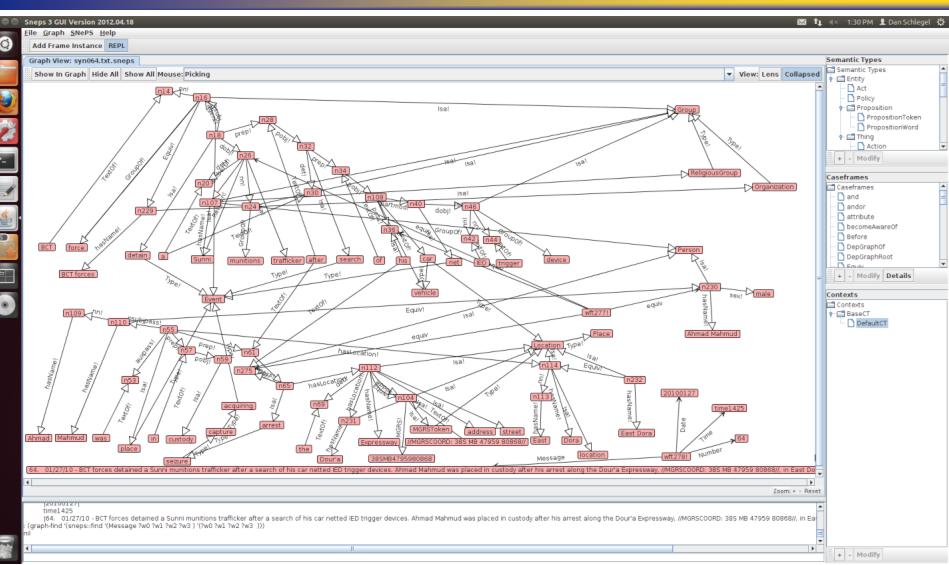




After subjEvent







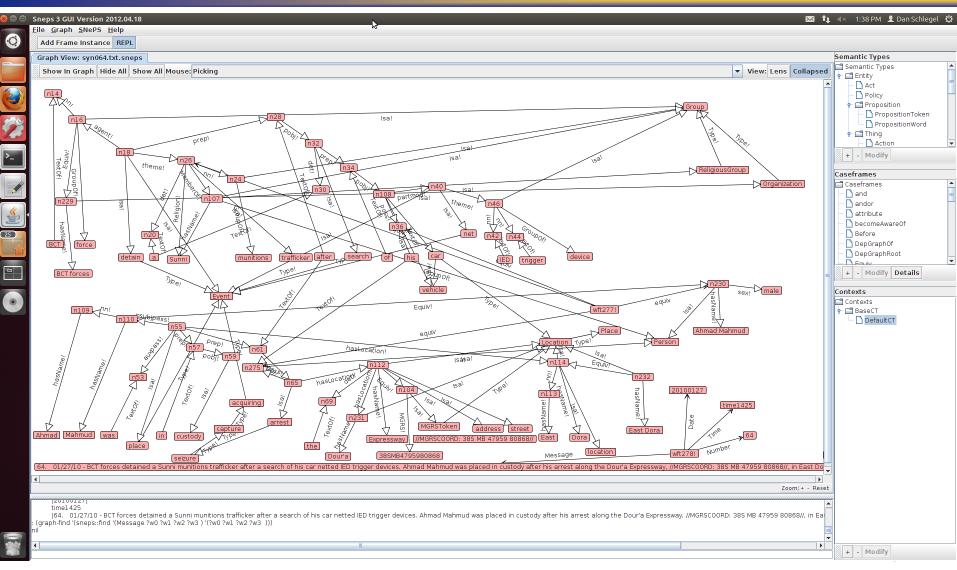


After dobjEvent







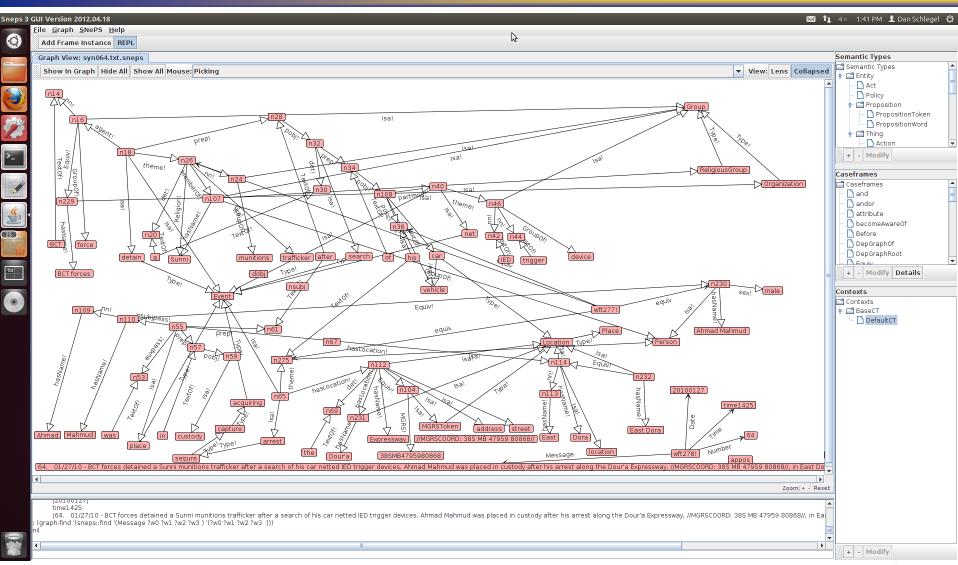




After possessEvent





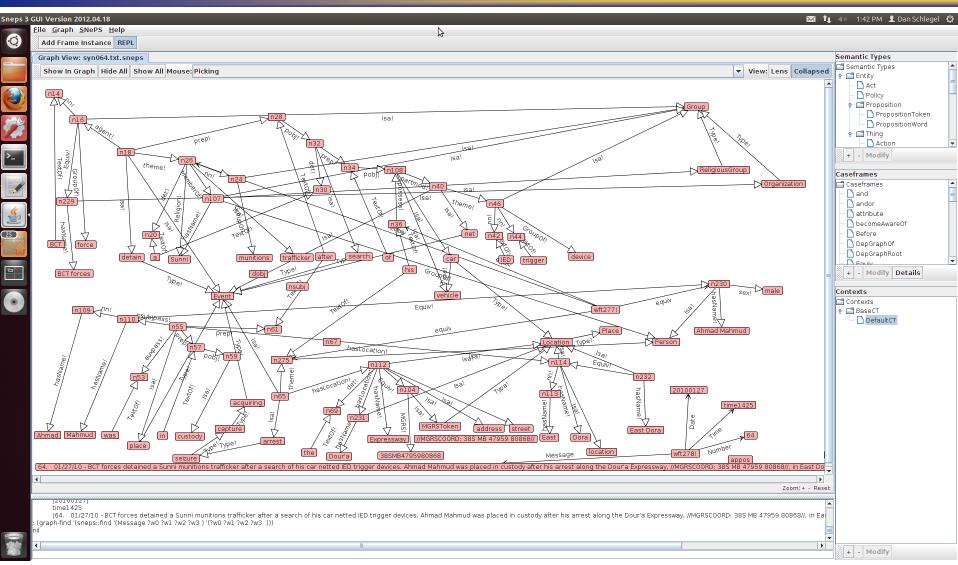




After possess





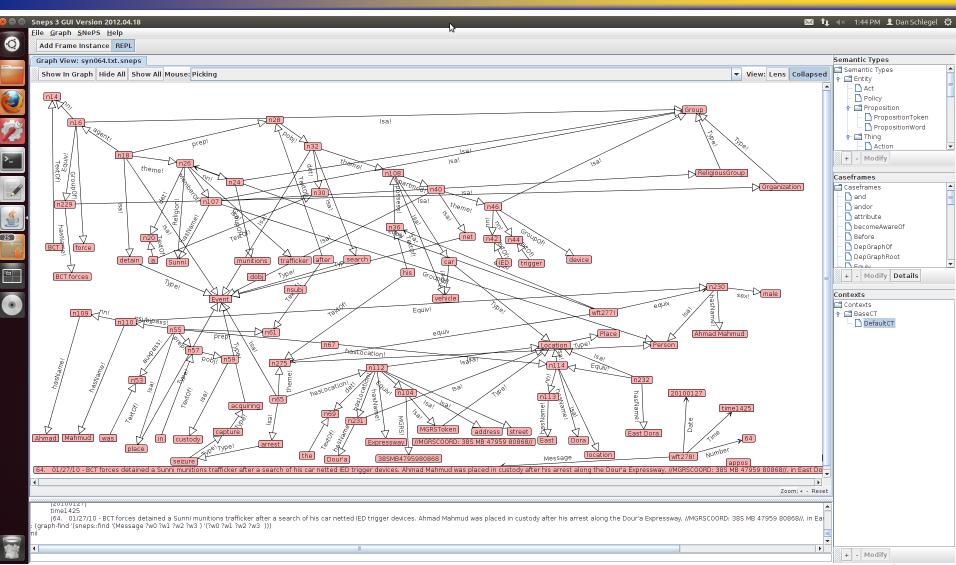




After of Event





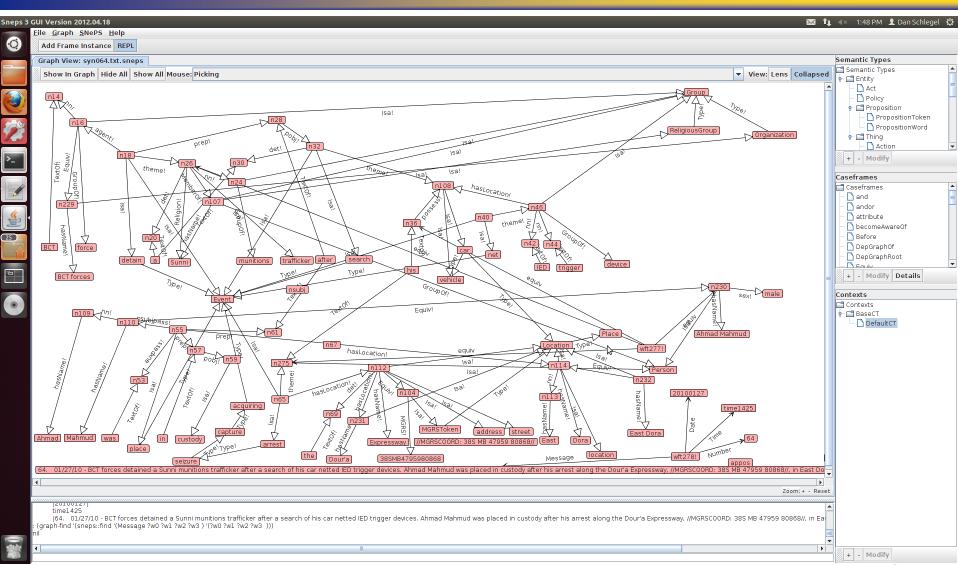




After searchNet





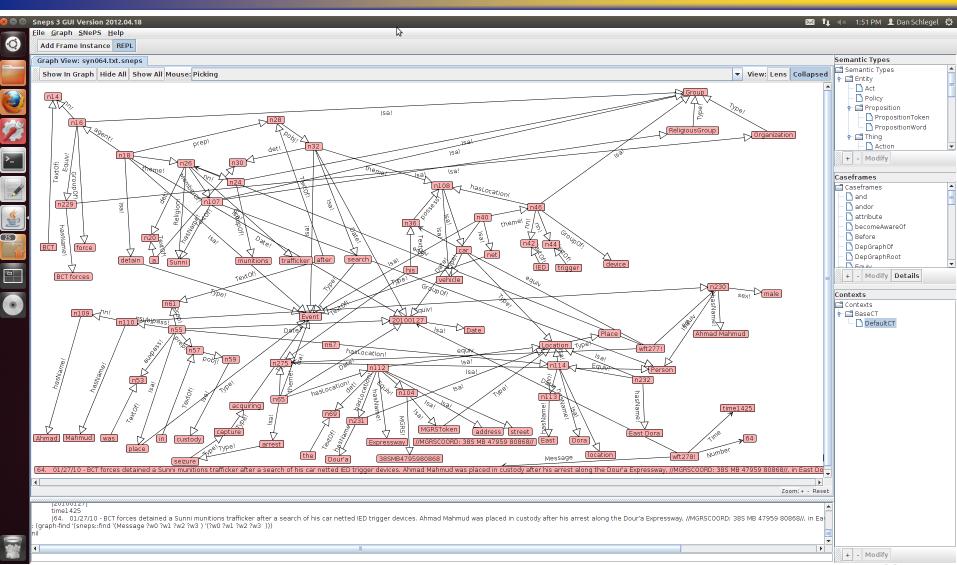




After eventDate







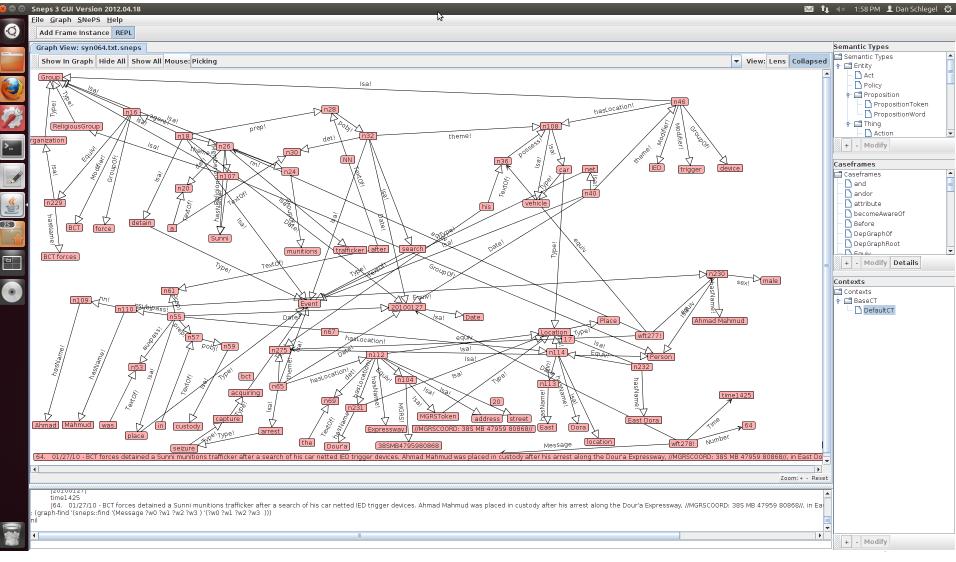


After nnToModifier









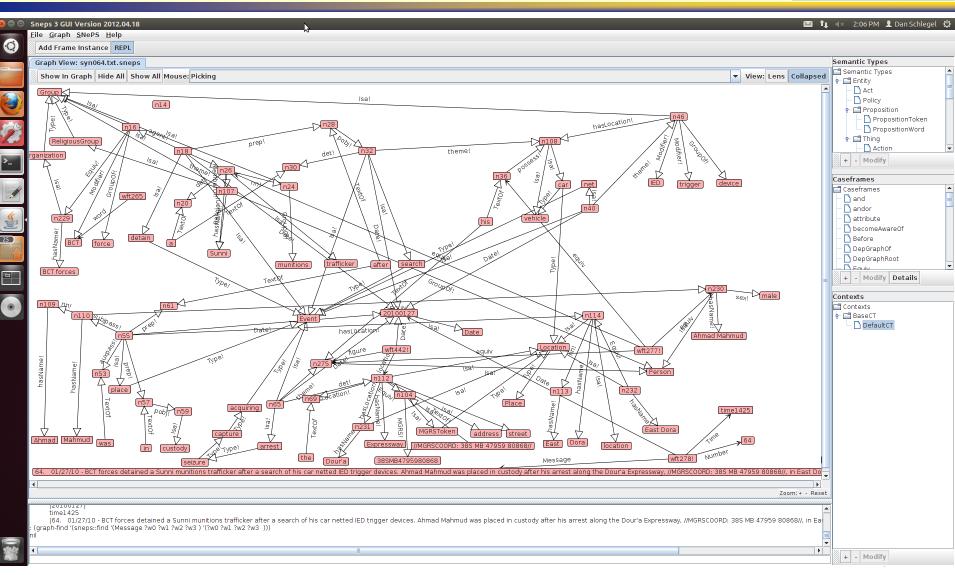


Semantic Propositional Graph











Description Facility Sample Output





|Ahmad Mahmud|

Instance of: (setof Person trafficker)

Ahmad Mahmud | has the relation possess to car108.

|Ahmad Mahmud|'s Religion is Sunni.

Ahmad Mahmud is located at Expressway's Date is 20100127.

|Ahmad Mahmud| is a member of the group, Sunni.

|Ahmad Mahmud| is located at Expressway.

|Ahmad Mahmud|'s sex is male.

arrest65 has the relation theme to |Ahmad Mahmud|.

detain18 has the relation theme to |Ahmad Mahmud|.

Coreference Chain:

n26 55-65 "trafficker"

n36 84-87 "his"

n230 120-132 "Ahmad Mahmud"

n110 126-132 "Mahmud"

n275 161-164 "his"

Arrest65

Instance of: arrest

arrest65 has the relation theme to |Ahmad Mahmud|.

arrest65's Date is |20100127|.

arrest65 is located at Expressway.

Coreference Chain:

n65 165-171 "arrest"



Finding & Describing **Connections** Example: syn089





89. 02/03/10 - The trigger devices which were netted in the arrest of Dhanun Ahmad Mahmud Ahmad on 01/27/10 match materials which were found in the truck of arrested ISG affiliate, Abdul Wahied. BCT analysts are attempting to establish their link.

snuser(78): (describeConnection 'Ahmad 'Wahied)

arrest29 has the relation theme to | Dhanun Ahmad Mahmud Ahmad | net23 has the relation in to arrest29 net23 has the relation theme to Group17 Group51, and Group17 have the relation match to each other find57 has the relation theme to Group51 find 57 has the relation in to truck 63 |Abdul Wahied | has the relation possess to truck63