"DANDE: Deductive Anomaly Detection with Program Synthesis"

January 14, 2003

Sponsored by

Defense Advanced Research Projects Agency (DOD)
(Controlling DARPA Office)

ARPA Order K475/47

Issued by US Army Aviation and Missile Command Under

Contract No. DAAH01-02-C-R

Stephen Westfold
Kestrel Technology LLC
3260 Hillview Ave.
Palo Alto, CA 94304
(650) 320-8888

Effective Date of Contract: 04/09/02
Contract Expiration Date: 12/10/02
Reporting Period: 08/01/02 thru 12/10/02

DISCLAIMER

Deductive Detection of Qualitative Anomalies

Stephen Westfold
Richard Waldinger
Kestrel Technology
3260 Hillview Avenue
Palo Alto, CA 94304

January 10, 2003

Objectives: Detecting Qualitative Anomalies Deductively

The core problem we are facing in this project is detecting qualitative anomalies. Qualitative anomalies are combinations of conditions or events that violate some standards of safety or normalcy. Examples of qualitative anomalies are

- An immigrant with an expired visa applies for a position as director of security at a large hydroelectric power plant.
- A person seeking employment in quality control at a large pharmaceutical company presents a resume that contains employment, residence, and educational history that is not substantiated by available documents.
- A plane is significantly off course and is not in communication with the control tower.
- A company reports dealings with another company which does not exist or that reports no such dealings.
- A person who looks similar on one photographed meeting with a member of Al Qaeda appears as a visitor at one of the largest meat-packing plants in the US.
- A meeting is held between an Iraq official and a member of a terrorist organization.

Such anomalies are distinct from statistical anomalies, and their detection requires symbolic rather than numerical methods.

Similar anomalies appeared before the September 11 attacks, but were only detected after the fact. Such anomalies may be obvious after they are pointed out, but can be buried in large masses of information.

The essence of the current project is to coordinate multiple online data sources, including public records and Web-sites, using a theory of the relevant world knowledge, and deductive inference methods. The theory and methods are incorporated into Specware, the Kestrel software-development environment, Data sources include employment histories and qualifications, educational histories, Visa trails, organizations and affiliations, residence addresses, immigration records, radar tracks, security cameras, and flight-departure information. These sources, which are continually being updated, are linked to the Specware theory via a procedural attachment mechanism; this means that the inference mechanism can draw conclusions from the data sources just as if they were incorporated into the Specware theory. For example, a predicate symbol "Address" could be linked to several online directories, so that if a proof required looking up the address of Mr. John Doe, the system would behave as if an axiom containing the address of John Doe were included in the theory.

In the system we envision, a number of norms have been described in English by intelligence specialists and translated into the Specware theory by means of a natural language parser. The inference mechanism constantly peruses the theory, including the norms and online data sources, in search of contradictions. Contradictions need not be immediately obvious; they may depend on a long chain of inferences. The mechanism is not geared to looking for any one kind of anomaly; even a previously unanticipated combination of conditions that violates a norm leads to the discovery of a contradiction that raises an alarm, which can be examined by a human intelligence specialist, who can weigh their significance. If too many false positives are obtained, the specialist can qualify one or more of the norms to weed some of them out.

In this report, we conclude that deductive methods are very promising for detecting qualitative anomalies, particularly when recognition of the anomaly depends on world knowledge that other approaches may lack.

Approach: Deductive Methods

Our methods employ a specific technology: *automatic deduction* with *witness-finding* and *procedural attachment*. Automatic deduction, or theorem proving, involves drawing logical conclusions from sets of sentences; for this application, we use automatic deduction to try to find contradictions in large sets of sentences. The sequence of inferences leading to the contradiction is called a refutation. Witness-finding, or answer-extraction, is a method for obtaining factual information from a refutation. An explanation of the deductive anomaly could be developed from this mechanism.

The deduction is in the context of an axiomatic theory, a list of logical sentences assumed to be true in the world. This theory has properties of geographical space, time, events, and agents, including people and organizations. Large axiomatic theories are being constructed at several institutions, which can be brought to bear on the problem of finding qualitative anomalies.

Yet it would be foolhardy to rely entirely on facts that have been expressed in a logical language—the effort of translation would be overwhelming, and reasoning in an axiomatic theory is the most time-consuming phase of deductive anomaly detection. Instead, we rely on the technique of procedural attachment, which allows us to tie symbols in an axiomatic theory with external knowledge sources, including data bases and computational procedures. Axioms within the theory advertise the capabilities of the external knowledge sources; when such an axiom is used in a refutation, the corresponding knowledge source is invoked.

For example, the Alexandria Digital Library Gazetteer contains information and maps for about six million places on earth; it would be a massive effort to translate all this into logic, and the Gazetteer is constantly being expanded. By introducing a procedural attachment between our axiomatic theory and a the ADL Gazetteer, and axioms that advertise the capabilities of the Gazetteer, we can access all this information without translating it into logic.

Computations as well as data can be procedurally attached. We can access theorem provers that have special decision procedures for reasoning about time and space. For instance, the Allen temporal procedure allows us to reason about time intervals without representing their properties as axioms. A similar spatial reasoning procedure, RCC8, allows us to reason about geographical regions.

The principal thrust of our work has been the detection of anomalies. The same approach, however, can be applied to answering queries posed by an analyst. The query is phrased as a theorem that is proved by the theorem prover, using the same theory, procedural attachments, and answer extraction mechanism we use for qualitative anomaly detection.

Description of System

The system we are developing, then, has the following structure. A large axiomatic theory of the world is formed, including axioms developed locally and at other institutions. Norms are formulated within this theory. Procedural attachments are developed for external knowledge sources, and axioms are introduced that advertise their capabilities. A class of theorem provers are let loose on this theory, and any contradictions or anomalies that are detected are reported to an analyst, who can judge their significance.

So far we have been experimenting with the following components:

Specware The principal embodiment of the Kestrel's specificationtocode technology, Specware [Kestrel] has advanced capabilities for theory formation and the invocation of theorem provers. Its category-theory base provides a clear conceptual foundation for the combination of multiple theories, which may have disparate vocabularies or ontologies. It is linked to a number of advanced theorem provers, including SNARK and Gandalf. It contains a user interface that allows the knowledgeable Specware user to control the strategies of the various theorem provers, providing high performance in selected subject domains.

SNARK SNARK [Stickel] is an open-source automatic full first-order logic theorem prover developed at SRI International for application in software engineering and knowledge representation. It has a witness-finding mechanism, a procedural attachment mechanism, and special decision procedures for the Allen temporal interval calculus and the RCC8 spatial region calculus. It has strategic controls that enable us to tune it to exhibit high performance in selected subject domains.

Gandalf A highly ranked theorem prover in the biannual competition held at CADE, the International Conference on Automated Deduction, Gandalf [Tammet] exhibits high performance and has a unique ability to cycle through many different combinations of strategies during a single search for a refutation. It was developed at Chalmers University in Gothenburg, Sweden. It has a witness-finding mechanism; a procedural attachment mechanism is being implemented but is not yet available.

The Allen Temporal Interval Calculus Given two temporal intervals, the Allen calculus [Allen] allows us to tell if they overlap, if one precedes the other, etc. Given facts about some temporal intervals, it allows us to deduce other facts. The calculus is based on the thirteen possible primitive relationships between intervals.

The RCC8 Spatial Reasoning Calculus Like the Allen Calculus, but for space rather than time, RCC8 [RCC] allows us to reason about about eight relationships, including overlapping, bordering, discreteness, and inclusion.

The Alexandria Digital Library Gazetteer A repository of information for about six million places on Earth, including countries, cities, airports, factories, military installation, and harbors—for each place, the ADL Gazetteer [ADL] gives a latitude and longitude or a bounding box (its north-, south-, east-, and west-most latitudes and longitudes), and other information. It was developed by the University of California at Santa Barbara, and is accessible via a Web site.

The CIA World Factbook The Factbook [CIA] is an almanac of information about more than two hundred countries, including geographic, economic, governmental and military. We access subsections of the Factbook that have been parsed and translated into SNARK and, soon, DAML. It is available on the CIA Webpage and updated annually.

The TerraVision 3D Terrain Viewer A system for visualizing the terrain of earth, as if in a flight simulator, Terravision [TerraVision] contains data from satellite imagery and elevation measurements. Open source, it was developed at SRI International.

GDACC Mapping Agent GDACC [Goddard] is a system for displaying information based on NASA data stored at the Goddard Space Flight Center. While Terravision shows labeled imagery, the GDACC agent shows maps and line-drawings.

NIMA Mapping Agent The NIMA mapping agent displays NIMA maps, which do not display the same variety of geographic features as the GDACC maps, but are typically of higher quality. They include road maps, terrain maps, and maps of political boundaries.

Geographic Computation Agents A number of agents exist for performing geographic computations, such as given two lat/long pairs, find the distance between them; or given a lat/long, find the lat/long so many miles to the north. Some of these agents are local, but one is accessed via a Web site at Northern Arizona University.

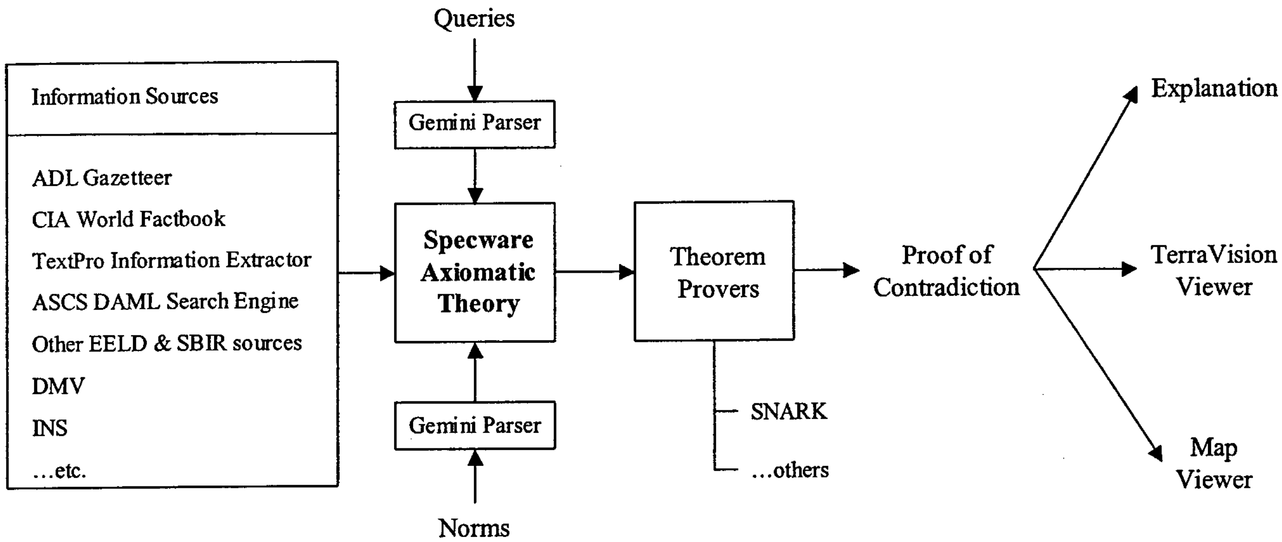

Figure 1: Overview of System Architecture

DAML Agent Semantic Communication Service (ASCS) This [ASCS] is a search engine, developed by Teknowledge under the DARPA DAML program, that allows us to query all pages available on the Web annotated with mark-up in the DAML language.

TextPro Information-Extraction System This [Discern] allows us to search large bodies of text for facts that participate in qualitative anomalies; developed by Discern.

Other Agents to be Included Several other agents will be included in continuation of this work.

- The Gemini [Gemini] natural language understanding system, a unification-based grammar developed at SRI, translates text into a logical form; it will allow us to introduce into the axiomatic theory facts and norms expressed in English and translated automatically into logic.
- License-Plate Recognition software developed at SRI.
- Web-based agent for finding addresses of people, given their names.
- Web-based agent for finding the lat/long corresponding to a given address.
- Sample employment, education, drivers license, and immigration data.

A view of the architecture of the experimental system we have implemented occurs in Figure 1.

Example

Here is an example that our current experimental system can carry out. Suppose we include in our axiomatic theory the norm that members of terrorist organizations do not meet with government officials. (This does not mean that such meetings never occur, only that we want to be alerted if they do.) Suppose we have access, via procedural attachments to external sources, the following facts:

- Al Qaeda is a terrorist organization.
- Mohammed Atta is a member of Al Qaeda.
- Mohammed Atta was at the Ruzyne Airport, Czech Republic, June 2-3, 2000.
- Ahmad Khalil Ibrahim Samir al Ani was in Prague between March, 1999, and April 22, 2001.
- Ahmad Khalil Ibrahim Samir al Ani was employed by the Mukhabarat, the Iraq Intelligence Service.

From this information and procedurally attached sources, the experimental system is able to detect an anomaly, which depends on the fact that the two time intervals overlap (deduced by the Allen interval calculus) and that Ruzyne Airport and Prague are only seven miles apart (deduced from latitudes and longitudes supplied by the ADL gazetteer and distances computed by the Northern Arizona University Web site.) The locations of Prague and Ruzyne are displayed by Terravision and the GDACC mapping agent.

This example is described in more detail in the Appendix: Detection of the Mohammed Atta anomaly.

Clearly, similar reasoning could be used to determine that two people could have met if they both were in the same restaurant at the same time, or that two people could have known each other if they attended the same university in overlapping years. Other inferences can detect people with terrorist history occupying safety-critical roles, and other anomalous situations.

False Positives

A problem with automated approach can occur if too many "anomalies" are discovered that, on closer inspection, are not at all ominous. If two many such *false positives* are obtained, the system's valid warnings will tend to be disregarded. It will require too much human effort to sort through them to determine which merit further investigation.

In our system, false positives can occur if a norm is formulated too strongly and requires qualification. For instance, whereas it may be anomalous for a person's home address to be greatly distant from his or her work address, telecommuting is becoming more common and is not evidence of falsification of documents.

Whereas a hard-coded system is difficult to modify, an axiomatic system is relatively easy. Another way of putting this is that reprogramming can occur at the specification level rather than at the software level. Norms are expressed as axioms and can easily be changed. For instance, the norm that people live relatively near their place of employment can be qualified to exclude telecommuters, migrant workers, and other common exceptions, with very little effort.

Use with other approaches

We have present this work as complementary to statistical and pattern-matching approaches. In fact, it is natural to use our approach in combination with other approaches.

- A statistical or pattern-matching based system can be used as one of the information sources, providing components of a larger anomaly. For instance, suppose a source has discovered a statistically significant concentration of airline security workers at a particular airport who have privately purchased infrared night-viewing glasses. The significance of this fact in isolation may be difficult to appreciate. However, the fact can be input to the theory via a procedural attachment and used in the discovery of a larger qualitative anomaly involving a sabotage scheme.

- Our approach can be used to fill in gaps if another approach has discoverd a partial anomaly, but hasn't been able to complete it. For instance, a mass murder may be considered to have political significance if the murderer had regular meetings with a member of a terrorist group, a qualitative anomaly.

Metrics

There are a number of natural metrics for the evaluation of this work. They include:

- Number of (a) planted (b) natural anomalies detected. (A "planted" anomaly is one inserted deliberately into the knowledge sources for the purpose of evaluating the detection capability.)

- Seriousness of the anomaly detected. (This can be judged in consultation with an intelligence analyst.)

- Number of false positives detected.

- Time required to detect anomalies.

- Number of knowledge sources integrated into the system.

One potential metric for DANDE that includes a number of the key aspects above is the so-called Receiver Operating Characteristic (ROC). The ROC is commonly used as a framework for comparison in decision processes where measurement noise is present. It may be meaningful to extend the basic idea to the problem of normative presentations of various anomaly detection systems. For that matter, it could also be used in EELD applications too, as well as in combinations of systems, e.g., where one system uses another

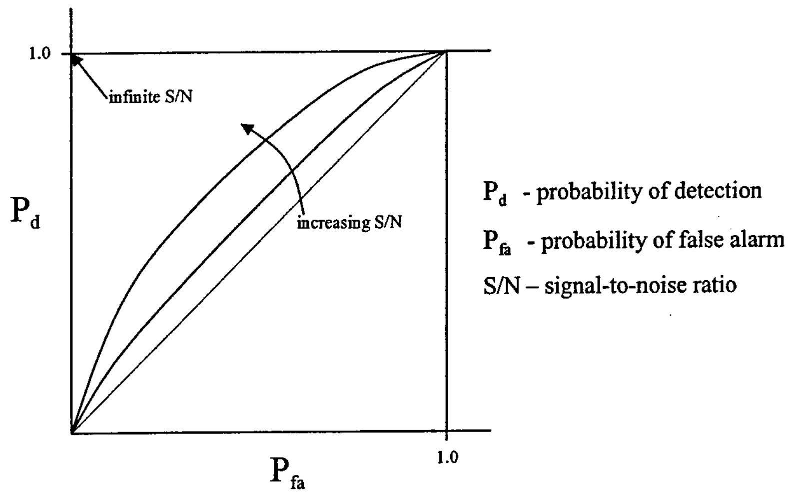

Figure 2: Receiver Operating Characteristic

as a knowledge source. The basic idea is that for any particular configuration, one presents the locus of probability of detection vs. the probability of false alarm (in our case, false positive). This relationship is parametrized on a appropriate measure of the "strength" of the "signal" we want to detect compared with the "strength" of the uncertainty (noise) in the environment in which we are looking for "signal". This is illustrated in Figure 2.

One of the benefits of the ROC as a metric is that it can often be generated analytically from the underlying statistics. Even when not, monte carlo methods can often suffice.

Comparison with Other Approaches

Most other approaches to the problem of anomaly detection are based on either statistical or pattern-matching methods; the one outlined here is based on logical inference. Statistical methods are appropriate for finding statistical anomalies but not qualitative ones. Neither approach is better than the other; they are complementary. Pattern-matching approaches are likely to be faster than those based on inference, but not so likely to find deeper or more subtle anomalies, which depend on world knowledge or the statements of norms.

Perhaps closest to our approach is that based on Cyc [Cycorp], the knowledge repository that has been under development for many years by CycCorp. The key difference between the Cyc approach and this is that Cyc emphasizes knowledge, while we emphasize inference. Cyc's knowledge-base is quite large and many aspects of its representation are well thought out. Our own knowledge-base incorporates much of Cyc's upper ontology. But the Cyc system does not have a powerful general inference mechanism. Rather, its knowledge base is subdivided into many micro-theories. Cyc incorporates special-purpose mechanisms for doing fast inference within a particular micro-theory. This makes Cyc better at doing "theory-specific" inference, which works within a single micro-theory, than "cross-theory" inference, which links many seemingly unrelated sub-theories.

For the purpose of detecting previously unanticipated anomalies, however, it is essential to do cross-theory inference. For example, there might be one micro-theory dealing with employment, and other dealing with immigration, but to discover anomalies that involve both requires reasoning across the two micro-theories. In general, while theory-specific reasoning may be faster for answering question within a single micro-theory, more general high-performance cross-theory reasoning may be necessary to uncover surprises.

Future Possibilities

In our future research we will need to address the following issues.

- Procedural attachment of new knowledge sources. Development of tools to make this easier.
- Combination of multiple ontologies and theories. We have already incorporated much of the Cyc Upper Ontology, but we would also like to introduce the Teknowledge Suggested Upper Ontology [Teknowledge], which has more axioms. The capabilities of Specware will be required here.
- We will not want to search large data sets during the theorem-proving process. Rather we would like to synthesize a program to search the data, optimize the program, and execute it after the refutation is complete. This requires Specware software development and optimization capabilities.
- If norms are carelessly formulated, we may find too many false positives. There must be tools by which an analyst can then elaborate on a norm to make it more realistic. For instance, a norm that people live near their place of employment may be violated by the rising prevalence of telecommuting.
- Some data may be classified or difficult to access. The use of Specware program analysis technology will enable us to prove that the accessed data will only be used for particular purposes and will not be accessible otherwise. This would be based on Kestrel's FLAWS technology. This is decribed further in the following section. At any rate, Kestrel Technology has a facility clearance to the Top Secret level, and applications pending for SCI clearances, and hence can deal with some of those sources.
- Accessing some of this data may violate privacy concerns. Use of Specware technology will enable us to prove that the accessed data will only be used for a specified purpose and will not be accessible for other reasons. This also would exploit Kestrel FLAWS technology. In general, the use of inference should enable us to uncover more anomalies without violating privacy rights.
- Efforts will be made to relate our activities with those of other SBIR and EELD participants.

Use of FLAWS Technology

Kestrel has developed a technology under the auspices of an NSA project that could be applied effectively in the EELD program, in particular, within a Phase II SBIR. This technology, called FLAWS, was originally intended to help an analyst determine whether a Java application could be exploited over the network. One example threat would be hostile applets. The FLAWS technology tackles the problem of uncovering feature interactions within an application, where no single feature is adequate for an exploit, but where the combination can open the unwitting host to serious damage. An important observation is that FLAWS's design supports the technology's use in applications where privacy protection could be a concern. Specifically, it is possible to build a Java program and to prove that it will do A, B, and C, but not X, Y, and Z. This capability depends upon being able to express A, ..., Z in suitable formal representations. However, we believe that many important cases can be handled with this approach, and that a search warrant could be granted based upon the mathematical assurances of correct behavior

Conclusions

Our preliminary conclusion is that deductive methods are extremely promising for the detection of qualitative anomalies. Although these methods do not detect statistical anomalies, they can accept as knowledge sources other packages, which notice statistical anomalies. And, though they may not be as fast as pure pattern-matching methods, they may detect anomalies that pattern-matching methods will miss, particularly when recognition of an anomaly depends on world knowledge.

Appendix: Detection of the Mohammed Atta Anomaly

This gives a detailed logical description of the detection of a anomaly: a meeting between an Iraqi intelligence officer and a member of a terrorist organization.

Suppose we wish to be alerted if any member of a terrorist organization is in a position to meet with an Iraqi intelligence agent. Then we assert the following norm:

```
(assume '(not
           (and
            (could-have-met-in-place
              ?person1 ?person2 ?time-interval ?region)
            (employed-by ?person1 ?organization1)
            (terrorist ?organization1)
            (employed-by ?person2 ?organization2)
            (subsidiary-of ?organization2 (government-of iraq))
            (intelligence-service ?organization2)))
        :answer
        '(val ?person1 of ?organization1
              could have met iraqi official
```

```
?person2 of ?organization2
in ?region at ?time-interval))
```

In other words, it is regarded as anomalous if two people could have met if one of them is employed by a terrorist organization and the other is employed by the Iraq intelligence service. If this should occur, we wish to be notified of the place and time that meeting could have occurred.

The relation 'could-have-met-in-place' obeys the following axiom:

```
(assert
 '(implied-by
   (could-have-met-in-place
     ?person1 ?person2 ?time-interval ?region1)
   (and
    (in ?person1 ?region1 ?time-interval1)
    (in ?person2 ?region2 ?time-interval2)
    (near ?region1 ?region2)
    (time-ii-intersects ?time-interval1 ?time-interval2)
    (time-ii-intersects ?time-interval1 ?time-interval)
    (time-ii-intersects ?time-interval2 ?time-interval)))
 :documentation
 "Two people in nearby regions at the same time
  could have met."
 :name 'people-in-same-place-at-same-time-could-have-met-there)
```

In other words, if one person is at a certain place and time, and a second person is at another place and time, and if the places are nearby and the time intervals in question intersect each other, then they could have met at that place and time.

The 'near' relation is defined by the following axiom:

```
(assert
   '(iff
     (near ?region1 ?region2)
     (=< (distance-between ?region1 ?region2) (miles 50)))
   :name 'definition-of-near)
```

In other words, two places are regarded as near if they are within 50 miles of each other.

Now assume that we learn the following facts, which are potentially accessible from information extraction systems:

Mohammed Atta, a member of Al Qaeda (a terrorist organization), went to Ruzyne, Czech Republic, on June 2, 2000, stayed over, and departed on June 3.

```
(assert
 '(in
   mohammed-atta
   (feature populated-place ruzyne czech-republic)
```

```
    (date-interval 2000 6 2 :until 2000 6 3))
  :documentation
  "Mohammed Atta stopped over at
   Ruzyne, Czech Republic, June 2-3, 2000."
  :name 'mohammed-atta-was-at-ruzyne)

(assert '(employed-by mohammed-atta al-qaeda)
        :documentation
        "Mohammed Atta was employed by al Qaeda"
        :name 'mohammed-atta-employed-by-al-qaeda)

(assert '(terrorist al-qaeda)
        :documentation
        "Al Qaeda is a terrorist organization"
        :name 'al-qaeda-is-terrorist)
```

Furthermore, Ahmad Khalil Ibrahim Samir al Ani was at the Iraqi embassy between March, 1999, and April 22, 2001.

```
(assert '(at ahmad-khalil-ibrahim-samir-al-ani
             (embassy-of iraq czech-republic)
             (date-interval 1999 3 :until 2001 4 22))
        :documentation
        "Ahmad Khalil Ibrahim Samir al Ani
         was at the czech embassy to iraq
         from March, 1999, until April 22, 2001"
        :name 'al-ani-was-at-embassy)
```

Mr. al Ani was a member of the Mukhabarat, the Iraq Intelligence Service.

```
(assert '(employed-by ahmad-khalil-ibrahim-samir-al-ani mukhabarat)
        :documentation
        "Ahmad Khalil Ibrahim Samir Al-Ani
         is employed by the Mukhabarat."
        :name 'ahmad-employed-by-mukhabarat)

(assert '(and
           (subsidiary-of mukhabarat (government-of iraq))
           (intelligence-service mukhabarat))
        :name 'mukhabarat-is-iraqi-intelligence-service
        :documentation
        "The Mukhabarat is the Iraq intelligence service.")
```

Although in our initial experiments these facts were expressed in logic, in more recent experiments we have extracted some of them from online text, using the TextPro information-extraction system.

Note that syntactic information in not enough to detect an anomaly here. None of the information we are given so far tells us that Ruzyne is anywhere near Prague. Furthermore, semantic processing of dates is necessary to determine that the stay of Atta in Ruzyne overlaps temporally with al Ani's term in Prague. Nevertheless, deductive inference, with SNARK geographic and temporal reasoning are enough to establish the anomaly.

Decomposing the assumed norm and employing the axiom for the relation couldhave-met, SNARK first determines that Mohammed Atta was a member of al Qaeda and visited the airport at Ruzyne, Czech republic. In more recent experiments, this fact has been obtained by information extraction from online textual documents by TextPro:

```
person_in_place('Mohamed Atta',
 airport,'Ruzyne',_9559,'Czech Republic')
```

Note that the information was obtained using a common alternative spelling of Atta's name.

The Alexandria library is invoked, by procedural attachment, to determine the bounding box of the Czech republic and, then, the latitude and longitude of Ruzyne.

```
place_to_latlong('Czech Republic',countries,'Czech Republic',countries,
 5.1419998E+01,4.823E+01,1.9379999E+01,1.213E+01)

place_to_latlong_partof_bounds
 ('Ruzyne','populated places', 'Czech Republic',
  '51.42','48.23','19.38','12.13',
  'Ruzyne','populated places','Czech Republic',
  5.0083332E+01,5.0083332E+01,1.4316667E+01,1.4316667E+01)
```

It is similarly determined that al-Ani was a member of the Secret Service and was stationed at the Iraqi Embassy to the Czech Republic. It is general knowledge that embassies are located in the capitals of their host countries; this is expressed by an axiom. The CIA World Factbook, accessed through the ASCS system, determines that the capital of the Czech republic is Prague.

```
ascs_query_conjunct('421','A33249',
 'http://www.daml.org/2001/12/factbook/factbook-ont',
 name,'2','Prague')
```

Finally, the Alexandria Digital Library determines the latitude and longitude for Prague.

```
place_to_latlong_partof_bounds
 ('Prague','populated places','Czech Republic',
  '51.42','48.23','19.38','12.13',
  'Praha',capitals,'Czech Republic',
  5.0083332E+01,5.0083332E+01,1.4466667E+01,1.4466667E+01)
```

Then the Northern Arizona University lat/long computation Web site is invoked to determine that Prague and Ruzyne are actually less than seven miles apart.

```
lat_long_dist('50.083332N','14.316667E','50.083332N','14.466667E',
 '6.6576')
```

This is well within the fifty-mile limit that occurs in the definition of "near."

SNARK's implementation of the Allen temporal interval calculus determines that the time intervals in question do indeed overlap, and hence the meeting could have occurred.

References

[ADL] Hill, L. L., Zheng, Q. Indirect geospatial referencing through place names in the digital library: Alexandria Digital Library experience with developing and implementing gazetteers. *Proceedings of the American Society for Information Science Annual Meeting*, Washington, D.C. (Oct. 31–Nov. 4, 1999), 57–69. http://www.alexandria.ucsb.edu/~lhill-paper_drafts/ASIS99_confpaper2_final.pdf

[Allen] J. F. Allen, Time and Time Again: The Many Ways to Represent Time, *International Journal of Intelligent Systems*, Vol. 6, No. 4 (July 1991), pp. 341–355.

[ASCS] Teknowledge. DAML Agent Semantic Communication Service. http://-reliant.teknowledge.com/DAML

[CIA] Central Intelligence Agency. *The World Factbook (2002).* http://-/www.cia.gov/cia/publications/factbook/index.html

[Tammet] Tammet, T. Gandalf. *Journal of Automated Reasoning* **18** 2 (1997) 199–204.

[Gemini] Dowding, J., Gawron, J. M., Appelt, D., Bear, J., Cherny, L., Moore, R., and Moran, D.: GEMINI: A natural language system for spoken-language understanding. *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics* (June 1993) 54–61.

[Cycorp] Lenat, D. B., and Guha, R. V.: Enabling Agents to Work Together. *Communications of the ACM* **37** 7 (July 1994)

[Goddard] Goddard Space Flight Center. Earth Sciences Distributed Archive Center. http://daac.gsfc.nasa.gov

[RCC] Randell, D. A., Cui, Z., Cohn, A. G., A Spatial Logic Based on Regions and Connection. *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning*, Morgan Kaufmann (1992) 165–176.

[Stickel] Stickel, M., Waldinger, R., Chaudhri, V. *A Guide to* SNARK. SRI International Artificial Intelligence Center (2000). www.ai.sri.com/hpkb/-snark/tutorial/tutorial.html

[Kestrel] Srinivas, Y. V., and Jullig, R. Specware: Formal Support for Composing Software. *Proceedings of the Conference on Mathematics of Program Construction*, Kloster Irsee, Germany, July 1995.

[Teknowledge] Pease, A., Niles, I. IEEE Standard Upper Ontology: A Progress Report. *Knowledge Engineering Review, Special Issue on Ontologies and Agents* **17** (2002).

[TerraVision] M. Reddy, Y. G. Leclerc, L. Iverson, and N. Bletter. TerraVision II: Visualizing Massive Terrain Databases in VRML. *IEEE Computer Graphics and Applications (Special Issue on VRML)* **19** 2 (1999) 30–38.

[Discern] Appelt, D., Martin, D., Named Entity Extraction from Speech: Approach and Results Using the TextPro System. *Proceedings of the DARPA Broadcast News Workshop,* Herndon, Virginia, February 28–March 3, 1999. `www.nist.gov/speech/publications/darpa99/`