Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2014-022 October 8, 2014

An Analyst's Assistant for the Interpretation of Vehicle Track Data

Gary Borchardt, Boris Katz, Hong-Linh Nguyen, Sue Felshin, Ken Senne, and Andy Wang

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

An Analyst's Assistant for the Interpretation of Vehicle Track Data

Gary Borchardt, Boris Katz,
Hong-Linh Nguyen, Sue Felshin
MIT Computer Science and Artificial Intelligence Laboratory

Ken Senne, Andy Wang
MIT Lincoln Laboratory

1. Overview

Interpretation of sensor data is a challenging task, whether attempted by man or machine. The volume of data presented for interpretation can often preclude the use of time-intensive analysis techniques. Sensor channels by their very nature focus on targeted types of information and are unable to obtain complementary information that might aid in interpretation. Additionally, human experience and world knowledge are often required to associate sensed phenomena with likely occurrences in the world.

This report describes a software system, the Analyst's Assistant, created in a project effort involving MIT CSAIL and MIT Lincoln Laboratory from 2008 to 2012. Design and implementation of the Analyst's Assistant has explored synergies possible in collaborative, language-interactive human–system interpretation of sensor data, specifically targeting interpretation of events involving vehicles in urban and rural settings on the basis of vehicle track data—sequences of timestamped positions for sensed vehicles. We have used this project as a means of elaborating and demonstrating a number of emerging intelligent systems techniques, both individually and in combination. These are:

- a strategy for performing automated event recognition by matching scene information to language-based representations of the temporal unfolding of events,
- a technique for summarizing sequences of events by minimizing redundancy in the coverage of scene information,
- a user-driven approach to partitioning and scheduling subtasks in collaborative user–system interpretation of data, facilitated by a robust natural language interface, and
- a suite of capabilities for coordinating the handling of natural language and GUI references to scene entities during an analysis session.

Together, these techniques enable the Analyst's Assistant to benefit from both human and software contributions to interpretation. In large part, this is made possible by infusing natural language throughout the design of the Analyst's Assistant. Natural language is used not only as a key channel of interaction between the human and system, but as well, the internal representation of the system is designed from the lowest level using language-motivated concepts, and knowledge of what happens during particular types of events is also coded in a language-motivated way. This allows the internal operation of the Analyst's Assistant to be intuitively understandable to humans, so that human and system may interact on a number of levels, not just concerning the end products of interpretation.

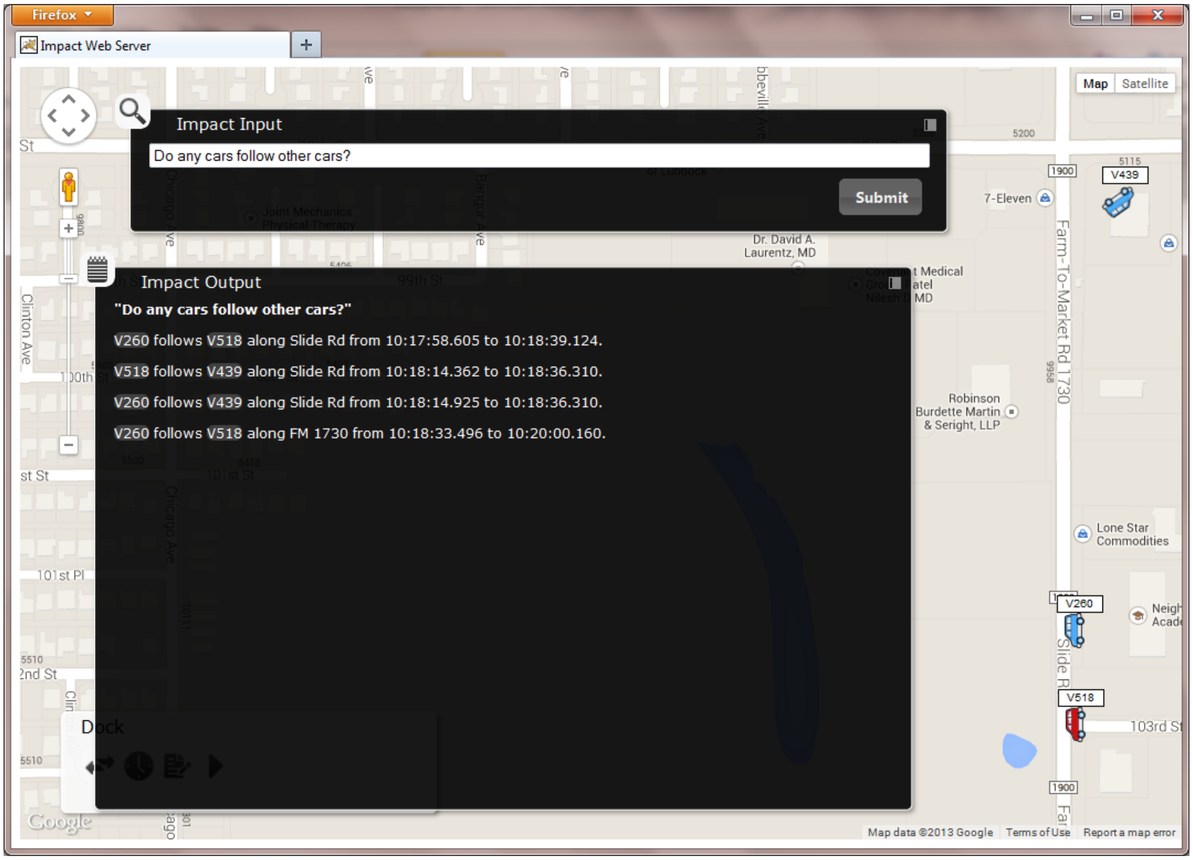
Within this report, Section 2 outlines the functionality and architecture of the Analyst's Assistant, which integrates and extends functionality provided by the IMPACT reasoning system, the START natural language processing system, and a Web-based GUI interface. Next, Section 3 presents specific contributions of the Analyst's Assistant concerning event recognition, summarization of events, language-facilitated user–system collaboration, and support for handling multi-modal references to scene entities. Section 4 summarizes the effort to date, and the report ends with two appendices: a listing of "event models" used for event recognition within the Analyst's Assistant, and a record of an extended interaction session with the system regarding the uncovering of an IED emplacement activity sequence carried out by vehicles within the dataset utilized by the Analyst's Assistant.

2. Functionality and Architecture

The Analyst's Assistant is designed to support collaborative interpretation of end-to-end geo-spatial track data. This type of data can be collected from a variety of sources, including remote sensing and by cooperative transmissions. The particular dataset used for the development and demonstration of the Analyst's Assistant was collected by Lincoln Laboratory in 2007 during an experiment which involved dozens of vehicles engaged in scripted activities in urban and rural areas in and around the Lubbock, Texas area. Added background clutter, in the form of hundreds of vehicle tracks in the vicinity of the scripted activities, was hand selected for a 2 1/2 hour period. In addition, the Analyst's Assistant has been supplied with a database of publicly available geographic information system (GIS) data, which specifies approximately 25,000 named and categorized road segments (highway, local road, other) for the region in and near Lubbock, Texas.

A typical interpretation goal for analysts working with vehicle track data is to identify suspicious activities on the part of a few vehicles within a much larger context of everyday activity conducted by the majority of vehicles. The Analyst's Assistant is designed to help its user uncover and explain these suspicious activities, either forensically or in real time as the activities unfold. To this end, the Analyst's Assistant provides a natural language interface that allows its user to both control the focus of attention of the system and pose specific requests like "Do any of those cars make a long stop?", "Where do the possible perpetrators travel from?", "Do any of these cars meet?", "Where do the meeting cars go?" and "Do any of these cars follow each other?". In this manner, the user and system are able to elaborate a network of vehicles and places involved in a suspicious activity. Along the way, bindings may be assigned to names like "the detonation site", "the meeting site", "the possible perpetrators" and "the meeting cars", and the user can refer to these sets by name or refer to recently-mentioned entities as "those cars", "that time", and so forth. The Analyst's Assistant also provides a summarization capability, where the system recognizes all instances of several dozen types of events for a focus spatial region or focus set of vehicles over a particular interval of time, then filters the list of recognized events to remove redundant information.

Figure 1 illustrates the system's response to a user request in a sample analysis sequence. At this point in the interpretation, the user has identified a vehicle of interest, V518, and has focused the system's interpretation on a short interval of space and time in which that vehicle and 16 other vehicles appear. The user asks "Do any cars follow other cars?" and is presented with a list of four instances of following that have been identified by the system.



Impact Input

Do any cars follow other cars?

Submit

Impact Output

"Do any cars follow other cars?"

V260 follows V518 along Slide Rd from 10:17:58.605 to 10:18:39.124.

V518 follows V439 along Slide Rd from 10:18:14.362 to 10:18:36.310.

V260 follows V439 along Slide Rd from 10:18:14.925 to 10:18:36.310.

V260 follows V518 along FM 1730 from 10:18:33.496 to 10:20:00.160.

Figure 1: An interaction with the Analyst's Assistant.

The Analyst's Assistant is constructed using three components: the IMPACT reasoning system, the START information access system, and a Web-based GUI interface. These components are depicted in Figure 2.

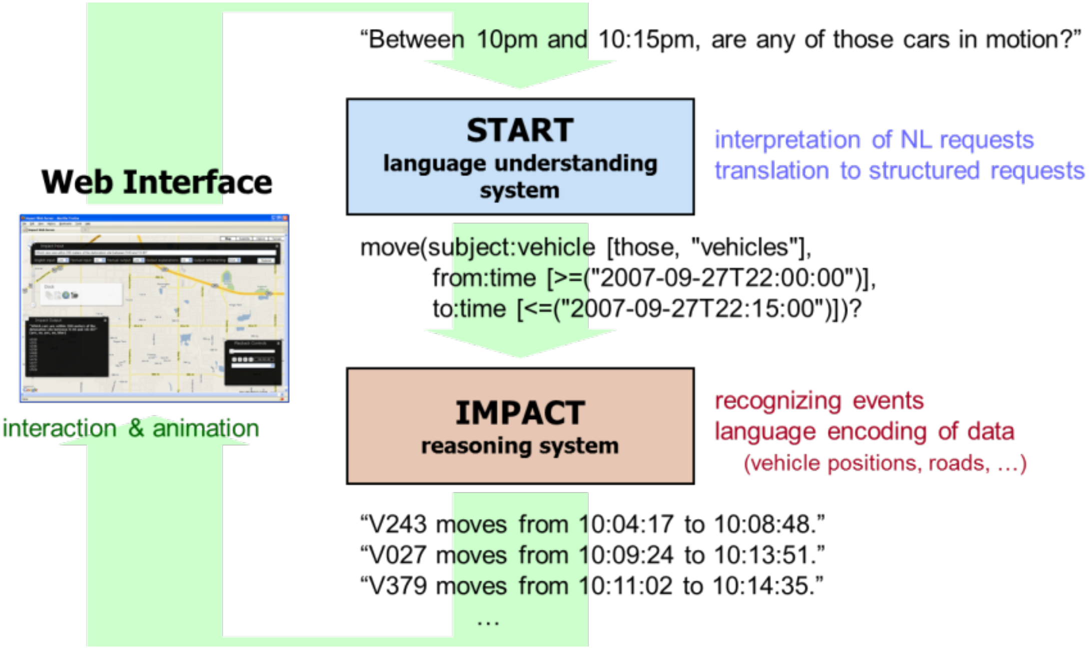

"Between 10pm and 10:15pm, are any of those cars in motion?"
START
language understanding system
interpretation of NL requests
translation to structured requests
Web Interface
move(subject:vehicle [those, "vehicles"],
from:time [>=("2007-09-27T22:00:00")],
to:time [<=("2007-09-27T22:15:00")])?
interaction & animation
IMPACT
reasoning system
recognizing events
language encoding of data
(vehicle positions, roads, …)
"V243 moves from 10:04:17 to 10:08:48."
"V027 moves from 10:09:24 to 10:13:51."
"V379 moves from 10:11:02 to 10:14:35."
…

Figure 2: Major components of the Analyst's Assistant.

The user interacts directly with the Web Interface. Through this interface, the user can view a static or animated map of vehicles on roads, enter natural language requests, view and select elements from natural language responses to requests, select positions in the map, and operate a number of GUI controls such as buttons and sliders. Some user requests are handled directly within the user interface, while other requests are forwarded to START and IMPACT. START interprets natural language requests and encodes them in a structured form for submission to IMPACT. IMPACT maintains a database of road coordinates plus timestamped vehicle positions and responds to a range of requests concerning this data, including requests about the occurrence of specific types of events. The user interface then presents the results of requests to START and IMPACT, either graphically or in language.

In the design of the Analyst's Assistant, an effort has been made to integrate language at a deep level within data analysis procedures. By this strategy, language is used not only to express the results of analysis—recognized instances of events—but also lower-level aspects of the scene: individual objects and abstractions involved in the activities, attributes of those objects and abstractions, and instantaneous values and changes in those attributes. In this manner, language serves as a framework for structuring the interpretation, aligning the interpretation with human conceptualization of scene entities and activities. This in turn can help facilitate system–human interaction in collaborative interpretation of the scene activity.

The Analyst's Assistant also addresses considerations of scalability. The dataset operated on by the system contains approximately 2 million timestamped vehicle positions and 25 thousand road segments, yet user interaction with the system must occur in real time for the human user. In addition, attention has been given to questions of generality of the approach, so that similar capabilities can potentially be demonstrated on other types of datasets. A subset of the techniques developed for Analyst's Assistant has recently been applied in an exploratory manner in the recognition of human-oriented events in ground-level surveillance video and also to the envisioning of actions postulated as steps in observed plans.

2.1 IMPACT

IMPACT is a system for recognizing events and reasoning about events, given low-level, timestamped information about an unfolding scene, plus models of what typically happens during different types of events, and, in some applications, externally-supplied assertions that particular events are occurring. The design of IMPACT builds on the presumption that language is critically involved in the process of reasoning about events, not only governing the top-level expression of event occurrences like “proceeding through an intersection” and “following another vehicle”, but also much lower-level assertions about states and changes in a scene, such as “ceasing to be in an intersection” or “speed exceeding a previous speed”. IMPACT consists of approximately 50,000 lines of Java code operating in conjunction with a PostgreSQL database with PostGIS geometric utilities, and it contains facilities for interacting through several external interfaces, maintaining information alternatively in database and main-memory formats, and processing data through a range of operations and matching capabilities.

In keeping with the position that language is critical to the process of reasoning about events, all information within the IMPACT system is depicted in language, as are all requests to the system and all responses issued by the system. This is facilitated by a representation called Möbius [Borchardt, 2014], which encodes simple language in parsed form, including both syntactic and semantic information. Möbius is intended to be a streamlined, intuitively-usable representation, incorporating just enough syntactic and semantic expressiveness to facilitate the representation of simple assertions. As an example, a Möbius assertion of the occurrence of an event might be as follows:

```
proceed(
  subject:vehicle "V253",
  through:intersection intersection(
    article: the, of: (:road "Indiana Ave", and:road "96th Street")),
  from:time "2013-10-14T11:47:35",
  to:time "2013-10-14T11:47:39").
```

whereas a lower-level assertion comparing speeds might be as follows:

```
exceed(
  subject:attribute speed(
    article: the, of:vehicle “V253”, at:time “2013-10-14T11:35:25”),
  object:attribute speed(
    article: the, of:vehicle “V253”, at:time “2013-10-14T11:35:24”)).
```

Within IMPACT, Möbius serves as a substrate for encoding a cognitively-motivated and language-based representation of “what happens” during particular types of events. This representation is called *transition space*, and was originally described in [Borchardt, 1992] and [Borchardt, 1994], with elaborations in [Katz *et al.*, 2005] and [Katz *et al.*, 2007]. A description of the encoding of transition space assertions within Möbius appears in [Borchardt, 2014].

In the transition space framework, “what happens” during events is described using a hierarchy of five quantities:

objects are concrete or abstract entities of importance in a situation, represented as parsable strings—e.g., the road “Chicago Ave”, the location “(-101.932745, 033.509693)”, the vehicle “V418”, or the time “2007-09-27T10:20:00.458”,

attributes are language-motivated properties, relationships, and other functions of the participants—e.g., “position” or “speed” of a vehicle, “distance” between two vehicles, a vehicle being “on” a road, or a vehicle being “at” an intersection,

states are instantaneous values of attributes—e.g., the speed of a vehicle at a particular time, or whether or not a vehicle is on a road at a particular time,

changes are comparisons of attribute values between time points—e.g., an “increase” or “decrease” in the distance between two vehicles, a “change” in a vehicle’s heading, or a vehicle “ceasing to be” at an intersection, and

events are collections of changes and states brought into focus for a particular analysis—e.g., “V507 meets with V509 from 09:58:12.319 to 10:08:19.531.”.

All five types of quantities draw their members from language. For instance, if, in language, we are able to express a relationship “in front of”, then in the representation, we also allow this relationship as an attribute. Objects include vehicles, locations and roads, for example, but also other entities that appear in language, such as intersections, time intervals, areas and events.

In addition to language-inspired constraints on the representation, there are also cognitively-inspired constraints. These are motivated by several considerations from the cognitive psychology literature (e.g., [Miller and Johnson-Laird, 1976]): time is depicted as a sequence of moments, attributes are taken to apply to one or two objects and have a range that is either qualitative or quantitative, attributes are typically compared for relative values, changes are assessed by looking at two moments, and events are characterized as complex patterns of change. Given these constraints, if we assume that each attribute has a “null” value plus one or more non-null values that can be compared qualitatively or quantitatively, we are presented with a set of ten varieties of change assessed between pairs of time points, as depicted in Figure 3.

appear	not appear	*boolean attributes (e.g., a vehicle being on a road or at an intersection)*
disappear	not disappear	
change	not change	*qualitative attributes (e.g., a vehicle's position or heading)*
increase	not increase	*quantitative attributes (e.g., a vehicle's speed, or distance between vehicles)*
decrease	not decrease	

Figure 3: Ten varieties of change.

Using these varieties of change, we construct *event models* that depict the temporal unfolding of particular types of events. For instance, the event of a vehicle turning left or right at an intersection might be described as follows:

- between "time 1" and "time 2", the vehicle enters the intersection (being "at" the intersection *appears,* while the vehicle's speed *does not disappear*),
- between "time 2" and "time 3", the vehicle may or may not come to a temporary stop (put simply: being at the intersection *does not disappear*),
- between "time 3" and "time 4", the vehicle executes its turning activity (being at the intersection *does not disappear*, speed and turning rate *do not disappear*, and turning direction *does not change*), and
- between "time 4" and "time 5", the vehicle exits the intersection (being at the intersection *disappears* while speed *does not disappear*).

Given a library of event models for common types of events in a domain, it is possible to recognize instances of those events through pattern matching between the models, on one hand, and, on the other hand, changes and states that have been identified in a scene. IMPACT performs this matching in such a way as to allow individual time intervals in the models (e.g., from "time 1" to "time 2" in the above model) to be matched to any number of sampling intervals of scene data, as needed to facilitate a complete match for the model. In this manner, an event model such as the one described above might be matched in one instance to a short sequence of data in the scene—say 3–5 seconds for a vehicle that turns without stopping at an intersection—and in another instance to a much longer sequence of data in the scene—20 seconds or more for a vehicle that stops before turning at an intersection.

2.2 START

START is a language-based information access system that has been in development for approximately 25 years [Katz, 1990; Katz, 1997; Katz *et al.*, 2006]. In its most general question-answering application, START is available as a public server at http://start.csail.mit.edu/. START answers questions in a range of domains including geography, arts and entertainment, history, and science, handling over a million requests per year from users around the world. In addition to the general-purpose public START server, several special-purpose servers have been created for specific topic areas. Also, several strategies pioneered by the START system contributed to the performance of IBM's Watson system in its 2011 *Jeopardy!* challenge.

In its traditional role, START accepts English questions, possibly containing grammatical or spelling errors, and offers responses that draw on information sources that include structured, semi-structured, and unstructured materials. Some of these materials are maintained locally and some are accessed remotely through the Internet. A particular emphasis of START is that of providing *high-precision* information access, such that the user may maintain a fair degree of confidence that a response, if returned by the system, is appropriate to the submitted question.

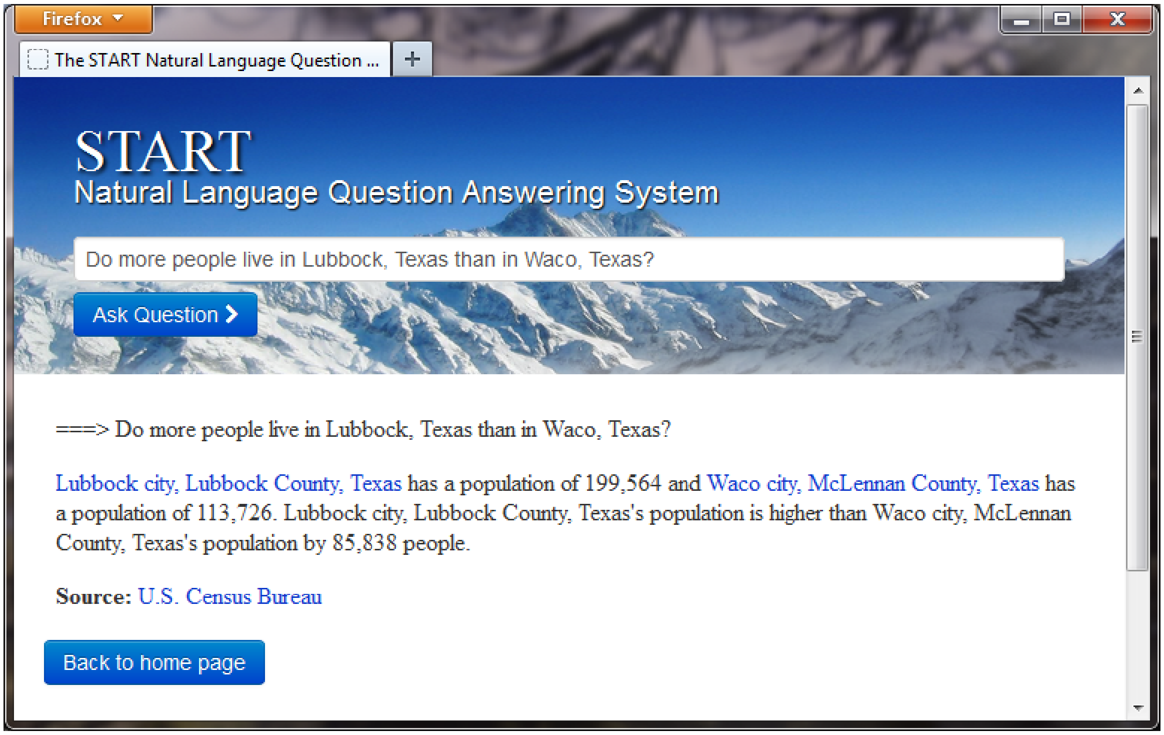
START uses tags called *natural language annotations* to index the information sources at its disposal [Katz, 1997; Katz *et al.*, 2006]. These tags are supplied by a knowledge engineer when START is configured for use. Natural language annotations are English phrases and sentences that may contain typed variables, as, for example

> *number* people live in *city*.

as an annotation for a table of city populations. Natural language annotations describe the content of the information sources they are associated with, using language as a representation of that content.

START precompiles its base of natural language annotations into *nested ternary expressions*, which are constituent–relation–constituent triples that capture much of the most salient information in a syntactic parse tree in a form well-suited to matching [Katz, 1990]. When a user submits an English question to START, the question is also translated into nested ternary expressions, and START uses the nested ternary expression representation as a basis for matching the submitted question to stored annotations, assisted by an array of techniques that include matching through synonymy and hyponymy, the application of structural transformation rules [Katz and Levin, 1988], a reference resolution mechanism, an external gazetteer for matching terms to typed variables, and a mechanism for decomposing complex questions syntactically into subquestions for independent processing [Katz *et al.*, 2005]. Once a question has been matched to one or more natural language annotations, the information sources associated with those annotations can be accessed to provide responses to the user.

Figure 4 illustrates an example of the public START server in use.



START
Natural Language Question Answering System

Do more people live in Lubbock, Texas than in Waco, Texas?

===> Do more people live in Lubbock, Texas than in Waco, Texas?

Lubbock city, Lubbock County, Texas has a population of 199,564 and Waco city, McLennan County, Texas has a population of 113,726. Lubbock city, Lubbock County, Texas's population is higher than Waco city, McLennan County, Texas's population by 85,838 people.

Source: U.S. Census Bureau

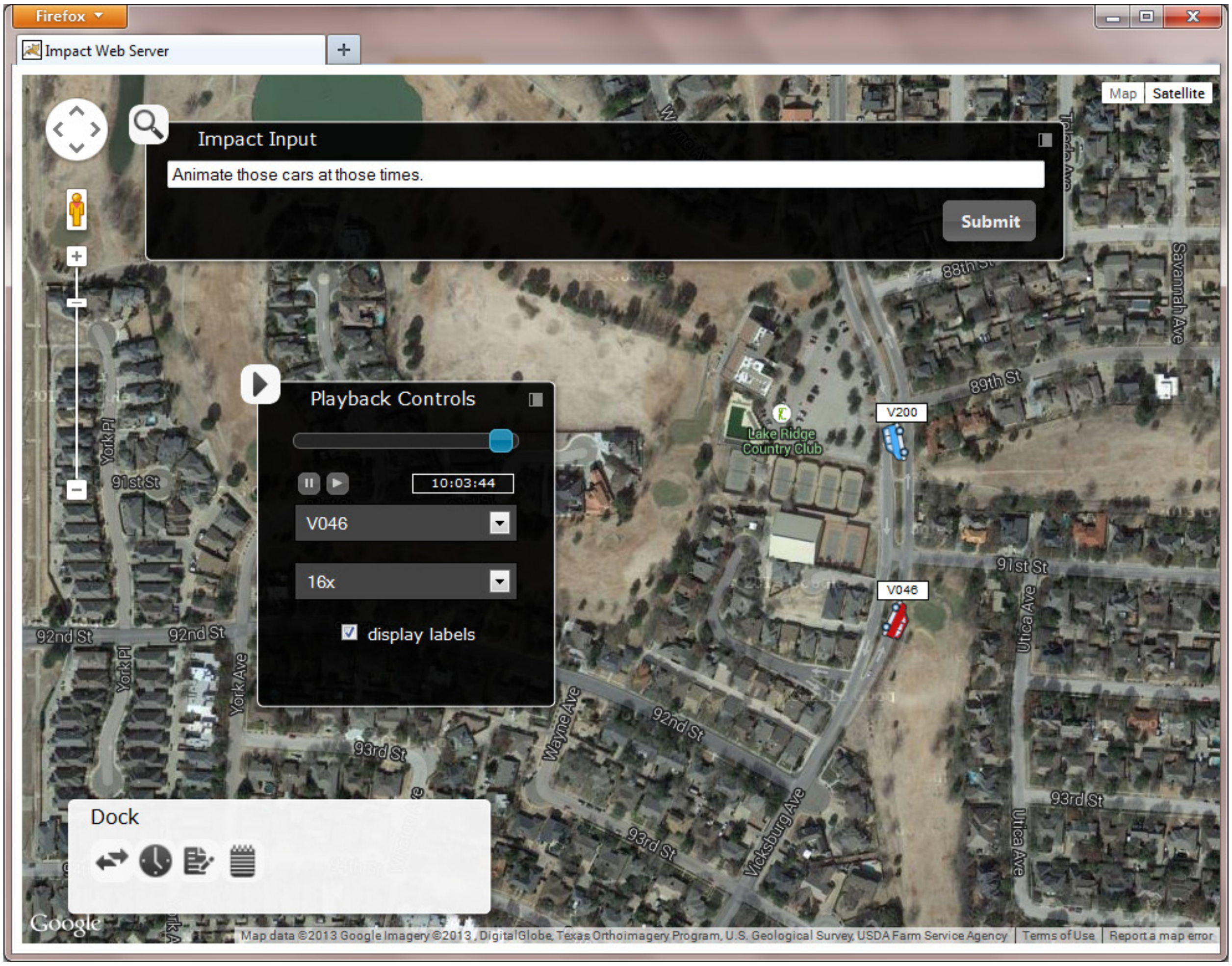
Figure 4: START handling a comparative question about cities' populations.

For structured materials such as databases, START can be configured to issue SQL requests in response to user queries, or it can be configured to issue symbolic requests in a format suitable for other, external information servers or systems. For the Analyst's Assistant, START has been configured to issue Möbius language requests to IMPACT. This linkage enables the user to pose a wide range of natural language requests to the system: requests for state information ("How fast is car 125 going at 12:03?"), searches ("Which vehicles were within 500 meters of the detonation site between 10:30 and 11:30?"), processing directives ("Please focus on V200 between 9:30 and 10:30." or "Animate those vehicles at those times."), and requests concerning events ("Do any cars travel north on Chicago Avenue between 11:15 and 11:20?").

2.3 Graphical User Interface

The graphical user interface for the Analyst's Assistant is a standalone Web application coded in Ruby on Rails. The interface utilizes Google Maps utilities to enable display, zoom and pan of underlying roads and geographical features, plus overlay of flags and icons for individual vehicles in displays and animations. Local operations can be carried out within the interface, including the running and re-running of animations and browsing within a history of request/response interactions, and AJAX technology is used to facilitate interactions between the interface and START plus IMPACT.

The graphical user interface for the Analyst's Assistant displays a toolbar from which several interaction panes can be accessed. These include panes for: input of natural language requests, viewing of natural language responses from the system, buffering of input sequences of natural language requests, viewing the history of requests and responses, running animations, and viewing user-created spatial graphs of interconnected positions. Figure 5 provides a screen shot of the graphical user interface in action, running an animation involving two vehicles.



Impact Input

Animate those cars at those times.

Playback Controls

10:03:44

V046

16x

☑ display labels

Figure 5: Sample interaction with the graphical user interface.

3. Key Ideas

Construction of the Analyst's Assistant has enabled us to advance the development of several emerging intelligent systems capabilities: event recognition, summarization of events, collaborative and scalable user–system interpretation of data, and multi-modal labeling and reference resolution. The following sections describe these efforts.

3.1 Event Recognition using Language-Based Representations

Event recognition from video and other sensor data has received a fair amount of attention over a number of years. Reviews of work in video event recognition can be found in [Buxton, 2003], [Hu *et al.*, 2004], [Turaga *et al.*, 2008] and [Lavee *et al.*, 2009]. The analyzed data range from close-up indoor surveillance video to long-standoff aerial video, and the recognized events span from what might be called "poses" to "movements" to larger "actions" and "scenarios". Techniques used in these systems are varied: predominantly quantitative, although some symbolic and logic-based approaches are also employed. Specifically for aerial video, less work has been pursued regarding event recognition, primarily because automatically extracted vehicle tracks are often incomplete. Some approaches are [Burns *et al.*, 2006], where more localized events are detected (e.g., vehicles traveling in a convoy), and [Porter *et al.*, 2009], where localized events (meetings) are detected, and this information is used to infer attributes over collections of tracks.

To our knowledge, very little work has been done in coupling language primitives to interpreted video and other sensor data, and using these language primitives as a basis for recognizing events and forming higher-level language descriptions. One example is the work of Siskind and his colleagues [Siskind, 2003; Narayanaswamy *et al.*, 2014], which deals with close-up observation of human activities. Design of the Analyst's Assistant follows from the belief that language is critical to the recognition of events, that it is language that provides an appropriate, common medium for abstracting events across domains, contexts and sensor types and provides a basis for system generation of justifications and explanations of conclusions.

Using language as a framework for representing what happens during the temporal unfolding of events has, in our experience, greatly simplified the task of creating the functionality needed to provide coordinated event recognition, summarization and question answering capabilities regarding timestamped sensor data. In the development of these capabilities, what otherwise might be viewed as an extremely complex learning task requiring a considerable amount of training data can instead be approached by leveraging human guidance—introspection and interview—in articulating the objects, attributes, states and changes involved in the unfolding of particular types of events.

For the Analyst's Assistant, the approach to developing the event recognition, summarization and question answering capabilities has been largely iterative, proceeding from an enumeration of objects and attributes relevant to our expected question answering functionality, to the use of these objects and attributes to create event models that describe particular types of events relevant to our expected question answering, event recognition and summarization functionality. Event models initially inspired by human knowledge have then been refined through a "test, critique and modify" cycle as these models have been applied to the event recognition task. In some cases, refinement of our event models for event recognition has led to the introduction of new attributes. Also, in some cases, considerations of performance in the task of summarization have led to further refinements in our event models. This iterative process has been greatly facilitated by the fact that all elements of the representation, being articulated in language, are transparent to human understanding and scrutiny in assessing the suitability of their use in particular, tested instances of question answering, event recognition, or summarization by the system as a whole.

Commencing with the lowest level of the representation, Figure 6 lists the object types enumerated within the Analyst's Assistant as a result of this iterative process. In this list, each object type appears with a sample object instance as encoded in the Möbius language.

distance	`"0250 m"`
elapsed time	`"P0000-00-00T00:02:00"`

event	`make(subject:vehicle "V233",`
	`     object: stop(article: a, adjective: long))`

heading	`"145 deg"`

intersection	`intersection(article: the,`
	`     of: (:road "83rd St", and:road "Clinton Ave"))`

named heading	`west`
position	`"(-101.933206, 033.517764)"`
road	`"Indiana Ave"`
speed	`"10.2 m/s"`
time	`"2007-09-27T09:50:38.176"`
turning direction	`left`
turning rate	`"23 deg/s"`
vehicle	`"V127"`

Figure 6: Object types and sample values within the Analyst's Assistant.

Figure 7 lists attributes used within the Analyst's Assistant, using a notation that abbreviates the underlying Möbius encoding. Timestamped values for the first attribute, vehicle position, are supplied to the application. Values for the remaining attributes are computed by the application using this information plus information regarding the locations of road segments in the geographical region of interest for the application.

the position of <vehicle>	(qualitative)
the speed of <vehicle>	(quantitative)
the heading of <vehicle>	(qualitative)
the named heading of <vehicle>	(qualitative)
the turning rate of <vehicle>	(quantitative)

the turning direction of <vehicle>	(qualitative)
<vehicle> being on <road>	(boolean)
<vehicle> being at <intersection>	(boolean)
the distance between <vehicle> and <vehicle>	(quantitative)
<vehicle> being in front of <vehicle>	(boolean)
<vehicle> being in back of <vehicle>	(boolean)
<vehicle> being next to <vehicle>	(boolean)
the occurrence of <event>	(boolean)
the elapsed time from <time> to <time>	(quantitative)
the distance between <position> and <position>	(quantitative)
the turning direction between <heading> and <heading>	(qualitative)

Figure 7: Attributes within the Analyst's Assistant.

In turn, states and changes involving the object types and attributes of Figures 6 and 7 are then used to construct event models for 35 types of events within the Analyst's Assistant, as listed in Figure 8, again abbreviating the underlying Möbius encoding.

<vehicle> accelerates.
<vehicle> accelerates to <speed>.
<vehicle> decelerates.
<vehicle> hesitates.
<vehicle> makes a U-turn.
<vehicle> moves.
<vehicle> does not move.
<vehicle> does not turn.
<vehicle> travels <named heading>.
<vehicle> travels <named heading> on <road>.
<vehicle> travels on <road>.
<vehicle> turns sharply.
<vehicle> turns <turning direction>.
<vehicle> turns.
<vehicle> enters <intersection>.
<vehicle> exits <intersection>.
<vehicle> makes a U-turn at <intersection>.
<vehicle> proceeds through <intersection>.
<vehicle> stops at <intersection>.
<vehicle> turns <turning direction> at <intersection>.
<vehicle> makes a long stop.
<vehicle> makes a medium-length stop.
<vehicle> makes a short stop.
<vehicle> makes a stop at <position>.
<vehicle> makes a stop for <duration>.
<vehicle> makes a stop on <road>.
<vehicle> travels from the area within <distance> of <position> to the area within <distance> of <position>. *(4 models)*
<vehicle> travels from <position> to <position>.
<vehicle> approaches <vehicle>.
<vehicle> follows <vehicle> along <road>.
<vehicle> meets with <vehicle>.
<vehicle> pulls away from <vehicle>.

Figure 8: Event models within the Analyst's Assistant.

Complete descriptions of the 35 event models appear in Appendix A. As an example, Figure 9 depicts an event model for one vehicle approaching a second vehicle, using a graphical notation that organizes the presentation of changes, states and other assertions within the event model.

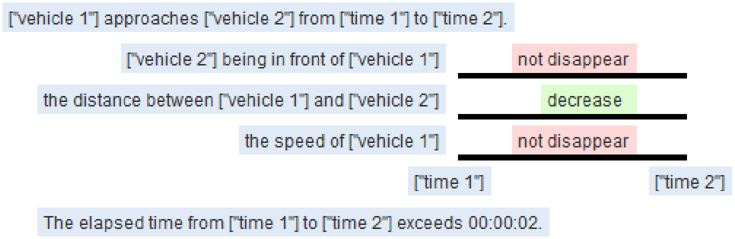

["vehicle 1"] approaches ["vehicle 2"] from ["time 1"] to ["time 2"].
["vehicle 2"] being in front of ["vehicle 1"]
not disappear
the distance between ["vehicle 1"] and ["vehicle 2"]
decrease
the speed of ["vehicle 1"]
not disappear
["time 1"]
["time 2"]
The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.

Figure 9: Event model for a vehicle approaching a vehicle.

In this event model, vehicle 1 maintains speed above a minimum threshold over an interval of at least 2 seconds, while vehicle 2 remains in front of vehicle 1 and the distance between the two vehicles decreases.

IMPACT matches event models to scene information using a strategy called *core–periphery matching*, which allows the system to stretch the intervals of event models (or "pattern events") in time, as necessary, to match the timing of observed changes in a scene. The core–periphery matching algorithm proceeds in a hierarchical manner. Given instructions to identify instances of a particular event model, the algorithm first identifies matches for individual changes in the first interval (or "transition") of the event model and combines these change matches to enumerate possible matches for that entire transition. Each subsequent transition in the event model is then matched in a similar manner, followed by an additional step in which candidate delimiting times are established between matches for the preceding transition and matches for the most recent transition.

Figure 10 illustrates the match of a single pattern change, an "increase", in an event model. Suppose this pattern change is being matched to a sequence of scene changes as indicated below this change. The "increase" can match a number of subsequences of this sequence of scene changes, within the region containing changes of type "increase" and "not change". Note that not all subsequences of this region will match the pattern change, however. In particular, the subsequences containing only changes of type "not change" do not imply an increase by themselves, but must be extended to include at least one scene change of type "increase".

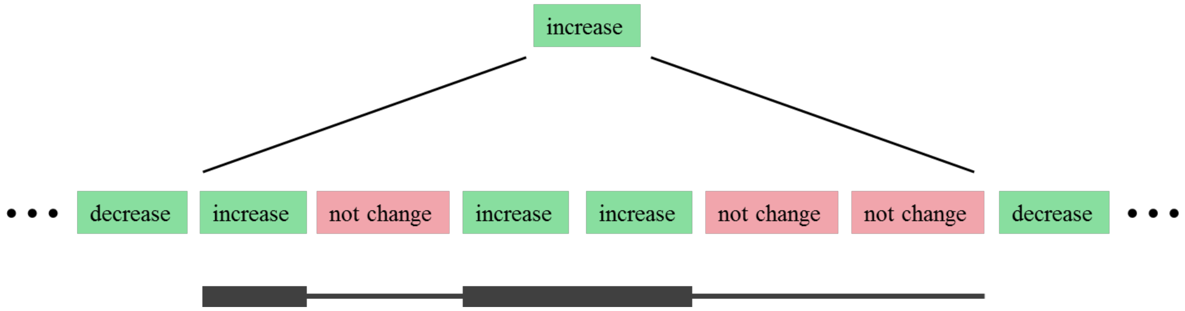


Figure 10: Matching a single pattern change to scene data.

A compact description of how a pattern change can match various subsequences of a sequence of scene changes is to first identify "cores" of a match: subsequences for which, in turn, every one of their subsequences is also a match to the pattern change. Match "cores" for the example in Figure 10 are illustrated as thick black bars under the scene changes. Next, surrounding each core, we can identify a "periphery": an extended subsequence for which any of its subsequences will be a match for the original pattern change only if at least some portion of a "core" is also present in that subsequence. Match "peripheries" for the example in Figure 10 are illustrated as thin black bars deemed to extend as well through the thick black bars.

The next step is to consider a column of a pattern event. Figure 11 presents a possible 3-element column of an event model, having an "increase" in some attribute of a first object or pair of objects (not shown), a "decrease" in some other attribute of some other objects, and a "disappear" in some other attribute of some other objects. We assume the matcher has matched these associated attributes and objects appropriately, maintaining consistent bindings for object variables that appear, and here we are only concerned with how to match the changes. The right side of Figure 11 contains a schematic illustration of some sample core–periphery matches for the three depicted changes. To form a composite match for the entire column of the event model, we need to find an earliest and latest time point for an interval in which all three changes can be matched. This is illustrated by blue vertical bars. We also want to calculate how much "give" there might be in the ending time point for the column match—how much earlier we might place this time point and still succeed in matching all of the changes. This is illustrated by a dotted blue line. If the ending time point for the column match is placed before this time, this will result in a failure for the second change match, involving the change "decrease".

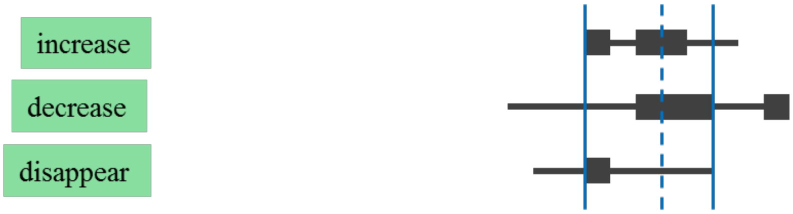


Figure 11: Combining change matches into a transition match.

Finally, suppose we have a second column of changes in the pattern event model—a second "transition"—specifying an "increase", a "not disappear" and an "appear" in the same three attributes. Suppose further that the core–periphery matches for those changes are as depicted by brown bars in Figure 12, using bindings that are consistent with the bindings applied in matching the first transition of the event model. The earliest and latest time points for a time interval supporting a match for the second transition are indicated by orange vertical bars. At this point, we would like to determine a suitable delimiting time to separate the match for the first transition from the match for the second transition of the event model. In this case we can choose the second transition's earliest match time as the delimiting time between the transitions, as this preserves all of the underlying change matches.

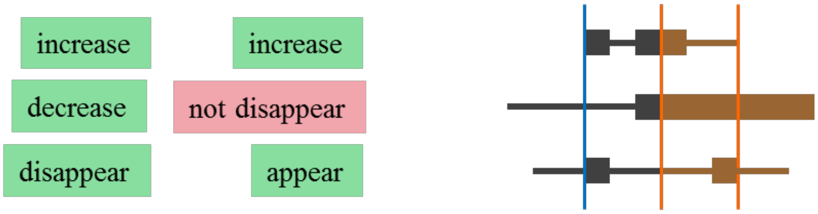


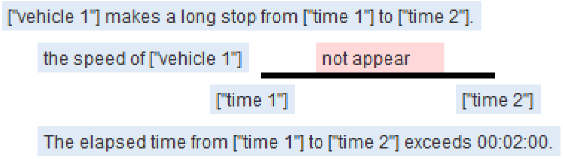
Figure 12: Combining transition matches.

This process repeats, progressing through the intervals from left to right in the event model. One useful heuristic we have found is to match all columns with "definite" changes (of the 10 possible changes, those that assure us that something has indeed changed—"appear", "disappear", "change", "increase" and "decrease") as tightly as possible, while matching columns without any definite changes as loosely as possible. This has the effect of generating tighter beginning and ending times for events which begin or end with definite changes, and it allows us to more readily isolate definite changes that appear within events.

The matcher is incomplete, as there are many additional matches to be found that have only slight differences in delimiting times. The matcher is sound in that all matches do faithfully match pattern changes to appropriate sequences of scene changes. The matcher is efficient in practice, as it abandons non-matches quickly and spends most of its time elaborating the details of ultimately successful matches.

Event recognition in the Analyst’s Assistant proceeds from largely top-down influences, starting with the direction to identify possible instances of a particular event type, say a “U-turn” event, which leads to the direction to identify possible matches for the first transition of that event, which in turn leads to the direction to identify possible matches for individual changes specified within the first transition of that event. Once appropriate change matches have been identified, they are assembled into appropriate transition matches and event matches in accordance with the structure of the targeted event models.

One aspect of the event recognition process that occurs in a strictly bottom-up manner concerns the “reification” of events—recognition of one particular type of event in a first cycle, leading to the system recording new state- and change-level information about the identified event occurrences, so that the matcher may be deployed in a second cycle to recognize additional “events about events”. Within the Analyst’s Assistant, this sort of approach was required to recognize instances of a vehicle traveling from an origin to a destination. Intuitively, traveling from an origin to a destination can be thought of making a fairly long stop at the origin, then proceeding for an interval of time while not making any long stops (possibly stopping briefly at a stop sign or traffic signal, for example), then making a fairly long stop at the destination. The Analyst’s Assistant recognizes instances of this sort of traveling by first recognizing instances of making long stops, using the event model illustrated in Figure 13.



["vehicle 1"] makes a long stop from ["time 1"] to ["time 2"].

the speed of ["vehicle 1"]

not appear

["time 1"]

["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:02:00.

Figure 13: Event model for making a long stop.

Once the system has identified instances of vehicles making long stops, it then creates a number of new “objects” to represent those events, objects like the following, in Möbius:

```
make(subject:vehicle "V345", object: stop(article: a, adjective: long)) .
```

The system then records information about the occurrence of the long stops by employing a new attribute “the occurrence of <event>”, or, in Möbius,

```
occurrence(article: the, of:event [])
```

applied to the new event-objects. This attribute specifies whether the occurrence of a particular event is, at any particular time point, present or absent. Having recorded the event occurrence information at the level of states and changes, the system is then in the position to recognize instances of traveling from an origin to a destination, using the event model illustrates in Figure 14.

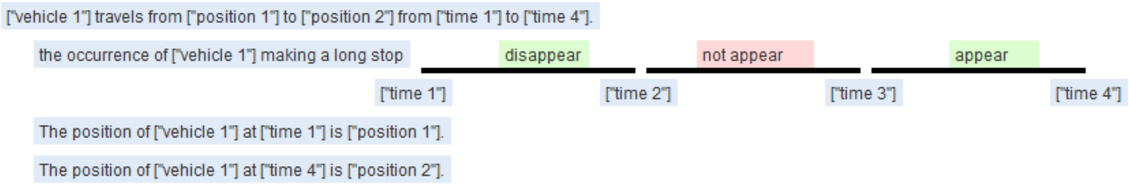

["vehicle 1"] travels from ["position 1"] to ["position 2"] from ["time 1"] to ["time 4"].
the occurrence of ["vehicle 1"] making a long stop
disappear
not appear
appear
["time 1"]
["time 2"]
["time 3"]
["time 4"]
The position of ["vehicle 1"] at ["time 1"] is ["position 1"].
The position of ["vehicle 1"] at ["time 4"] is ["position 2"].

Figure 14: Event model for traveling from an origin to a destination.

A match for this event model will begin with a vehicle culminating a first long stop, it will then proceed through an interval in which the vehicle is not making any long stops, and it will end with the vehicle initiating a second long stop.

3.2 Summarization by Covering Scene Information

The event recognition apparatus of the Analyst's Assistant enables the system either to identify instances of a single type of event—e.g., in response to a targeted query like "Do any vehicles make a long stop?"—or to identify instances of many types of events at once. When humans observe a scene and describe its activity, if instances of many types of events might be found in the scene, humans will nonetheless refrain from exhaustively listing all identified event instances in forming a description of the scene. Rather, they will offer a selection of event instances that summarize the observed situation according to accepted conversational conventions such as outlined by Grice [1975]:

The maxim of quantity: Provide as much information as is required for the current purposes of the exchange, but do not provide more information than is required.

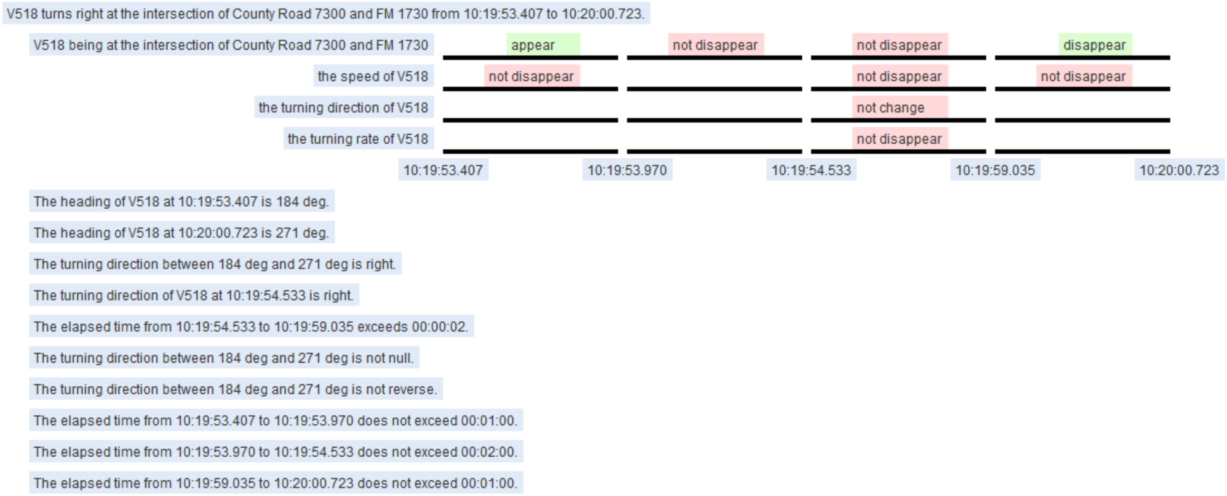
The maxim of quality: Try to make your contribution one that is true.

The maxim of relation: Be relevant.

The maxim of manner: Be perspicuous. (Avoid obscurity of expression, avoid ambiguity, be brief and be orderly.)

Implicit in both Grice's maxim of quantity and maxim of manner is the notion of avoiding redundancy. Regarding event recognition, it is of little use to mention event instances that provide no additional information about an unfolding scene relative to other event instances that describe the same phenomena. Within the Analyst's Assistant, we have implemented a summarization algorithm that follows this strategy. Each recognized event instance is said to "cover" a subset of the scene information corresponding to a set of assertions at the lowest level of the transition space representation—assertions that various quantities at various times "equal", "do not equal", "exceed" or "do not exceed" various other quantities at various times. To form a summary of observed activity, we exclude from the description particular event instances that fail to uniquely cover any scene information relative to the scene information covered by other event instances retained in the description. When considering event instances for exclusion, we start with events that have smaller coverage of scene information and proceed to events that have larger coverage of scene information.

As an example, in one interaction with the Analyst's Assistant, an instance of "turning right at an intersection" was recognized, as depicted in Figure 15.



V518 turns right at the intersection of County Road 7300 and FM 1730 from 10:19:53.407 to 10:20:00.723.

V518 being at the intersection of County Road 7300 and FM 1730	appear	not disappear	not disappear	disappear
the speed of V518	not disappear		not disappear	not disappear
the turning direction of V518			not change	
the turning rate of V518			not disappear	

10:19:53.407 10:19:53.970 10:19:54.533 10:19:59.035 10:20:00.723

The heading of V518 at 10:19:53.407 is 184 deg.

The heading of V518 at 10:20:00.723 is 271 deg.

The turning direction between 184 deg and 271 deg is right.

The turning direction of V518 at 10:19:54.533 is right.

The elapsed time from 10:19:54.533 to 10:19:59.035 exceeds 00:00:02.

The turning direction between 184 deg and 271 deg is not null.

The turning direction between 184 deg and 271 deg is not reverse.

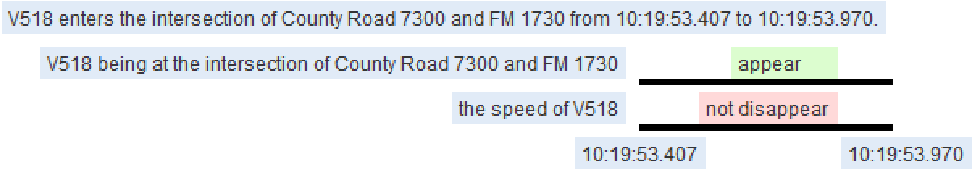
The elapsed time from 10:19:53.407 to 10:19:53.970 does not exceed 00:01:00.

The elapsed time from 10:19:53.970 to 10:19:54.533 does not exceed 00:02:00.

The elapsed time from 10:19:59.035 to 10:20:00.723 does not exceed 00:01:00.

Figure 15: Event instance of "turning right at an intersection".

Also, an instance of "entering an intersection" was recognized, as depicted in Figure 16.


V518 enters the intersection of County Road 7300 and FM 1730 from 10:19:53.407 to 10:19:53.970.

V518 being at the intersection of County Road 7300 and FM 1730
appear
the speed of V518
not disappear
10:19:53.407
10:19:53.970

Figure 16: Event instance of "entering an intersection".

Since the instance of entering the intersection does not cover any scene information that is not also covered by the instance of turning right at the intersection, the summarization algorithm excludes mention of the instance of entering the intersection in its generated description of the scene activity.

A complementary summarization strategy implemented within the Analyst's Assistant is to exclude particular attributes, or particular types of states or changes of particular attributes, from the calculations of scene information coverage for the purpose of retaining or excluding events from the description. This enables the generated summary to focus on events involving the most significant attributes, states and changes. By this strategy, the following are excluded from coverage calculations within the Analyst's Assistant:

- all constraints on the durations of time intervals within event models
- information concerning instantaneous values and across-time comparisons of speed for vehicles,
- information concerning instantaneous values and across-time comparisons of turning rate, plus information about the absence of a turning rate for a vehicle,
- information about the presence or absence, particular values, and across-time comparisons of distance between vehicles,
- information about the presence or absence, particular values, and across-time comparisons of vehicles being in front of or in back of other vehicles.

Using the two above-described strategies in tandem, the summarization capability of the Analyst's Assistant will typically eliminate approximately 2/3 of the recognized event instances from inclusion in its summary of an analyzed activity. For example, in one interaction for which the Analyst's Assistant has interpreted the activity within 500 meters of a specified position over an interval of 2 minutes, the full set of recognized event instances for vehicle V030 is as depicted in Figure 17.

V030 moves from 10:08:38.102 to 10:08:52.171.
V030 travels on County Road 7300 from 10:08:38.102 to 10:08:52.171.

V030 does not turn from 10:08:38.102 to 10:08:48.231.
V030 travels west from 10:08:38.102 to 10:08:48.231.
V030 travels west on County Road 7300 from 10:08:38.102 to 10:08:48.231.
V030 decelerates from 10:08:43.729 to 10:08:52.171.
V030 approaches V201 from 10:08:44.292 to 10:08:51.045.
V030 accelerates from 10:08:53.296 to 10:09:03.426.
V030 accelerates to 16.9 m/s from 10:08:53.296 to 10:09:03.426.
V030 proceeds through the intersection of County Road 1800 and County Road 7300 from 10:08:54.422 to 10:09:00.612.
V030 enters the intersection of County Road 1800 and County Road 7300 from 10:08:54.422 to 10:08:54.984.
V030 does not turn from 10:08:54.984 to 10:09:25.373.
V030 travels west from 10:08:54.984 to 10:09:25.373.
V030 travels west on County Road 7300 from 10:08:54.984 to 10:09:25.373.
V030 exits the intersection of County Road 1800 and County Road 7300 from 10:09:00.049 to 10:09:00.612.
V030 accelerates from 10:09:05.114 to 10:09:10.742.
V030 accelerates to 18.8 m/s from 10:09:05.114 to 10:09:10.742.
V030 decelerates from 10:09:10.742 to 10:09:14.681.
V030 accelerates from 10:09:14.681 to 10:09:17.495.
V030 accelerates to 18.2 m/s from 10:09:14.681 to 10:09:17.495.
V030 decelerates from 10:09:17.495 to 10:09:27.061.
V030 makes a U-turn from 10:09:25.936 to 10:09:32.126.
V030 turns left from 10:09:25.936 to 10:09:32.126.
V030 turns from 10:09:25.936 to 10:09:32.126.
V030 decelerates from 10:09:28.750 to 10:09:31.001.
V030 accelerates from 10:09:31.001 to 10:09:43.381.
V030 accelerates to 22.5 m/s from 10:09:31.001 to 10:09:43.381.
V030 does not turn from 10:09:36.628 to 10:09:50.134.
V030 travels east from 10:09:36.628 to 10:09:50.134.
V030 travels east on County Road 7300 from 10:09:36.628 to 10:09:50.134.
V030 decelerates from 10:09:43.944 to 10:09:46.195.
V030 accelerates from 10:09:46.195 to 10:09:48.446.
V030 accelerates to 22.2 m/s from 10:09:46.195 to 10:09:48.446.
V030 enters the intersection of County Road 1800 and County Road 7300 from 10:09:55.199 to 10:09:55.762.

Figure 17: All recognized event instances involving V030 in one example.

Following application of the summarization strategies, the reduced description provided by the system is as depicted in Figure 18. In particular, it should be noted that the inclusion of event instances involving traveling in a particular direction on a road has forced the exclusion of events involving traveling in that direction or simply moving, that inclusion of an event of "proceeding through an intersection" has forced the exclusion of events involving entering and exiting the intersection, and inclusion of an event describing a U-turn has forced the exclusion of an event involving the vehicle turning left.

V030 travels on County Road 7300 from 10:08:38.102 to 10:08:52.171.
V030 travels west on County Road 7300 from 10:08:38.102 to 10:08:48.231.
V030 accelerates to 16.9 m/s from 10:08:53.296 to 10:09:03.426.
V030 proceeds through the intersection of County Road 1800 and County Road 7300 from 10:08:54.422 to 10:09:00.612.
V030 travels west on County Road 7300 from 10:08:54.984 to 10:09:25.373.
V030 makes a U-turn from 10:09:25.936 to 10:09:32.126.
V030 accelerates to 22.5 m/s from 10:09:31.001 to 10:09:43.381.
V030 travels east on County Road 7300 from 10:09:36.628 to 10:09:50.134.
V030 enters the intersection of County Road 1800 and County Road 7300 from 10:09:55.199 to 10:09:55.762.

Figure 18: Summary of event instances involving V030 in the example.

3.3 Collaborative and Scalable User–System Interpretation of Data

Humans and software systems have distinct capabilities related to the analysis of sensor data, and it is important to leverage these relative capabilities in order to achieve high quality, time efficiency, and scalability in the interpretation of data. The Analyst's Assistant has been designed to partition analysis subtasks between the human and software system, with the human generally performing higher-level interpretive tasks, and the system performing lower-level interpretive tasks. Joint human–system interpretation of data is then facilitated by encoding all system data and partial results using language-based representations and by incorporating a robust natural language interface between the user and system, enabling the user to enter requests in language and view system responses rendered in language.

Particular tasks carried out by the human user of the Analyst's Assistant are:

- controlling the focus of processing for the analysis to cover particular spatial regions of the map or particular sets of vehicles, considered over particular intervals of time,
- steering the analysis between broad consideration of many event types to more limited consideration of specific types of events or drill-down to underlying information about states, and
- interpreting recognized events as elements of larger plans carried out by scene actors.

In return, the system carries out the following tasks:

- recognizing individual instances of events, based on scene information,

- forming summaries of event activity for a focus region or set of vehicles, and
- responding to a range of natural language requests placed by the user.

Allowing the user to control the focus of processing serves as one element of a larger strategy of timing the execution of various analysis tasks for efficiency; tasks whose results are likely to support multiple follow-on tasks are executed earlier, while tasks unlikely or unable to support follow-on tasks are executed as late as possible, only when needed. The Analyst's Assistant executes various analysis tasks at three distinct times:

- in a preprocessing step, vehicle position information is enhanced for use in multiple analysis sessions by multiple users,
- in a user-directed "focus" step during an analysis session, state and change information is calculated for a subset of the data, preparing the system for recognition of events on that data subset, and
- in an on-demand processing step during an analysis session, specific user requests are handled, using information prepared during the preprocessing and focus steps.

During the preprocessing step, accomplished before the first user analysis session, the Analyst's Assistant takes timestamped vehicle position information and road network data from its supplied dataset and calculates instantaneous speeds, headings, turning rates and turning directions for vehicles, plus instances of vehicles being on various roads. These state calculations are then integrated within the dataset operated upon by the system. The preprocessing step is designed to incorporate less computationally-expensive calculations that are broadly needed for follow-on calculations during data analysis and that could potentially be carried out in real time in a pipelined manner during data capture.

In contrast, the focus step and on-demand processing steps are designed to incorporate more computationally-expensive calculations that are only needed if they are of particular interest to the human user. The focus step is triggered by user requests to analyze subsets of the data covering particular spatial regions or particular sets of vehicles over specified intervals of time. This step prepares the system to perform on-demand event recognition on the focus region or set of vehicles.

The following are examples of natural language requests that trigger the focusing step. The START system, acting as the natural language interpretation component of the Analyst's Assistant, also accepts many natural language variants of each type of request:

> Focus on the area within 500 meters of 101.914764 W 33.511973 N between 10:01 and 10:04.
> Analyze the area within 1000 meters of site 21-1 from 10:17 to 10:22.
> Interpret the activity of V200 from 9:34 to 9:37.
> Focus on the suspicious vehicles between 10:28 and 10:35.

During the focus step, the system calculates instances of vehicles being “at” various intersections, spatial relationships (distance, being “in front of”, “in back of” or “next to”) for pairs of vehicles that are sufficiently close to one another, and relative changes for all attributes for which state information has been calculated. Also during the focus step, all relevant state and change information is assembled into a main-memory cache for quick matching to event models.

During the on-demand processing step, the user may pose targeted natural language requests to the system. Requests concerning the occurrence of specific types of events are processed relative to the current focus area or set of vehicles. Examples of these types of requests are:

> Are any cars following V518?
> Where do those cars go?
> Do any of these cars meet?

In response to these types of requests, the system identifies and lists occurrences of the specified types of events, restricting its search to the time interval and spatial region or set of vehicles specified by the user during the focus step.

During the on-demand processing step, the user also exerts control over another aspect of the data interpretation: breadth of the analysis, ranging from the consideration of many types of events to more limited consideration of individual types of events or even lower-level inspection of underlying states for vehicles involved in the activity. Consideration of many types of events occurs when the user submits a request to view all events identified by the system for the current focus region or set of vehicles, or if the user requests a summary of these events, or, optionally, if the user restricts the listing of all events or a summary of events to only those pertaining to a specified set of vehicles. The following are examples of these types of requests:

> Give me a complete list of the events.
> What are the main activities?
> What are all of the events that occur for V030?
> Summarize events for V507.

Requests for specific types of events can be as illustrated previously, or in the following examples:

> Which cars travel through the intersection of 90th St and Chicago Ave?
> Does V477 make a short stop?
> When does V201 turn left?

Requests for state information can be as in the following examples:

> Which cars are on 93rd Street between 9:40 and 9:41?
> What is the heading of V258 at 9:44:24?
> Where is the intersection of 82nd Street and Memphis Ave?
> Is V047's speed ever greater than 35 mph between 9:50 and 9:55?

Using a combination of requests that concern many types of events, requests that concern individual types of events, and requests that concern underlying states, the user can gain an understanding of significant activities taking place within a focus region or for a focus set of vehicles, plus underlying data support for those interpretations. Progressing to new focus regions, sets of vehicles and times, the user can gain an understanding of larger activities of interest. Typically, this might concern the execution of various plans on the part of adversaries. Appendix B illustrates use of the Analyst’s Assistant in uncovering an enacted IED emplacement and detonation sequence of activity captured within its supplied dataset.

To support the user in the interpretation of larger sequences of activity, the Analyst’s Assistant provides a record-keeping mechanism whereby the user can designate significant locations, vehicles and times, plus associations between these entities. This record is referred to as the “activity graph”. The user can specify significant entities using requests such as the following:

> Include V321 in the activity graph.
> Please put that location in the activity graph.

Examples of requests that designate associations between entities are:

> Associate the stationary vehicle with site 1.
> Link the stop site with 10:20 AM in the activity graph.

Once the user has specified a set of significant entities and associations between those entities, the user can query the system for elements of this record as illustrated in the following examples:

> What vehicles are connected with site 1?
> What locations are linked to V101?
> What times are associated with the stop site?

Also, the user can inspect the record of significant entities and associations graphically. This capability is illustrated in Appendix B, in the context of the IED emplacement interpretation example given there.

In response to the tasks carried out within an analysis session by the user of the Analyst's Assistant, the system carries out a complementary range of tasks, including major tasks such as performing event recognition, summarizing event sequences, shifting the focus of the system, and maintaining the activity graph, but also including supporting tasks such as opening and closing datasets, assigning and updating temporary names to scene entities and sets of scene entities, retrieval and search over state information for the dataset, and displaying animations of scene activity. Coordination of user and system tasks is facilitated by the natural language interface of the system, as well as the displays, buttons and menu commands of the system's graphical user interface.

3.4 Multi-Modal Labeling and Reference Resolution

The Analyst's Assistant is designed to employ language-motivated representations throughout its operation. This design strategy benefits user–system collaboration by ensuring that data entities operated on by the system can be easily articulated in language output generated by the system, and it also simplifies the task of interpreting the user's natural language input in terms of system data entities and operations on those entities.

One mechanism that contributes significantly to the system's ability to interpret natural language input from the user is a capability for flexible resolution of natural language and graphical references to data entities. Natural language references can be proper names or literal values like "Indiana Avenue" or "2:10 PM", they can be indirect references to recently-concerned elements using phrases such as "those roads" or "that time", or they can be user-defined names like "the suspicious vehicles" or "the meeting spot". Graphical references can be mouse clicks on the map, or they can be mouse selections of elements listed in natural language output of the system.

The most specific varieties of requests to the Analyst's Assistant contain proper names or literal values. An example of such a request is

> Is V149 on Chicago Avenue between 10:20 AM and 10:25 AM?

The START system, acting as the natural language front end of the Analyst's Assistant, interprets proper names and literal values directly, using pre-defined strategies for translating these natural language references to data entities. Many data entities can be expressed in multiple ways in natural language. For example, a vehicle might be "V248", "vehicle 248" or "car 248", a road might be "Chicago Avenue" or "Chicago Ave", and a time might be "10:20 AM" or "10:20:00.000".

In place of proper names and literals, the user may leave one or more elements of a request unspecified or specifically targeted by the request, as in the following examples.

Is V149 on Chicago Avenue at some time?
Which vehicles are on Chicago Ave between 10:20 AM and 10:25 AM?

In these examples, the phrases “some time” and “which vehicles” leave particular elements unspecified, to be filled by data values. Whether an element has been explicitly specified, as by a proper name or literal, or left unspecified, as by the references illustrated above, the Analyst’s Assistant will maintain a record of the sets of data entities that become associated with these references. This record is maintained within the IMPACT reasoning system, and it enables the user to employ references like “those roads” and “that vehicle” in subsequent requests. As an example, following the request “Is V149 on Chicago Avenue at some time?”, the user might submit a request

Is V149 driving faster than 40 miles per hour at those times?

In this instance, the Analyst’s Assistant will interpret “those times” to be the most recent times concerned in a request–response cycle, and as a result, this request will be interpreted as a request to determine if V149 has traveled faster than 40 mph at those particular times when it was on Chicago Avenue.

The IMPACT reasoning system retains sets of most-recently-concerned values for vehicles, roads, positions and times. An additional mechanism ensures that these values are re-used in the proper context, however. If the request “Which vehicles are on Chicago Ave between 10:20 AM and 10:25 AM?” is followed by the request

Please animate those vehicles at those times.

then the system must ensure that each vehicle is animated only at those times for which that particular vehicle was on Chicago Avenue, rather than at all of the times for which any vehicle was found to be on Chicago Avenue. IMPACT records these correspondences by maintaining a partially-resolved constraint system that includes combinations of values—particular vehicles associated with particular times, particular locations associated with particular vehicles and particular times, and so forth. This partially-resolved constraint system is updated with each subsequent request that refers to previous values, possibly becoming more constrained with fewer value combinations retained. In this manner, the system can process a sequence of requests such as the following:

Which vehicles are on Chicago Ave between 10:20 AM and 10:25 AM?
Are those vehicles traveling faster than 40 miles per hour at any of those times?
Please animate those vehicles at those times.

In processing this sequence, the system first identifies vehicles on Chicago Avenue during the specified interval of time. It then records a set of vehicle–time pairs: vehicles found to be on Chicago Avenue paired with times at which those vehicles were on Chicago Avenue. Next, the system constrains this list of pairs to include only those vehicle–time combinations for which the vehicle was traveling faster than 40 mph. Finally, the system animates the remaining, reduced set of vehicles and times.

The Analyst's Assistant also contains a mechanism for "assigning" sets of values to particular names. If the user enters a request such as

> Remember those positions as the "possible meeting sites".

then the IMPACT system will create a separate record of the association between the name "possible meeting sites" and the positions most recently considered within preceding requests. The user may then modify the set of values assigned to that name with requests like

> Include site 1 in the possible meeting sites.
> Retain only those locations in the possible meeting sites.
> Please remove the monitored positions from the possible meeting sites.

where "site 1" and "the monitored positions" are other, used-defined sets of positions. However, the associations of values and names such as these remain otherwise fixed, in line with common language usage. As an example, if the user has declared a name "possible lookout vehicles" to correspond to a set of vehicles

> Please store those cars as the "possible lookout vehicles".

and then the user submits two follow-on requests in sequence

> Which of the possible lookout vehicles are within 100 meters of site 10 between 10:00 AM and 10:10 AM?
> Which of the possible lookout vehicles are within 100 meters of site 15 between 10:30 AM and 10:40 AM?

the system will process each of these two requests independently, each time starting with the set of vehicles assigned to the name "possible lookout vehicles".

In addition to natural language reference handling for the Analyst's Assistant, the system also handles graphical references to locations. This occurs in two forms. First, the user may click on a map position, which automatically updates the system's record of most recently-concerned positions to correspond to that position. This enables the user to employ natural language expressions such as "that position" or "there" to refer to the clicked position in subsequent requests. Second, the user may click on a listed position in a natural language response generated by the system. This also updates the system's record of most recently-concerned positions. In this manner, the user may employ whichever mode of specification of positions is most suitable at a particular point in the analysis: literal coordinates, reference to a recently-concerned position, reference to a named position, clicked position on the map, or clicked position in a natural language response.

4. Conclusions

The Analyst's Assistant serves as a demonstration of functionality for collaborative human–system interpretation of sensor data, with the system's implementation grounded in language-based representations at all levels of its processing. From the supplied data expressed in terms of timestamped vehicle positions plus static road segment positions, the system computes states and changes in language-oriented attributes, and by comparing this computed scene information with language-based models of the temporal unfolding of various types of events, the system recognizes instances of those events. Operations for performing event recognition, summarization, focusing of attention for the system, and other tasks are expressed internally using language, supporting the interpretation of user-submitted, free natural language requests and the generation of descriptive natural language responses.

Two broad capabilities are exhibited by the Analyst's Assistant. The first concerns recognition of events and summarization of event sequences. For this capability, it has been found to be quite useful to construct event models using the transition space representation, then match these event models to scene data using the "core–periphery" matching scheme. Construction of the event models is straightforward, easily debugged, and sufficient to yield comprehensible explanations of the event recognition and summarization process applied to particular instances of scene data. The second broad capability concerns support for collaboration between a human user and the system. Providing a language-centered, mixed-modal user interface coupled with an appropriate request set allows the user to effectively partition human and software responsibilities according to relative abilities and context-dependent preferences of the user. In particular, the user is able to schedule processing tasks for efficient execution by choosing between focusing steps, which prepare system capabilities for targeted interpretation of particular subsets of the data, with specific interpretation requests chosen to expedite the interpretation of activities within those focus contexts. A flexible reference handling mechanism also helps reduce cognitive overload on the part of the user by making system-maintained records of recently-concerned data items and results easily accessible to the user in subsequent interpretation steps.

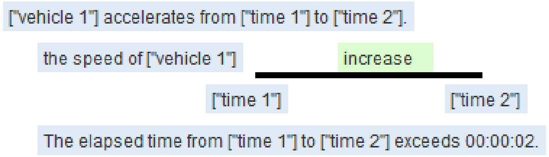
There are a number of unfinished aspects of the Analyst's Assistant. Provided with the opportunity, we hope to extend this work by completing the following:

- adding coverage of new event types: possible pick-ups and drop-offs of people and items, distinguishing "following" a vehicle from "tailing" or "chasing" it, interpreting larger activities like "canvassing" or "patrolling" an area;
- fine-tuning the event models for maximal performance, given additional dataset examples for development and testing;
- enabling the user to customize event models for recognition in particular situations and enabling the user to define new types of events for recognition, using language-motivated descriptions of those events;
- application of the Analyst's Assistant to related types of datasets: vehicle movement data acquired through other types of sensors, air/sea/rail transport, and overhead observation of human activities;
- generating more compact summaries, particularly with respect to reducing redundancy that arises from partially-overlapping event descriptions and incorporating natural language phrasings that lead to concise, readily comprehended event listings within summaries; also, user-controlled level of detail in summaries;
- constructing explanations of system reasoning, in particular concerning why the system has or has not concluded that particular types of events have occurred in particular situations;
- increasing the system's coverage of alternative ways in which the user may express requests in natural language;
- increasing the range of ways in which users may reference previously-mentioned quantities, including additional integration of reference-handling mechanisms for the natural language and graphical interaction modalities.

A. Event Models Used within the Analyst's Assistant

The following 35 event models are defined within the Analyst's Assistant. The event models are listed here in a form that converts parts of the underlying Möbius to text and presents a graphical arrangement of changes, states and other assertions within each event model. All event types except those describing travel from one location to another location are used by the summarization capability. Event types used to answer event-specific natural language queries are: accelerate or decelerate, move or not move, turn or not turn, travel a heading on a road, turn a direction, stop or proceed through or turn a direction at an intersection, make a short or medium-length or long stop, travel from the area within a distance of a position to the area within a second distance of a second position, follow a vehicle, and meet with a vehicle.

Movement Events

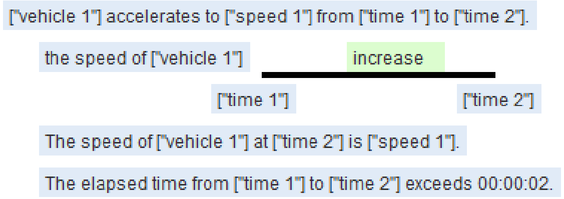


["vehicle 1"] accelerates from ["time 1"] to ["time 2"].

the speed of ["vehicle 1"] increase

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.



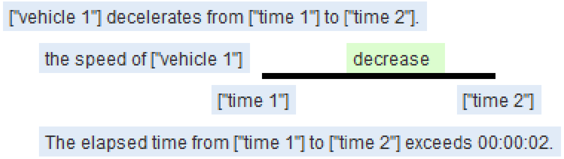
["vehicle 1"] accelerates to ["speed 1"] from ["time 1"] to ["time 2"].

the speed of ["vehicle 1"] increase

["time 1"] ["time 2"]

The speed of ["vehicle 1"] at ["time 2"] is ["speed 1"].

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.

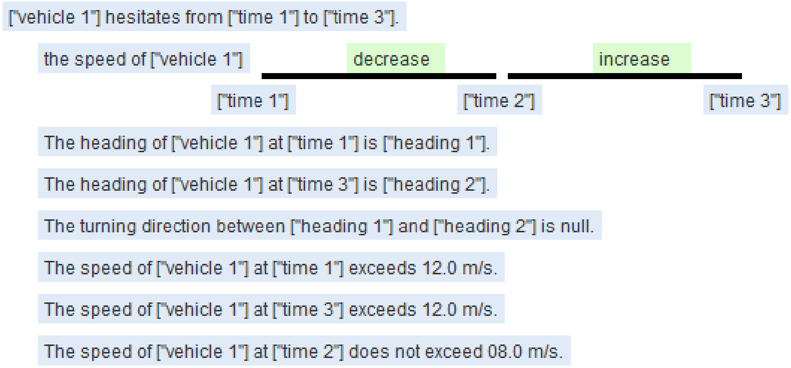


["vehicle 1"] decelerates from ["time 1"] to ["time 2"].

the speed of ["vehicle 1"] decrease

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.



["vehicle 1"] hesitates from ["time 1"] to ["time 3"].

the speed of ["vehicle 1"] decrease increase

["time 1"] ["time 2"] ["time 3"]

The heading of ["vehicle 1"] at ["time 1"] is ["heading 1"].

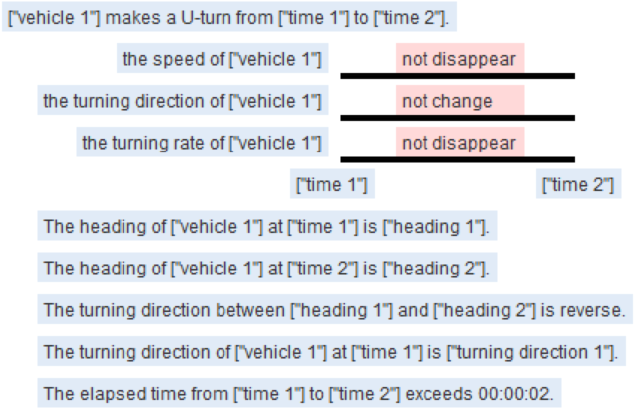
The heading of ["vehicle 1"] at ["time 3"] is ["heading 2"].

The turning direction between ["heading 1"] and ["heading 2"] is null.

The speed of ["vehicle 1"] at ["time 1"] exceeds 12.0 m/s.

The speed of ["vehicle 1"] at ["time 3"] exceeds 12.0 m/s.

The speed of ["vehicle 1"] at ["time 2"] does not exceed 08.0 m/s.



["vehicle 1"] makes a U-turn from ["time 1"] to ["time 2"].

the speed of ["vehicle 1"] not disappear

the turning direction of ["vehicle 1"] not change

the turning rate of ["vehicle 1"] not disappear

["time 1"] ["time 2"]

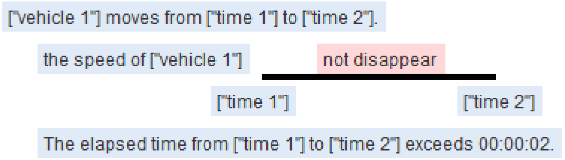
The heading of ["vehicle 1"] at ["time 1"] is ["heading 1"].

The heading of ["vehicle 1"] at ["time 2"] is ["heading 2"].

The turning direction between ["heading 1"] and ["heading 2"] is reverse.

The turning direction of ["vehicle 1"] at ["time 1"] is ["turning direction 1"].

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.

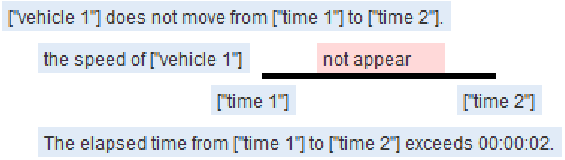


["vehicle 1"] moves from ["time 1"] to ["time 2"].

the speed of ["vehicle 1"] not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.

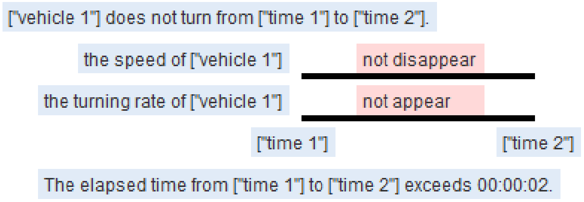


["vehicle 1"] does not move from ["time 1"] to ["time 2"].

the speed of ["vehicle 1"] not appear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.



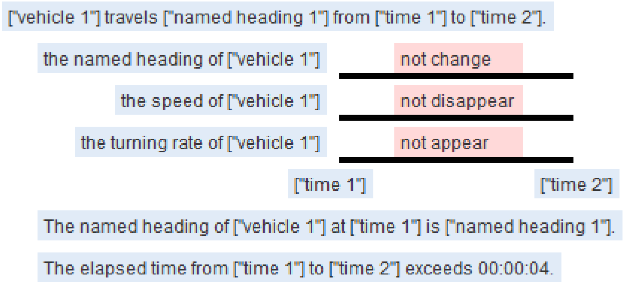
["vehicle 1"] does not turn from ["time 1"] to ["time 2"].

the speed of ["vehicle 1"] not disappear

the turning rate of ["vehicle 1"] not appear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.



["vehicle 1"] travels ["named heading 1"] from ["time 1"] to ["time 2"].

the named heading of ["vehicle 1"] not change

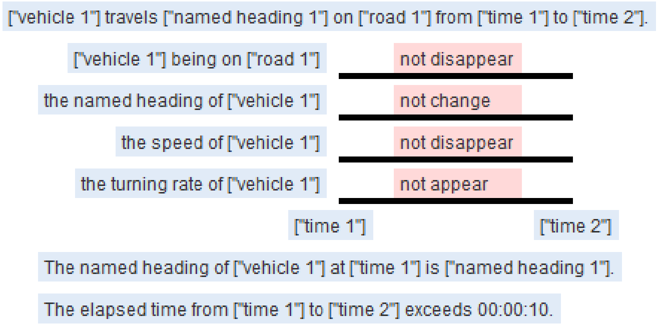
the speed of ["vehicle 1"] not disappear

the turning rate of ["vehicle 1"] not appear

["time 1"] ["time 2"]

The named heading of ["vehicle 1"] at ["time 1"] is ["named heading 1"].

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:04.



["vehicle 1"] travels ["named heading 1"] on ["road 1"] from ["time 1"] to ["time 2"].

["vehicle 1"] being on ["road 1"] not disappear

the named heading of ["vehicle 1"] not change

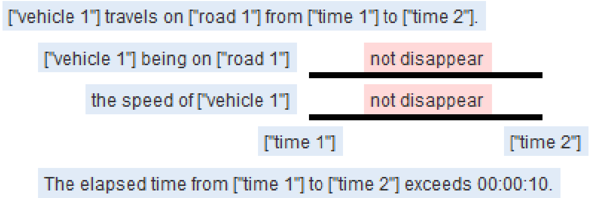
the speed of ["vehicle 1"] not disappear

the turning rate of ["vehicle 1"] not appear

["time 1"] ["time 2"]

The named heading of ["vehicle 1"] at ["time 1"] is ["named heading 1"].

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:10.



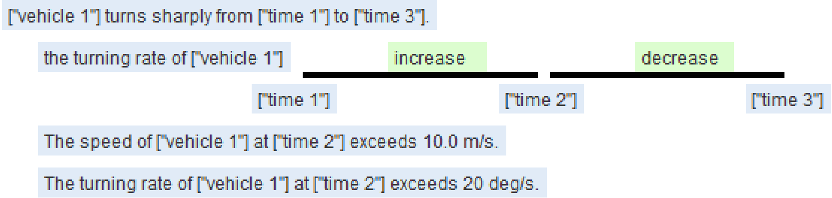
["vehicle 1"] travels on ["road 1"] from ["time 1"] to ["time 2"].

["vehicle 1"] being on ["road 1"] not disappear

the speed of ["vehicle 1"] not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:10.



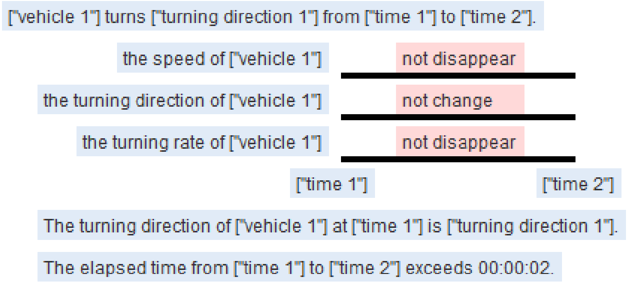
["vehicle 1"] turns sharply from ["time 1"] to ["time 3"].

the turning rate of ["vehicle 1"] increase decrease

["time 1"] ["time 2"] ["time 3"]

The speed of ["vehicle 1"] at ["time 2"] exceeds 10.0 m/s.

The turning rate of ["vehicle 1"] at ["time 2"] exceeds 20 deg/s.



["vehicle 1"] turns ["turning direction 1"] from ["time 1"] to ["time 2"].

the speed of ["vehicle 1"] not disappear

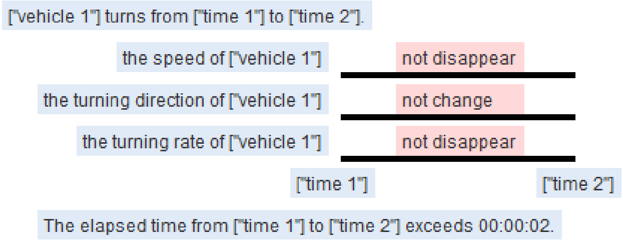
the turning direction of ["vehicle 1"] not change

the turning rate of ["vehicle 1"] not disappear

["time 1"] ["time 2"]

The turning direction of ["vehicle 1"] at ["time 1"] is ["turning direction 1"].

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.



["vehicle 1"] turns from ["time 1"] to ["time 2"].

the speed of ["vehicle 1"] not disappear

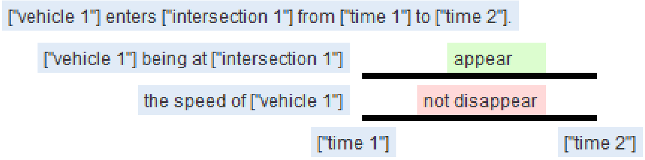
the turning direction of ["vehicle 1"] not change

the turning rate of ["vehicle 1"] not disappear

["time 1"] ["time 2"]

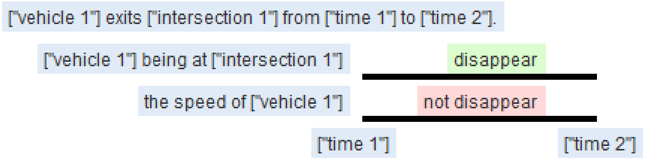
The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.

Intersection Events



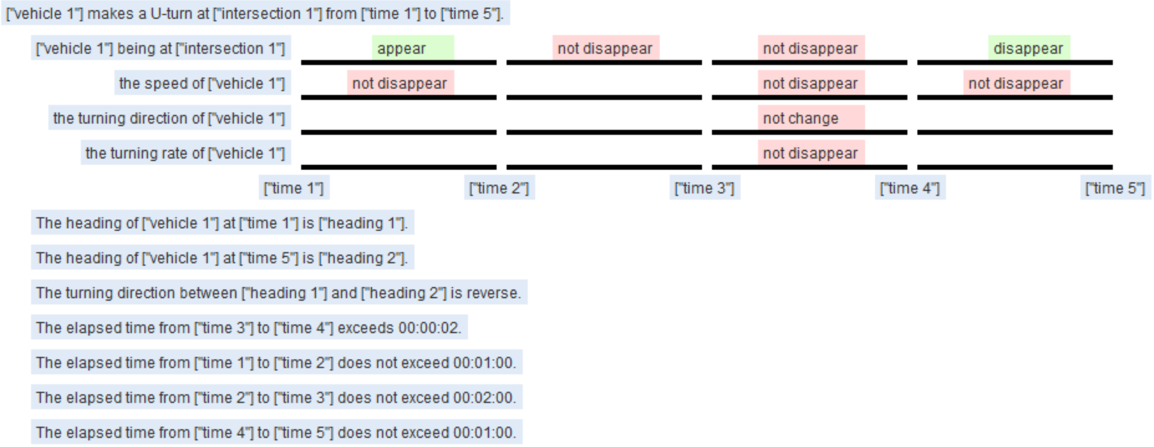
["vehicle 1"] enters ["intersection 1"] from ["time 1"] to ["time 2"].

["vehicle 1"] being at ["intersection 1"]
appear
the speed of ["vehicle 1"]
not disappear
["time 1"]
["time 2"]



["vehicle 1"] exits ["intersection 1"] from ["time 1"] to ["time 2"].

["vehicle 1"] being at ["intersection 1"]
disappear
the speed of ["vehicle 1"]
not disappear
["time 1"]
["time 2"]



["vehicle 1"] makes a U-turn at ["intersection 1"] from ["time 1"] to ["time 5"].

The heading of ["vehicle 1"] at ["time 1"] is ["heading 1"].

The heading of ["vehicle 1"] at ["time 5"] is ["heading 2"].

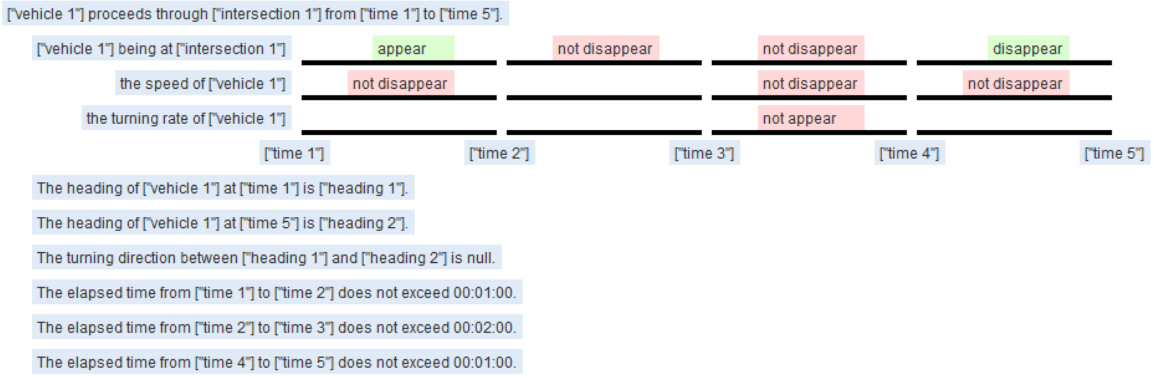
The turning direction between ["heading 1"] and ["heading 2"] is reverse.

The elapsed time from ["time 3"] to ["time 4"] exceeds 00:00:02.

The elapsed time from ["time 1"] to ["time 2"] does not exceed 00:01:00.

The elapsed time from ["time 2"] to ["time 3"] does not exceed 00:02:00.

The elapsed time from ["time 4"] to ["time 5"] does not exceed 00:01:00.



["vehicle 1"] proceeds through ["intersection 1"] from ["time 1"] to ["time 5"].

The heading of ["vehicle 1"] at ["time 1"] is ["heading 1"].

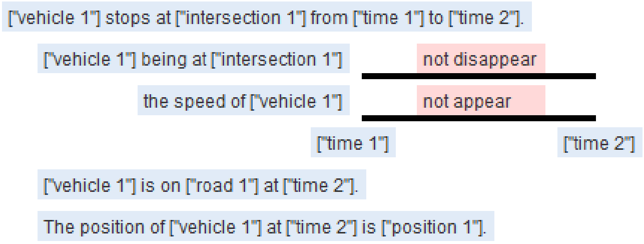
The heading of ["vehicle 1"] at ["time 5"] is ["heading 2"].

The turning direction between ["heading 1"] and ["heading 2"] is null.

The elapsed time from ["time 1"] to ["time 2"] does not exceed 00:01:00.

The elapsed time from ["time 2"] to ["time 3"] does not exceed 00:02:00.

The elapsed time from ["time 4"] to ["time 5"] does not exceed 00:01:00.

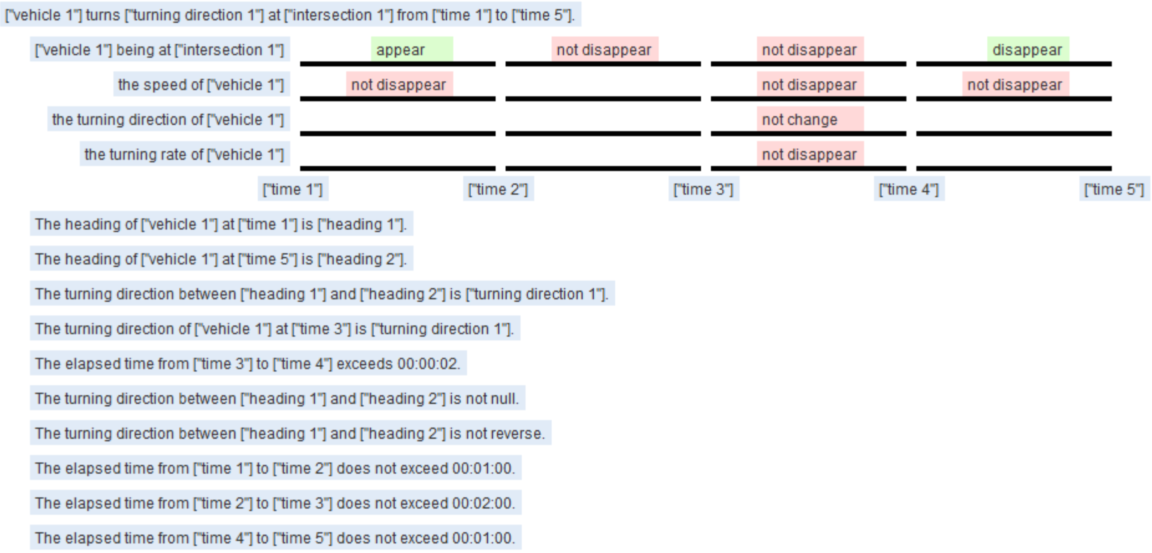


["vehicle 1"] stops at ["intersection 1"] from ["time 1"] to ["time 2"].

	["time 1"] – ["time 2"]
["vehicle 1"] being at ["intersection 1"]	not disappear
the speed of ["vehicle 1"]	not appear

["vehicle 1"] is on ["road 1"] at ["time 2"].

The position of ["vehicle 1"] at ["time 2"] is ["position 1"].



["vehicle 1"] turns ["turning direction 1"] at ["intersection 1"] from ["time 1"] to ["time 5"].

	["time 1"] – ["time 2"]	["time 2"] – ["time 3"]	["time 3"] – ["time 4"]	["time 4"] – ["time 5"]
["vehicle 1"] being at ["intersection 1"]	appear	not disappear	not disappear	disappear
the speed of ["vehicle 1"]	not disappear		not disappear	not disappear
the turning direction of ["vehicle 1"]			not change	
the turning rate of ["vehicle 1"]			not disappear	

The heading of ["vehicle 1"] at ["time 1"] is ["heading 1"].

The heading of ["vehicle 1"] at ["time 5"] is ["heading 2"].

The turning direction between ["heading 1"] and ["heading 2"] is ["turning direction 1"].

The turning direction of ["vehicle 1"] at ["time 3"] is ["turning direction 1"].

The elapsed time from ["time 3"] to ["time 4"] exceeds 00:00:02.

The turning direction between ["heading 1"] and ["heading 2"] is not null.

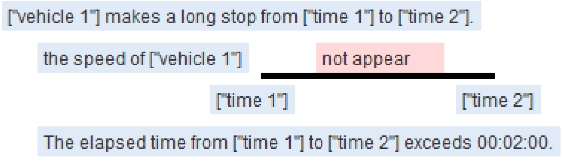
The turning direction between ["heading 1"] and ["heading 2"] is not reverse.

The elapsed time from ["time 1"] to ["time 2"] does not exceed 00:01:00.

The elapsed time from ["time 2"] to ["time 3"] does not exceed 00:02:00.

The elapsed time from ["time 4"] to ["time 5"] does not exceed 00:01:00.

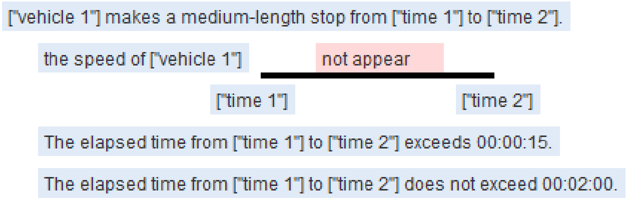
Stopping Events



["vehicle 1"] makes a long stop from ["time 1"] to ["time 2"].

	["time 1"] – ["time 2"]
the speed of ["vehicle 1"]	not appear

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:02:00.

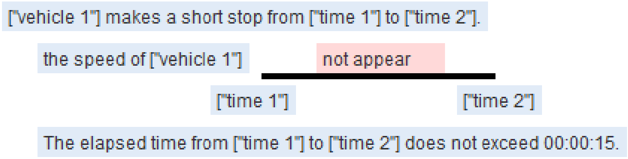


["vehicle 1"] makes a medium-length stop from ["time 1"] to ["time 2"].

	["time 1"] – ["time 2"]
the speed of ["vehicle 1"]	not appear

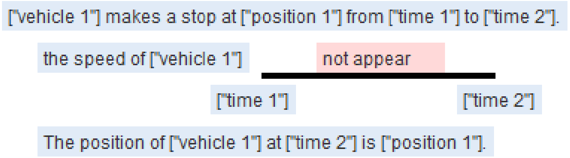
The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:15.

The elapsed time from ["time 1"] to ["time 2"] does not exceed 00:02:00.



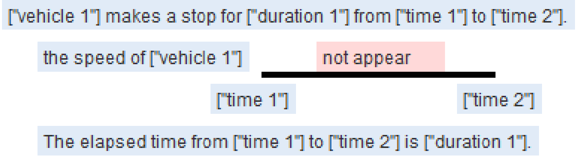
["vehicle 1"] makes a short stop from ["time 1"] to ["time 2"].

The elapsed time from ["time 1"] to ["time 2"] does not exceed 00:00:15.



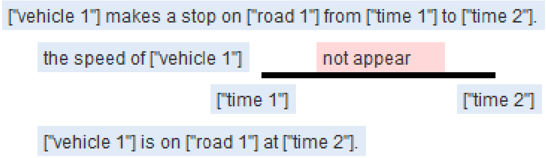
["vehicle 1"] makes a stop at ["position 1"] from ["time 1"] to ["time 2"].

The position of ["vehicle 1"] at ["time 2"] is ["position 1"].



["vehicle 1"] makes a stop for ["duration 1"] from ["time 1"] to ["time 2"].

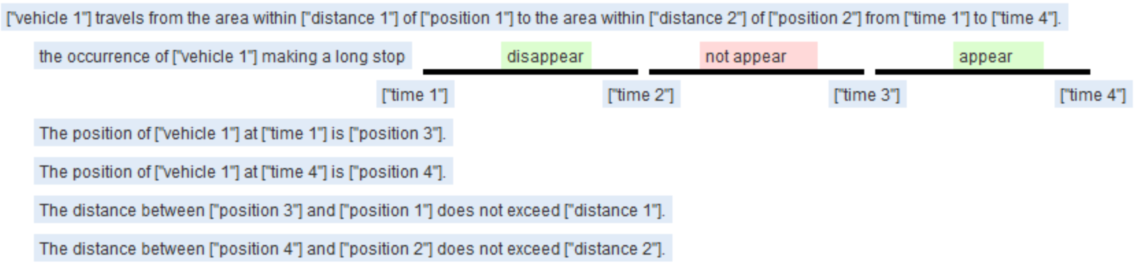
The elapsed time from ["time 1"] to ["time 2"] is ["duration 1"].



["vehicle 1"] makes a stop on ["road 1"] from ["time 1"] to ["time 2"].

["vehicle 1"] is on ["road 1"] at ["time 2"].

Traveling Events



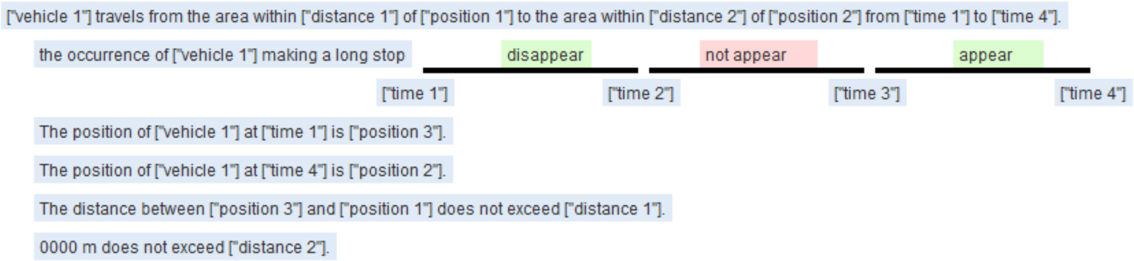
["vehicle 1"] travels from the area within ["distance 1"] of ["position 1"] to the area within ["distance 2"] of ["position 2"] from ["time 1"] to ["time 4"].

The position of ["vehicle 1"] at ["time 1"] is ["position 3"].

The position of ["vehicle 1"] at ["time 4"] is ["position 4"].

The distance between ["position 3"] and ["position 1"] does not exceed ["distance 1"].

The distance between ["position 4"] and ["position 2"] does not exceed ["distance 2"].



["vehicle 1"] travels from the area within ["distance 1"] of ["position 1"] to the area within ["distance 2"] of ["position 2"] from ["time 1"] to ["time 4"].

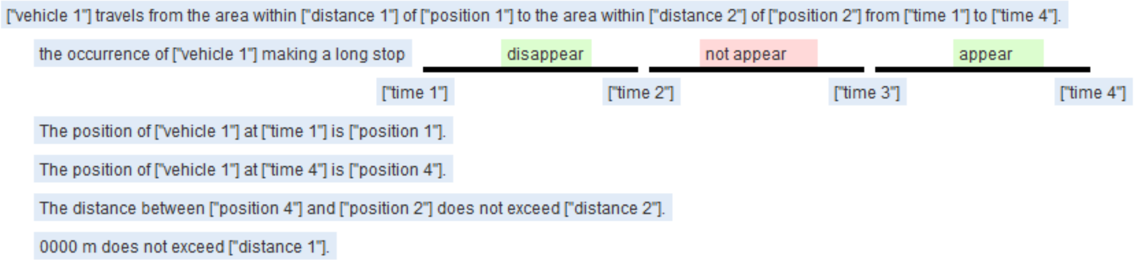
the occurrence of ["vehicle 1"] making a long stop

The position of ["vehicle 1"] at ["time 1"] is ["position 3"].

The position of ["vehicle 1"] at ["time 4"] is ["position 2"].

The distance between ["position 3"] and ["position 1"] does not exceed ["distance 1"].

0000 m does not exceed ["distance 2"].



["vehicle 1"] travels from the area within ["distance 1"] of ["position 1"] to the area within ["distance 2"] of ["position 2"] from ["time 1"] to ["time 4"].

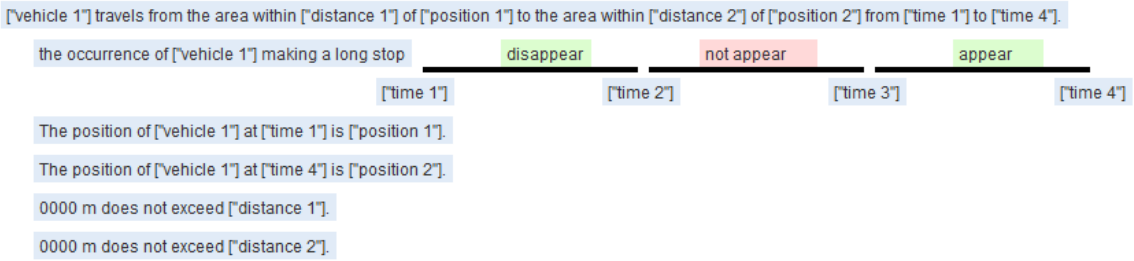
the occurrence of ["vehicle 1"] making a long stop

The position of ["vehicle 1"] at ["time 1"] is ["position 1"].

The position of ["vehicle 1"] at ["time 4"] is ["position 4"].

The distance between ["position 4"] and ["position 2"] does not exceed ["distance 2"].

0000 m does not exceed ["distance 1"].



["vehicle 1"] travels from the area within ["distance 1"] of ["position 1"] to the area within ["distance 2"] of ["position 2"] from ["time 1"] to ["time 4"].

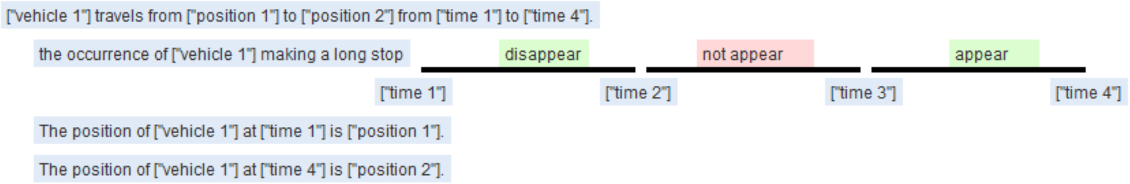
the occurrence of ["vehicle 1"] making a long stop

The position of ["vehicle 1"] at ["time 1"] is ["position 1"].

The position of ["vehicle 1"] at ["time 4"] is ["position 2"].

0000 m does not exceed ["distance 1"].

0000 m does not exceed ["distance 2"].



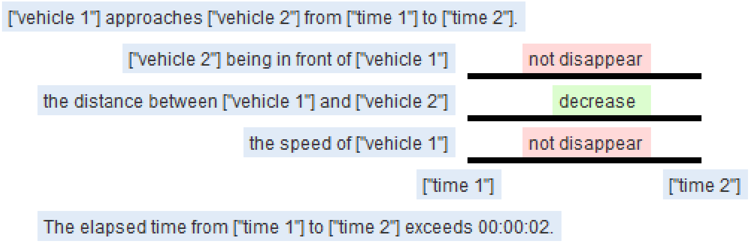
["vehicle 1"] travels from ["position 1"] to ["position 2"] from ["time 1"] to ["time 4"].

the occurrence of ["vehicle 1"] making a long stop

The position of ["vehicle 1"] at ["time 1"] is ["position 1"].

The position of ["vehicle 1"] at ["time 4"] is ["position 2"].

Two-Vehicle Events



["vehicle 1"] approaches ["vehicle 2"] from ["time 1"] to ["time 2"].

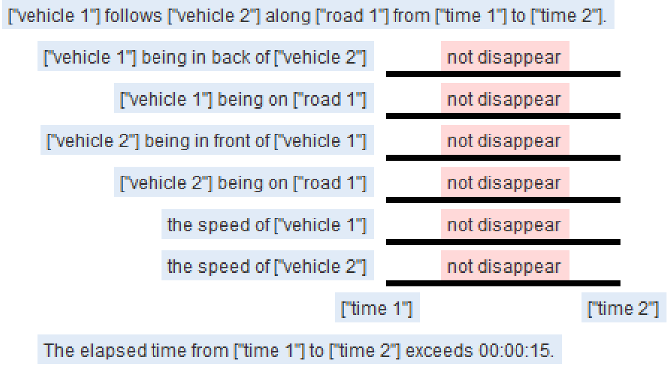
["vehicle 2"] being in front of ["vehicle 1"] — not disappear

the distance between ["vehicle 1"] and ["vehicle 2"] — decrease

the speed of ["vehicle 1"] — not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.



["vehicle 1"] follows ["vehicle 2"] along ["road 1"] from ["time 1"] to ["time 2"].

["vehicle 1"] being in back of ["vehicle 2"] — not disappear

["vehicle 1"] being on ["road 1"] — not disappear

["vehicle 2"] being in front of ["vehicle 1"] — not disappear

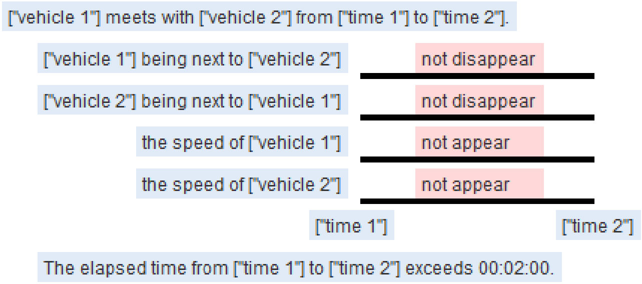
["vehicle 2"] being on ["road 1"] — not disappear

the speed of ["vehicle 1"] — not disappear

the speed of ["vehicle 2"] — not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:15.



["vehicle 1"] meets with ["vehicle 2"] from ["time 1"] to ["time 2"].

["vehicle 1"] being next to ["vehicle 2"] — not disappear

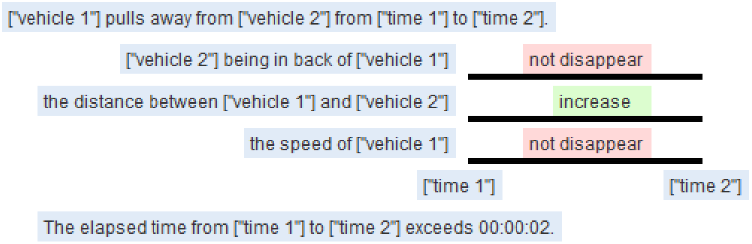
["vehicle 2"] being next to ["vehicle 1"] — not disappear

the speed of ["vehicle 1"] — not appear

the speed of ["vehicle 2"] — not appear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:02:00.



["vehicle 1"] pulls away from ["vehicle 2"] from ["time 1"] to ["time 2"].

["vehicle 2"] being in back of ["vehicle 1"] — not disappear

the distance between ["vehicle 1"] and ["vehicle 2"] — increase

the speed of ["vehicle 1"] — not disappear

["time 1"] ["time 2"]

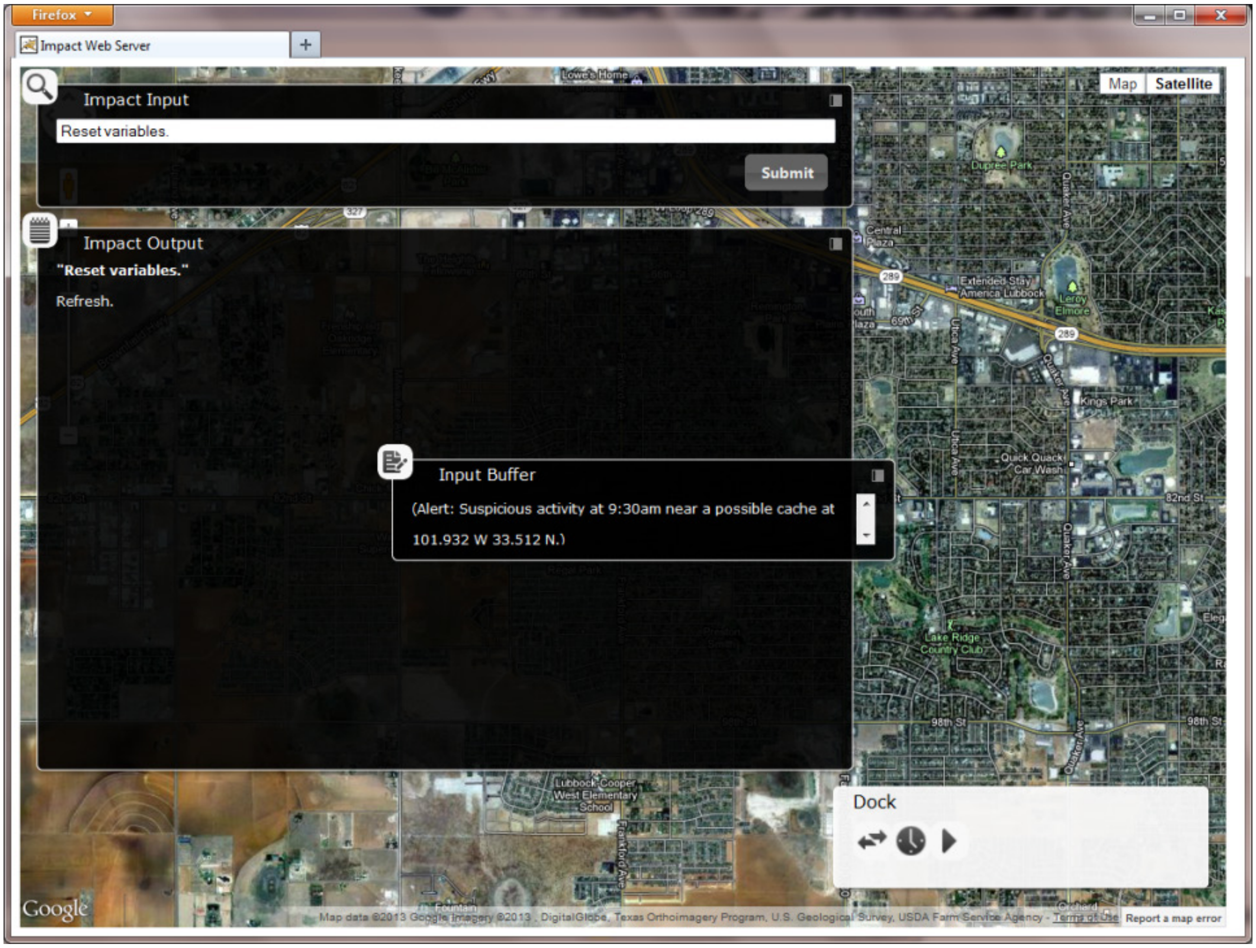
The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:02.

B. Sample Interpretation Sequence

Following is a recorded interaction sequence with the Analyst's Assistant, illustrating the interpretation of vehicle activities that have occurred within the 2 1/2 hour time period covered by the system's supplied dataset. Included within the analyzed data segment is a physically-enacted, simulated IED emplacement and detonation, and the goal of the interaction sequence here is to uncover portions of this included activity in a feed-forward manner, as if responding to real-time inflow of data.

The interaction sequence begins with the user having just initialized a new session. The system then presents an alert to the user:

> Alert: Suspicious activity at 9:30am near a possible cache at 101.932 W 33.512 N.


Firefox
Impact Web Server
Map
Satellite
Impact Input
Submit
Impact Output
Input Buffer
Dock
Google
Map data ©2013 Google Imagery ©2013 , DigitalGlobe, Texas Orthoimagery Program, U.S. Geological Survey, USDA Farm Service Agency - Terms of Use
Report a map error

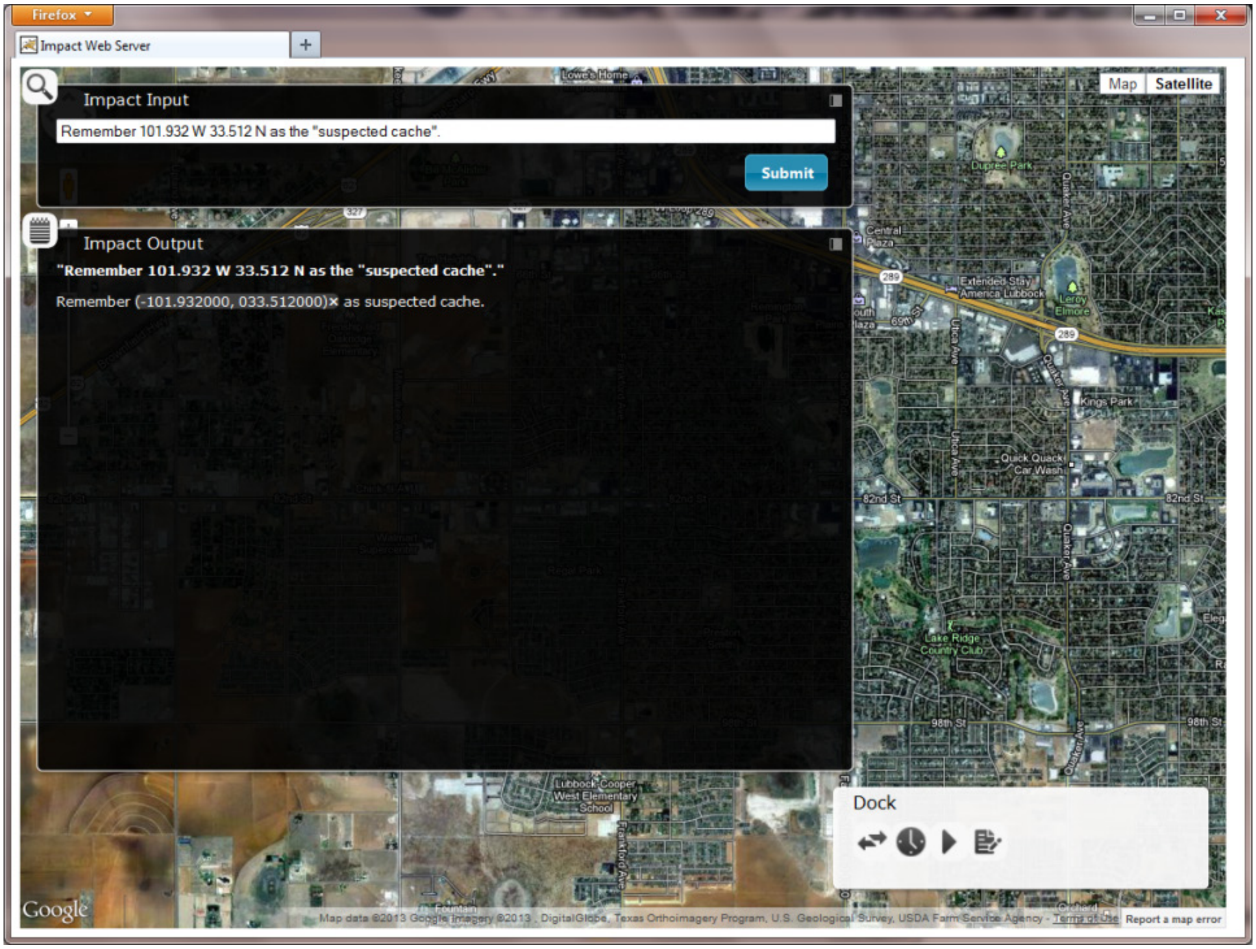
Reset variables.

"Reset variables."

Refresh.

(Alert: Suspicious activity at 9:30am near a possible cache at 101.932 W 33.512 N.)

The user starts the analysis by assigning a natural language name to the coordinate position indicated by the alert, calling this location the “suspected cache”. The Analyst’s Assistant allows its user to assign names to locations, vehicles, times, roads, and sets of these quantities. The names can then be used freely in subsequent natural language requests to the system.



Impact Input

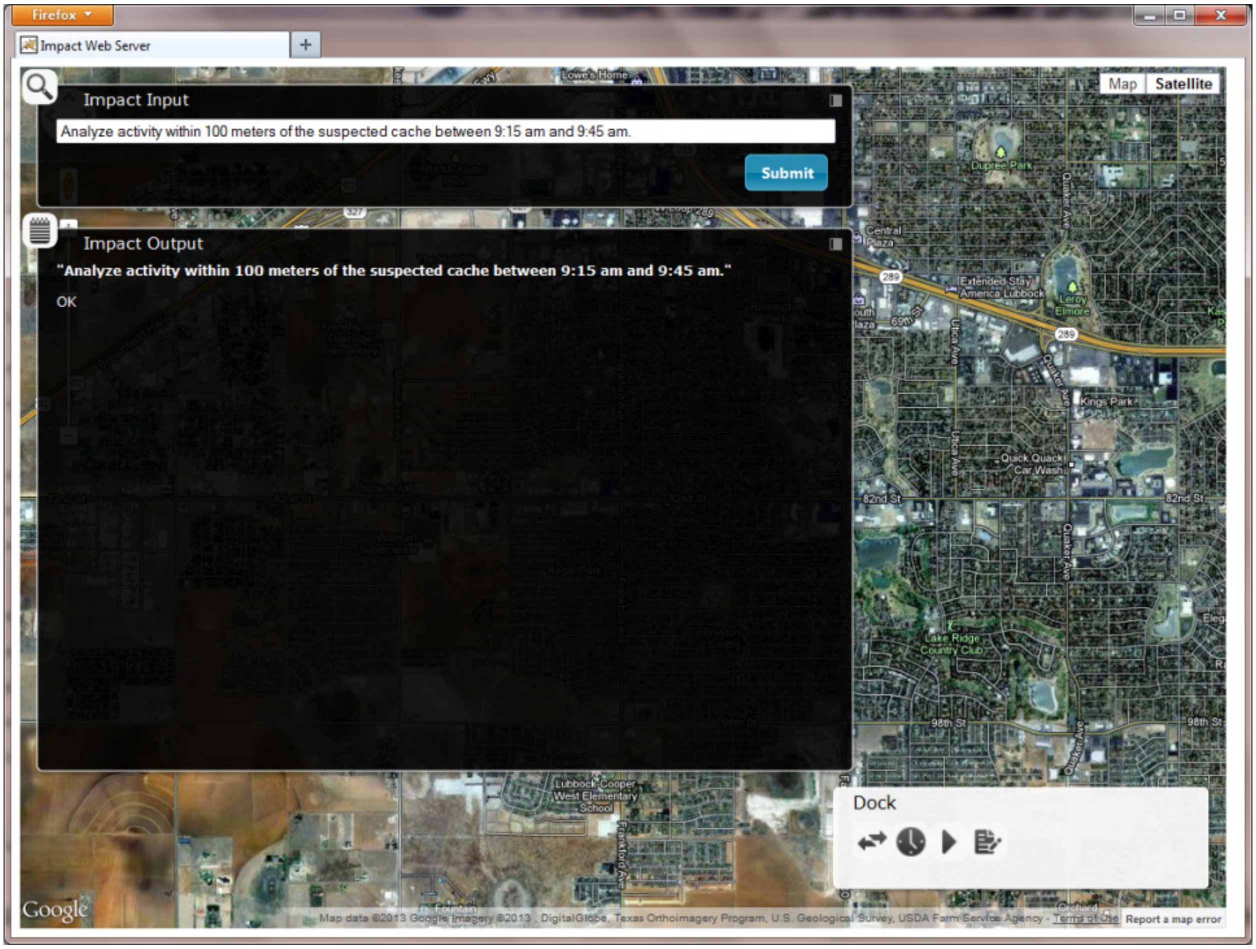
Remember 101.932 W 33.512 N as the "suspected cache".

Impact Output

"Remember 101.932 W 33.512 N as the "suspected cache"."

Remember (-101.932000, 033.512000)× as suspected cache.

Next, the user instructs the system to analyze all activity within a small radius of the suspected cache, over a 30-minute time period surrounding the time of the alert. This is the “focus” step described in Section 3.3, and it causes the system to calculate relevant attribute states and changes within the specified spatial region and time interval, in preparation for subsequent event recognition. Following this focus step, the system can perform either on-demand recognition of particular types of events in response to specific user requests, or it can perform broad event recognition and summarization in response to a more general request by the user. States and changes need not be recalculated for the specified region prior to each subsequent event recognition step concerning that region. The focus step carried out here takes approximately 12 seconds to complete.



Impact Input

Analyze activity within 100 meters of the suspected cache between 9:15 am and 9:45 am.

Impact Output

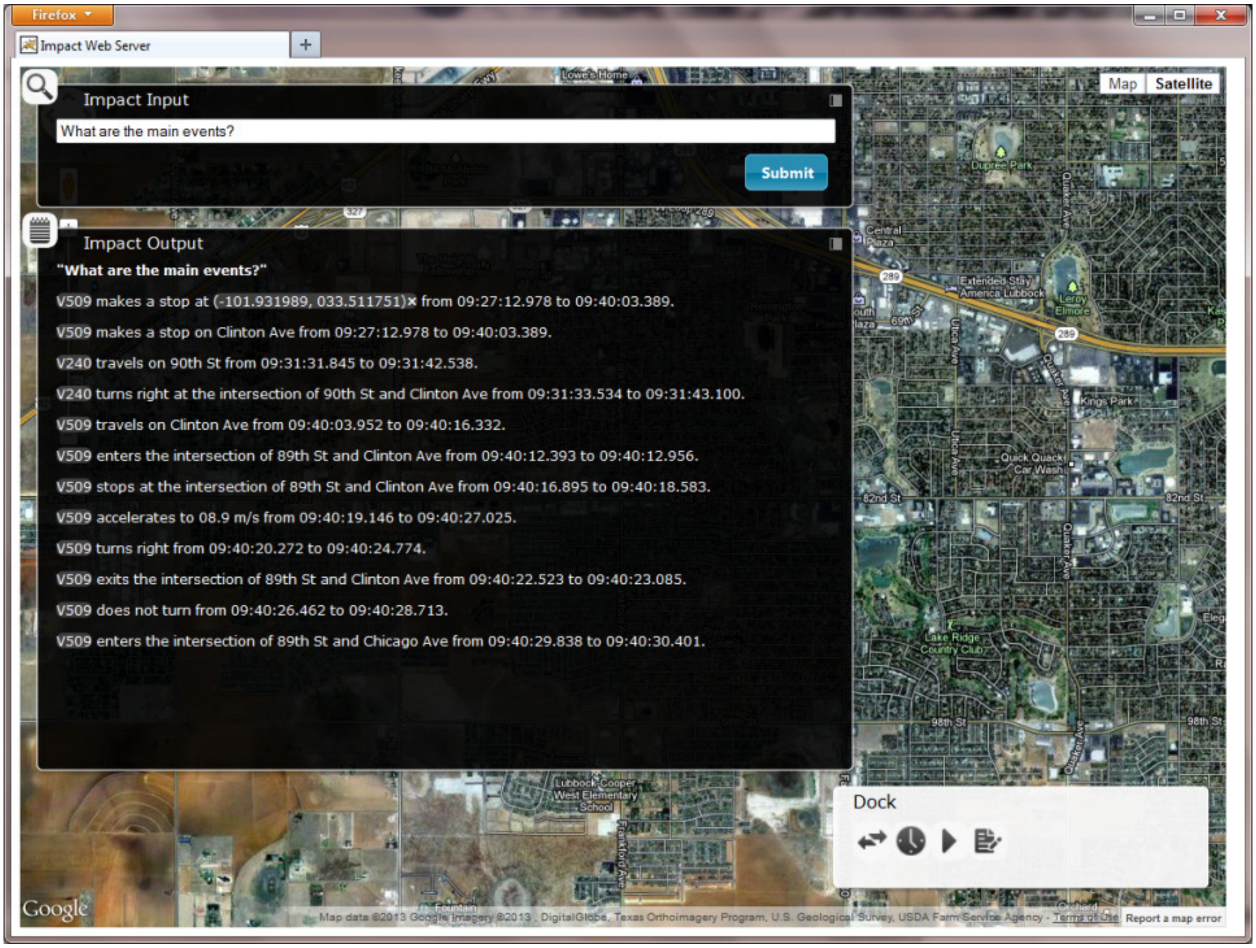
"Analyze activity within 100 meters of the suspected cache between 9:15 am and 9:45 am."

OK

Firefox
Impact Web Server
Map
Satellite
Submit
Dock
Google
Map data ©2013 Google Imagery ©2013 , DigitalGlobe, Texas Orthoimagery Program, U.S. Geological Survey, USDA Farm Service Agency - Terms of Use
Report a map error

Following this focus step, the user requests a broad summary of events identified within the specified region. The system searches for instances of 30 types of events, using the matching strategy described in Section 3.1, and then it applies the summarization strategy described in Section 3.2. The result is a list of 12 event instances involving two vehicles, V509 and V240. The event recognition and summarization steps, in combination, take approximately 10 seconds to complete.

Of interest among the listed events is the relatively long stop V509 has made on Clinton Ave, lasting about 13 minutes from 9:27 AM to 9:40 AM. The other vehicle, V240, passes through the focus region but does not stop.


Firefox
Impact Web Server
Map
Satellite
Submit
Dock
Google
Map data ©2013 Google Imagery ©2013 , DigitalGlobe, Texas Orthoimagery Program, U.S. Geological Survey, USDA Farm Service Agency - Terms of Use
Report a map error

Impact Input

What are the main events?

Impact Output

"What are the main events?"

V509 makes a stop at (-101.931989, 033.511751)× from 09:27:12.978 to 09:40:03.389.

V509 makes a stop on Clinton Ave from 09:27:12.978 to 09:40:03.389.

V240 travels on 90th St from 09:31:31.845 to 09:31:42.538.

V240 turns right at the intersection of 90th St and Clinton Ave from 09:31:33.534 to 09:31:43.100.

V509 travels on Clinton Ave from 09:40:03.952 to 09:40:16.332.

V509 enters the intersection of 89th St and Clinton Ave from 09:40:12.393 to 09:40:12.956.

V509 stops at the intersection of 89th St and Clinton Ave from 09:40:16.895 to 09:40:18.583.

V509 accelerates to 08.9 m/s from 09:40:19.146 to 09:40:27.025.

V509 turns right from 09:40:20.272 to 09:40:24.774.

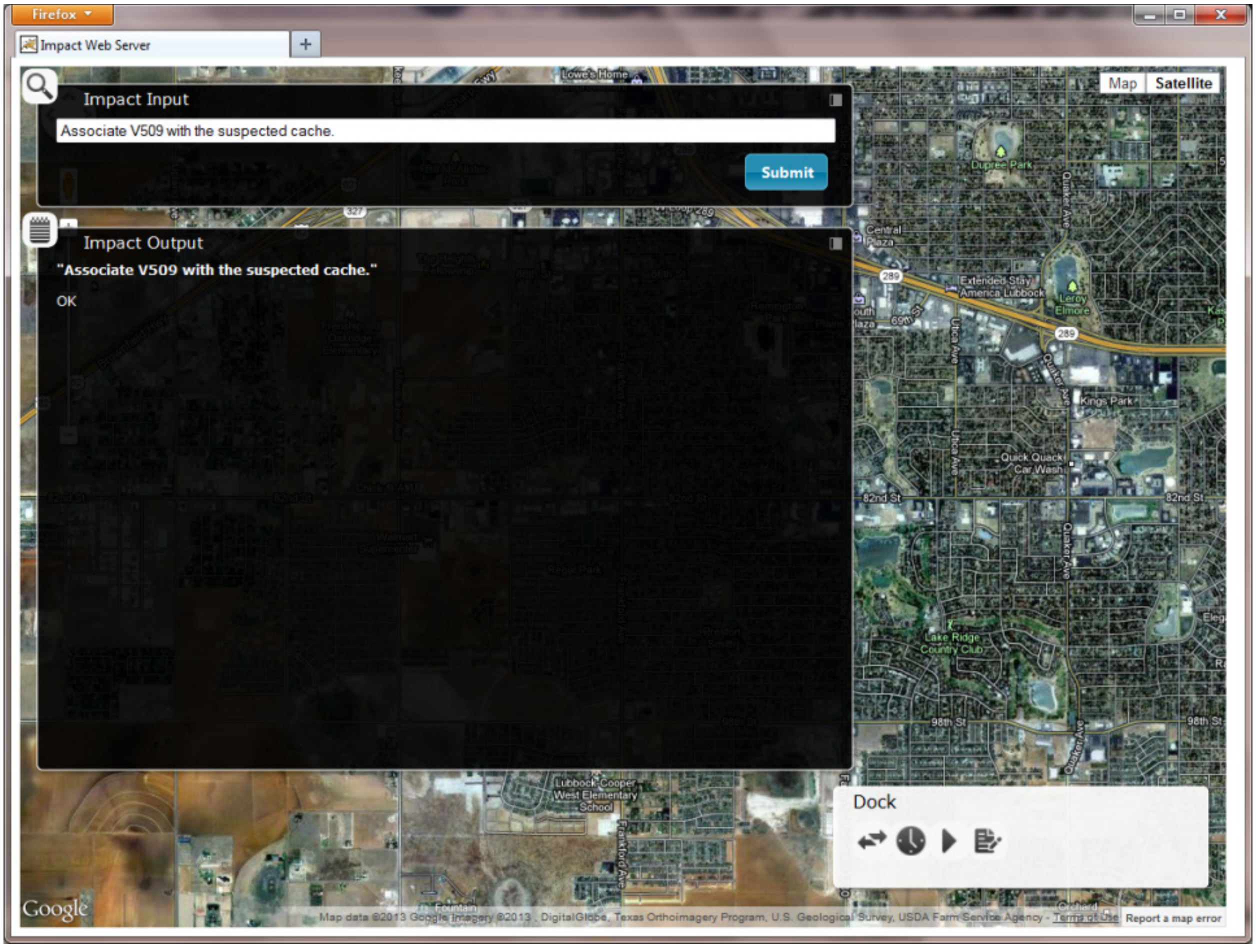
V509 exits the intersection of 89th St and Clinton Ave from 09:40:22.523 to 09:40:23.085.

V509 does not turn from 09:40:26.462 to 09:40:28.713.

V509 enters the intersection of 89th St and Chicago Ave from 09:40:29.838 to 09:40:30.401.

The next several steps illustrate a capability within the Analyst's Assistant that allows the user to note significant locations, vehicles and times and then specify associations between these quantities. The system can then display these associations later for review or reporting.

First, the user specifies an association between vehicle V509 and the suspected cache.

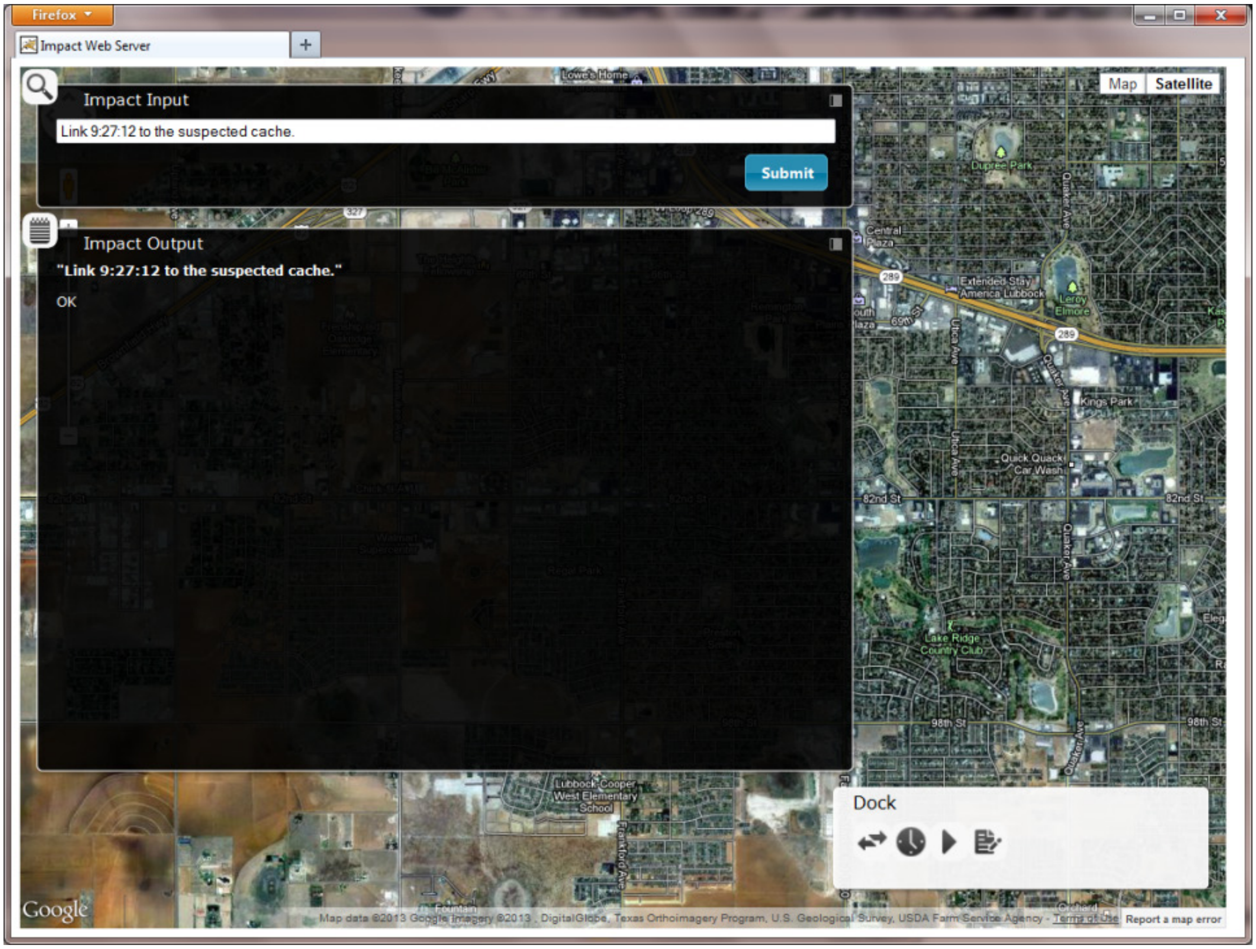


Associate V509 with the suspected cache.

"Associate V509 with the suspected cache."

OK

Next, the user associates the suspected cache with the starting time of V509's stop near that location.



Impact Input

Link 9:27:12 to the suspected cache.

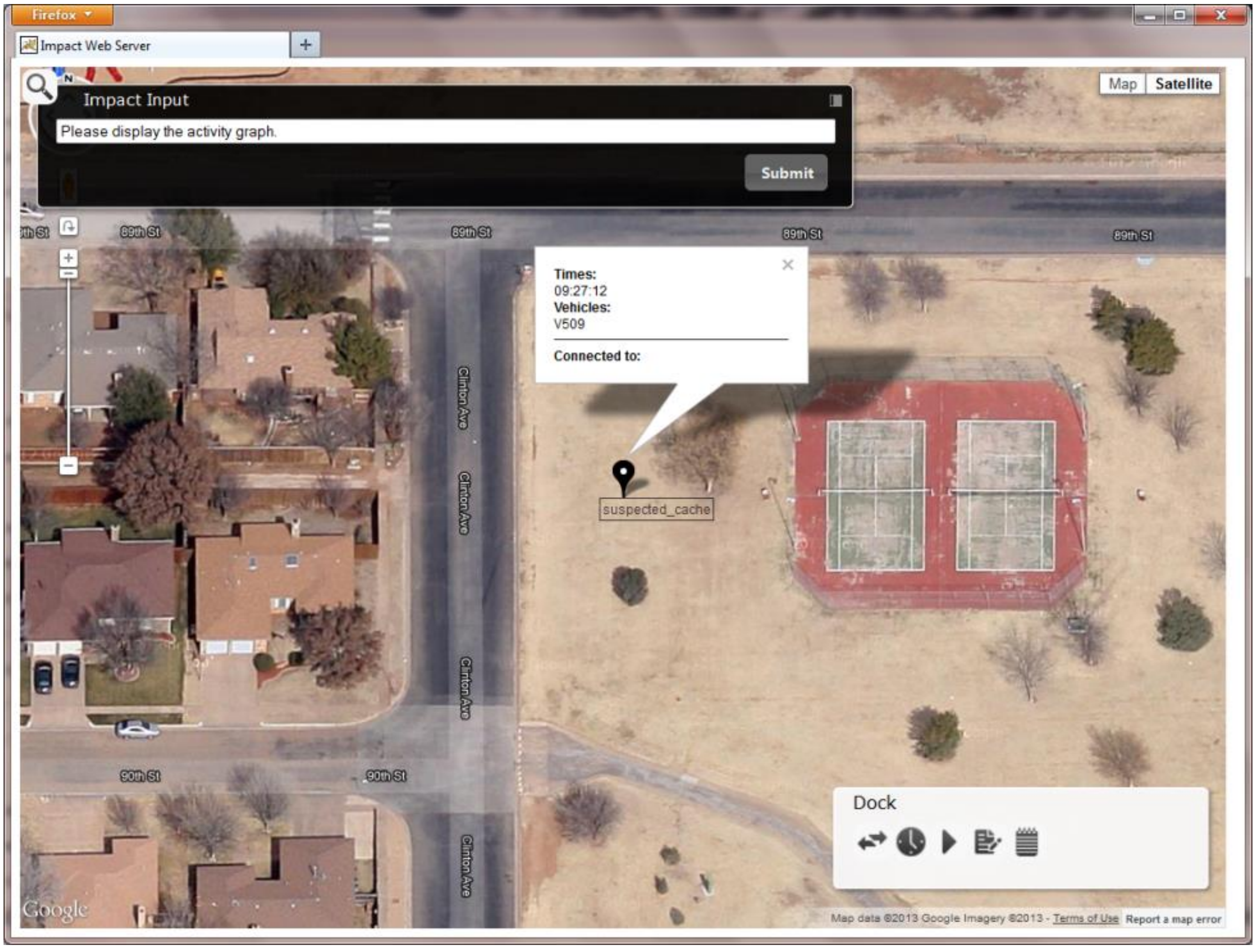
Impact Output

"Link 9:27:12 to the suspected cache."

OK

Firefox
Impact Web Server
Submit
Map
Satellite
Dock
Google
Map data ©2013 Google Imagery ©2013 , DigitalGlobe, Texas Orthoimagery Program, U.S. Geological Survey, USDA Farm Service Agency - Terms of Use
Report a map error

The user then requests a display of the “activity graph”—the accumulated set of significant items and their associations. In this case, the activity graph consists of simply the suspected cache and the associated vehicle and time.



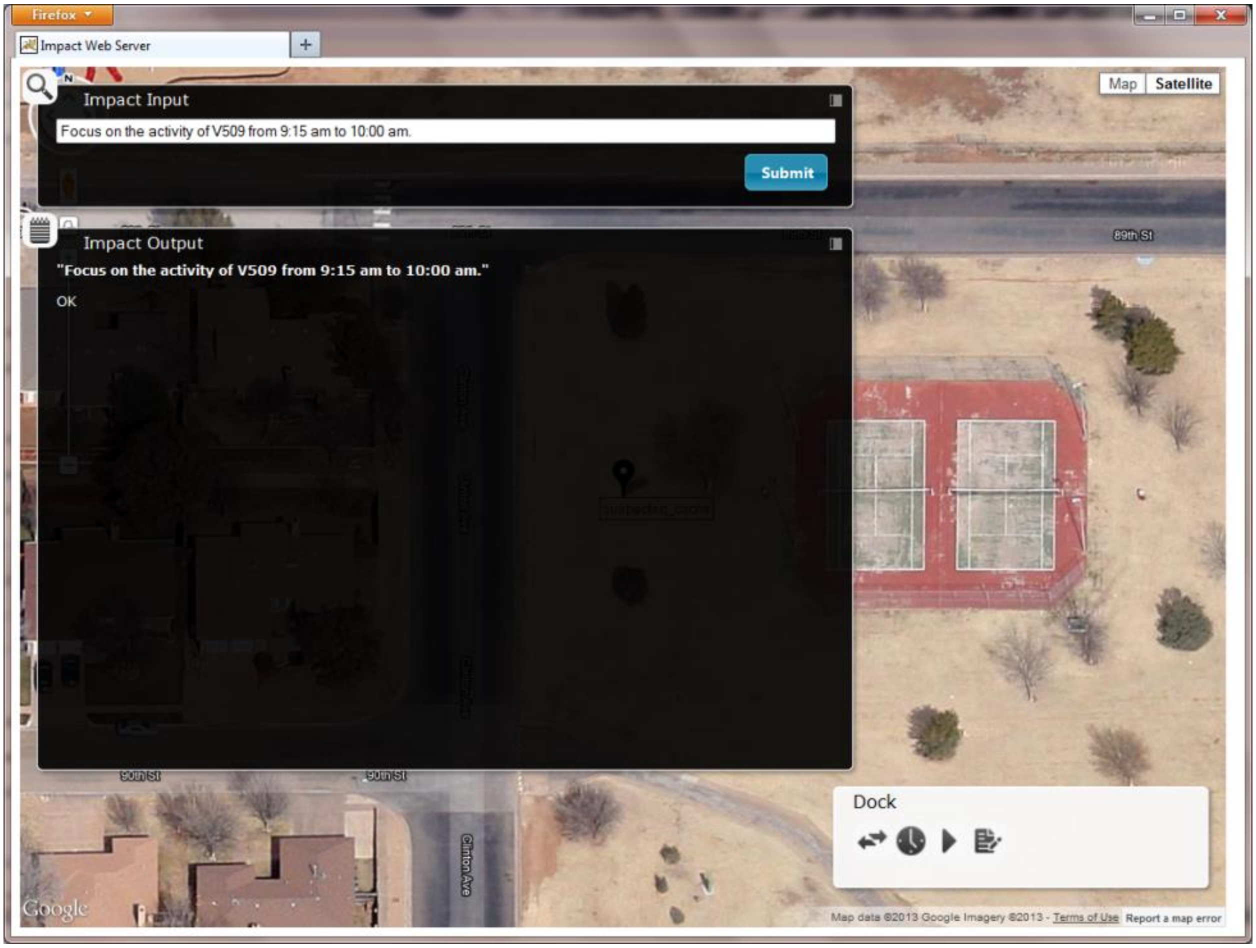
Impact Input

Please display the activity graph.

Times:
09:27:12
Vehicles:
V509

Connected to:

The user then instructs the system to refocus its analysis on vehicle V509 over a 45-minute time period starting before V509's stop on Clinton Ave and continuing for approximately 20 minutes after that stop. This type of "focusing" operation is not limited to a specific spatial region, but rather extends to cover all activities conducted by the specified vehicle or set of vehicles during the specified time period. For the request submitted here, the system takes about 17 seconds to complete its analysis.



Impact Input

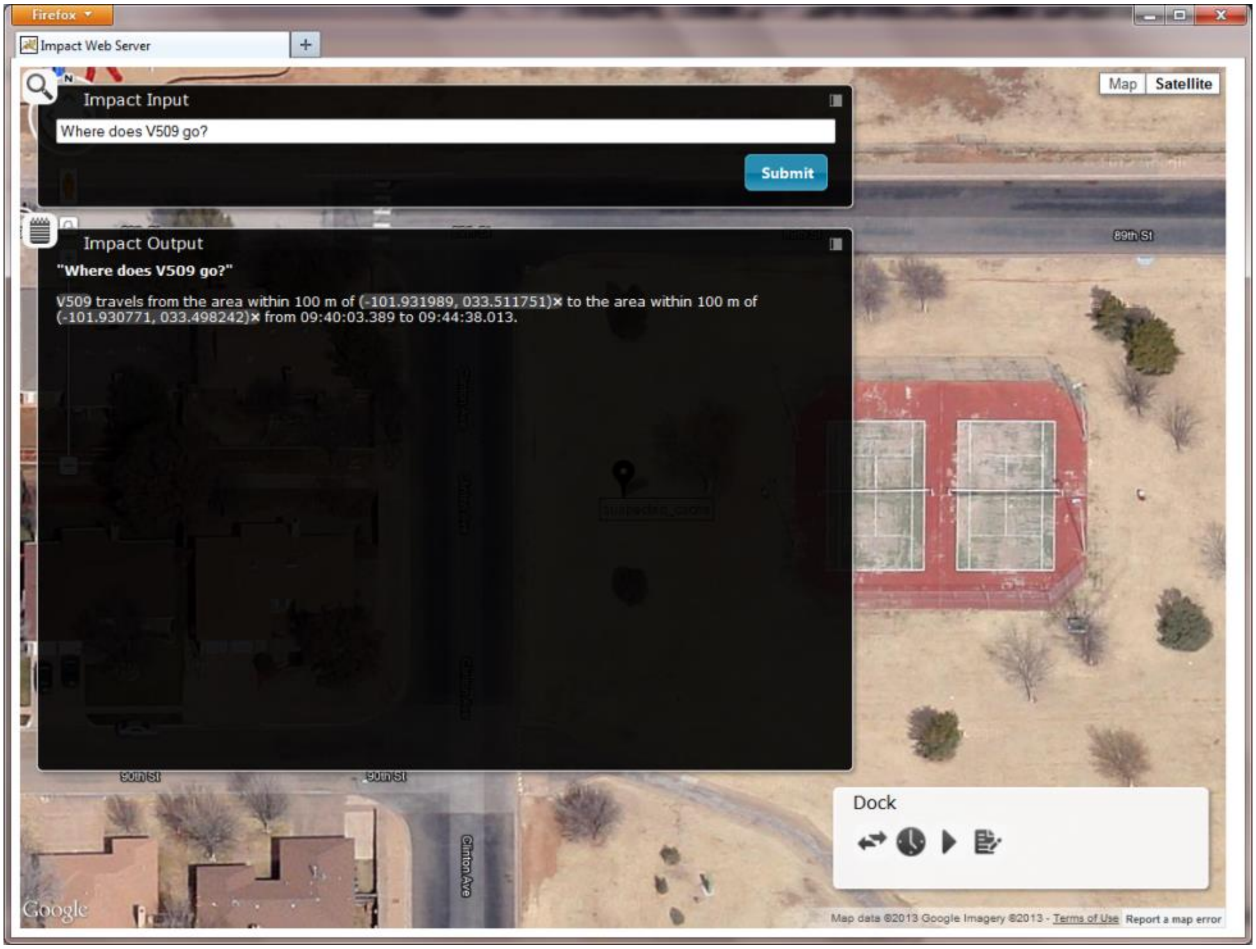
Focus on the activity of V509 from 9:15 am to 10:00 am.

Impact Output

"Focus on the activity of V509 from 9:15 am to 10:00 am."

OK

This time, instead of asking the system to summarize all events for the focus region, the user submits a targeted request for instances of one type of event: traveling from a relatively long stop at one location to a relatively long stop at another location. In response, the system reports that V509 has traveled from its original location near the suspected cache to a new location specified in coordinates. This system takes about 2 seconds to respond to this request.

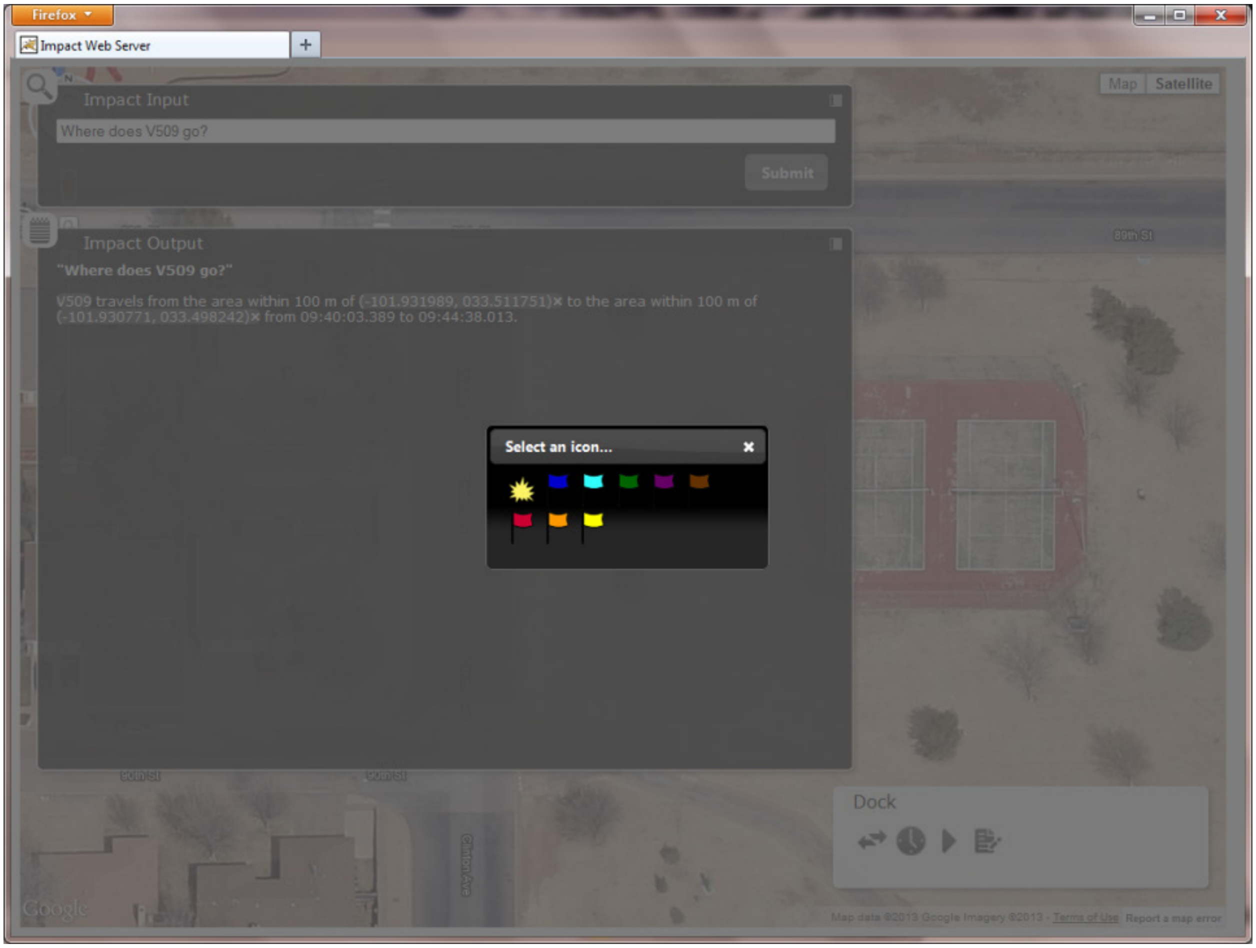


Where does V509 go?

"Where does V509 go?"

V509 travels from the area within 100 m of (-101.931989, 033.511751)× to the area within 100 m of (-101.930771, 033.498242)× from 09:40:03.389 to 09:44:38.013.

The user clicks on the travel destination coordinates in the system's response and is given the option of placing a colored flag on the map at that location. In this case, the user declines to place a flag on the map. However, selection of the travel destination coordinates in itself primes this location to serve as a target of subsequent natural language references to a location of current interest—references like "that location" or "there"—using mechanisms described in Section 3.4.



Impact Input

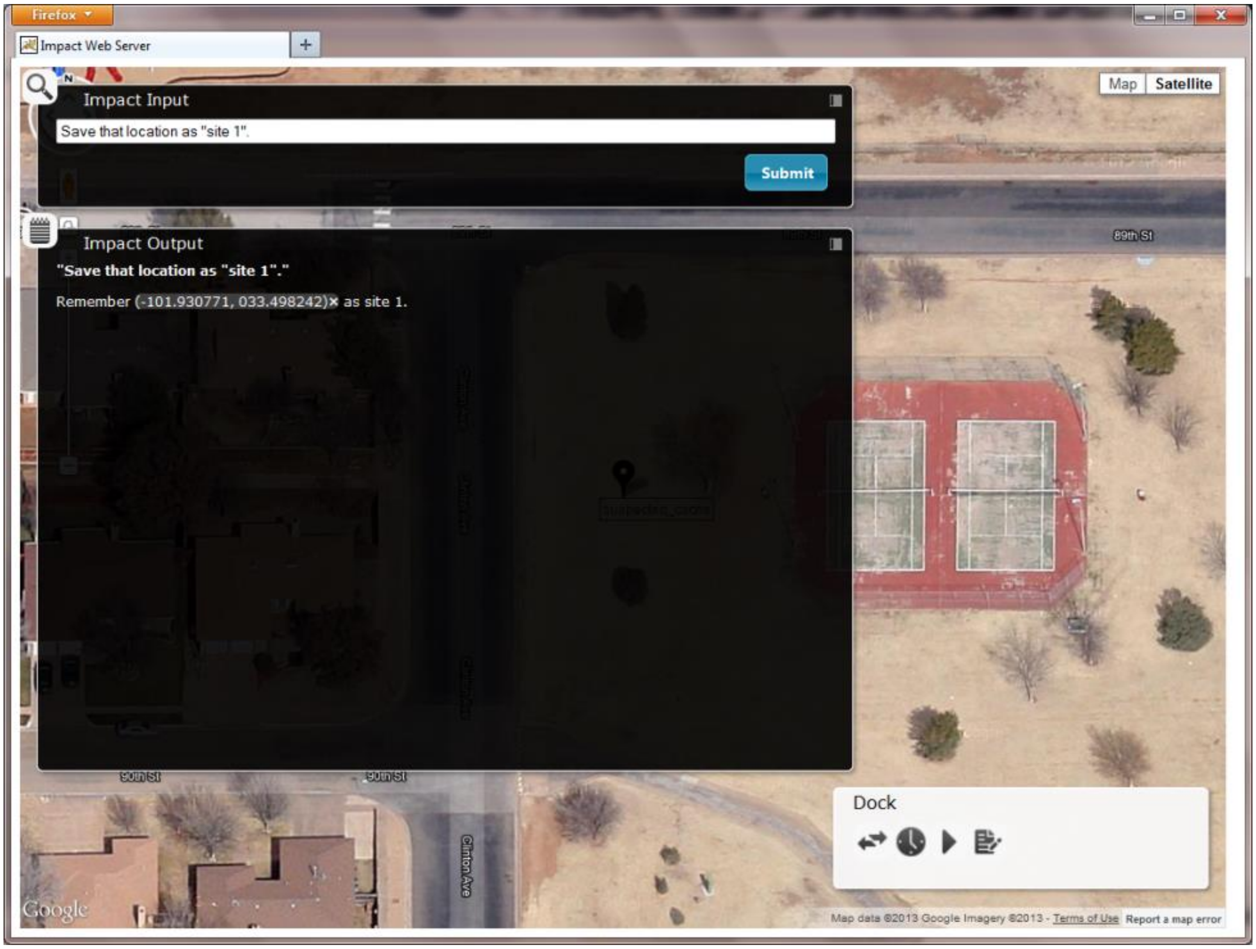
Where does V509 go?

Impact Output

"Where does V509 go?"

V509 travels from the area within 100 m of (-101.931989, 033.511751)× to the area within 100 m of (-101.930771, 033.498242)× from 09:40:03.389 to 09:44:38.013.

An explicit reference of this sort is made next, with the user requesting that the system record a permanent name for the position referred to as “that location”. The new name is “site 1”.



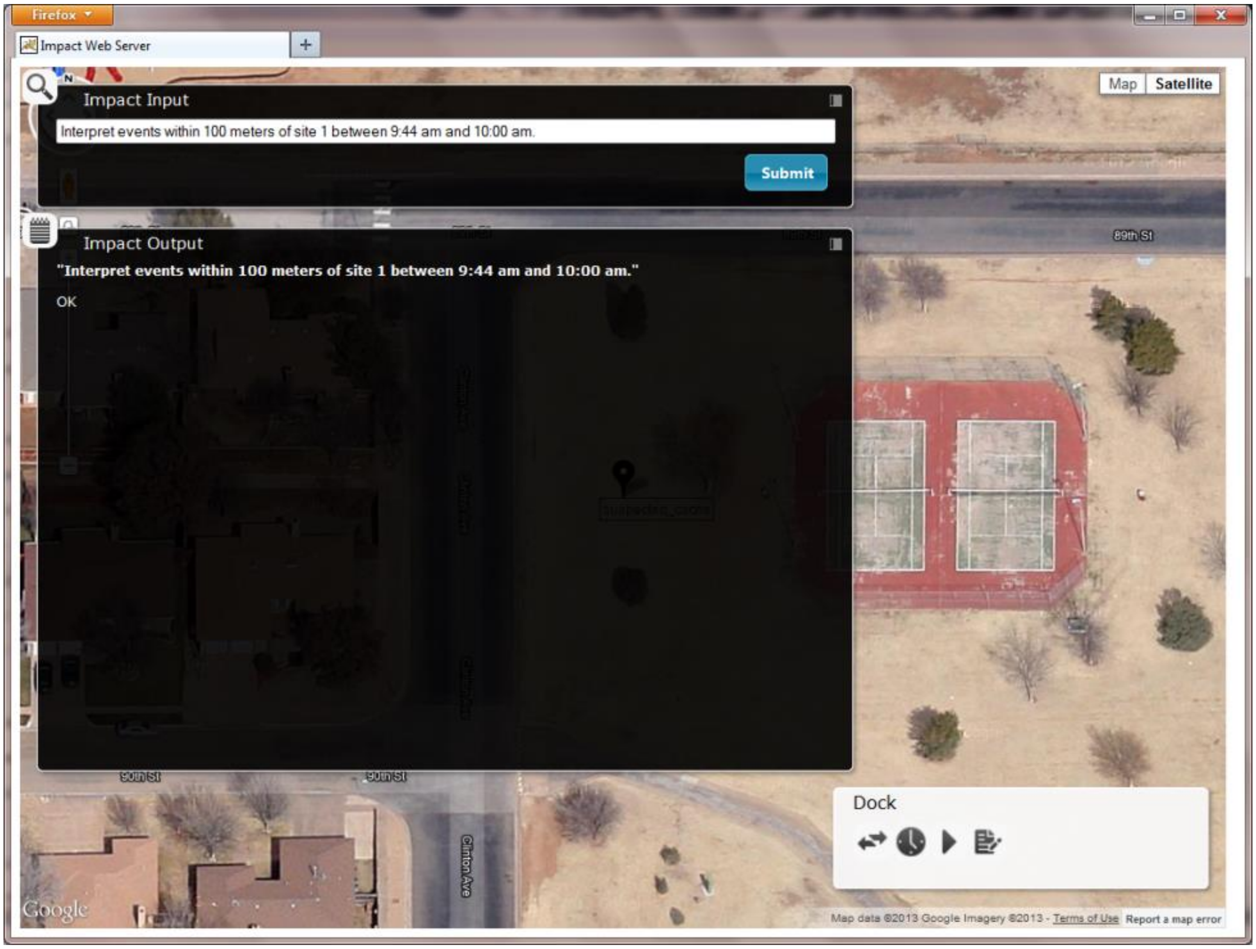
Save that location as "site 1".

"Save that location as "site 1"."

Remember (-101.930771, 033.498242)× as site 1.

89th St
90th St
90th St
Clinton Ave
Dock
Google
Map data ©2013 Google Imagery ©2013 - Terms of Use Report a map error

At this point, the user would like to discover what types of activities are taking place at V509's destination point, "site 1". The user submits a request to focus the system's analysis on a spatial region surrounding this location, for a period of time beginning just before V509 arrives at the site and continuing up to 10:00 AM.

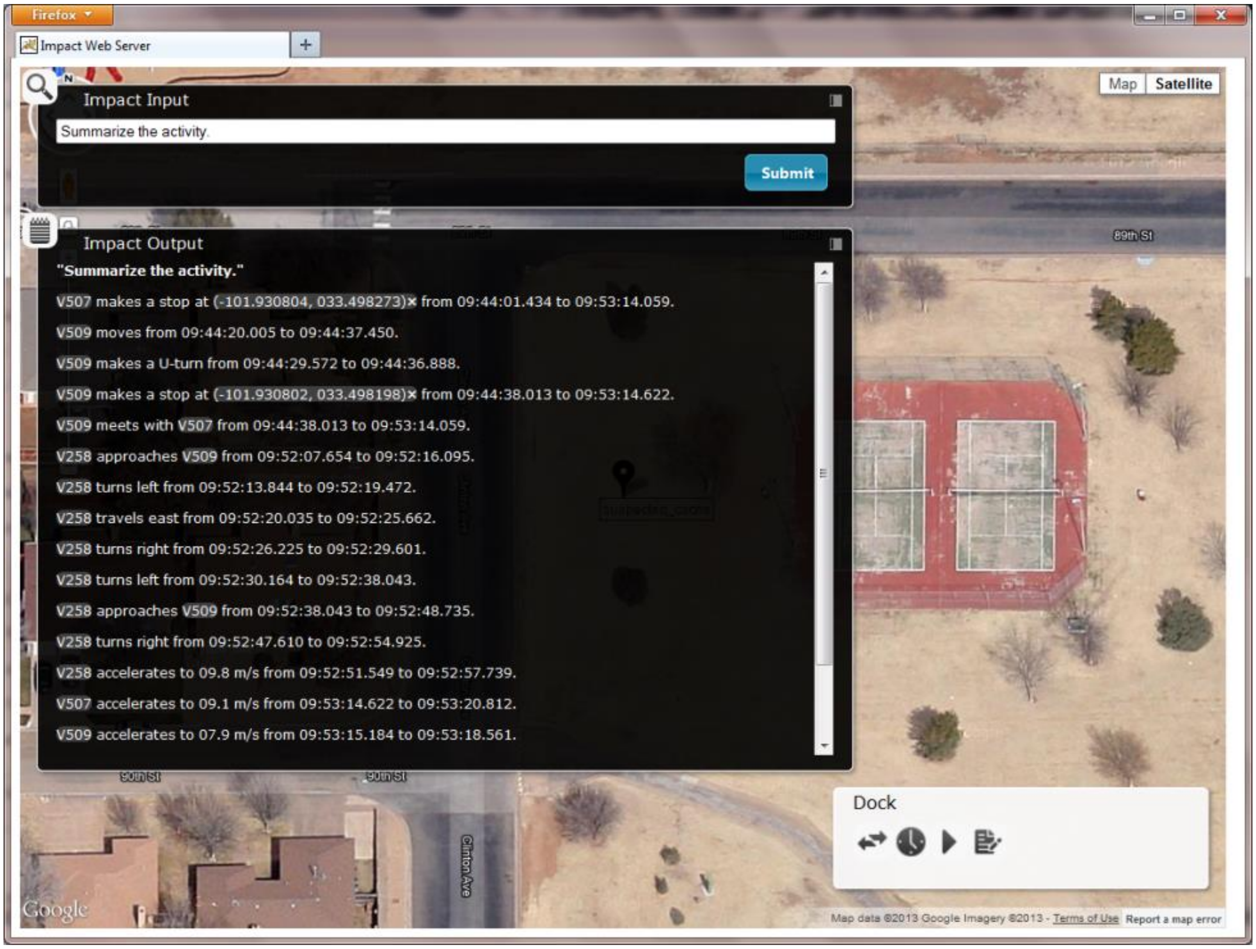


Interpret events within 100 meters of site 1 between 9:44 am and 10:00 am.

"Interpret events within 100 meters of site 1 between 9:44 am and 10:00 am."

OK

The user then asks the system to identify all instances of a range of event types within the focus region and time interval, then summarize its findings. The system reports 18 event instances, the last three viewable through scrolling of the output pane.



Summarize the activity.

Impact Output

"Summarize the activity."

V507 makes a stop at (-101.930804, 033.498273)× from 09:44:01.434 to 09:53:14.059.

V509 moves from 09:44:20.005 to 09:44:37.450.

V509 makes a U-turn from 09:44:29.572 to 09:44:36.888.

V509 makes a stop at (-101.930802, 033.498198)× from 09:44:38.013 to 09:53:14.622.

V509 meets with V507 from 09:44:38.013 to 09:53:14.059.

V258 approaches V509 from 09:52:07.654 to 09:52:16.095.

V258 turns left from 09:52:13.844 to 09:52:19.472.

V258 travels east from 09:52:20.035 to 09:52:25.662.

V258 turns right from 09:52:26.225 to 09:52:29.601.

V258 turns left from 09:52:30.164 to 09:52:38.043.

V258 approaches V509 from 09:52:38.043 to 09:52:48.735.

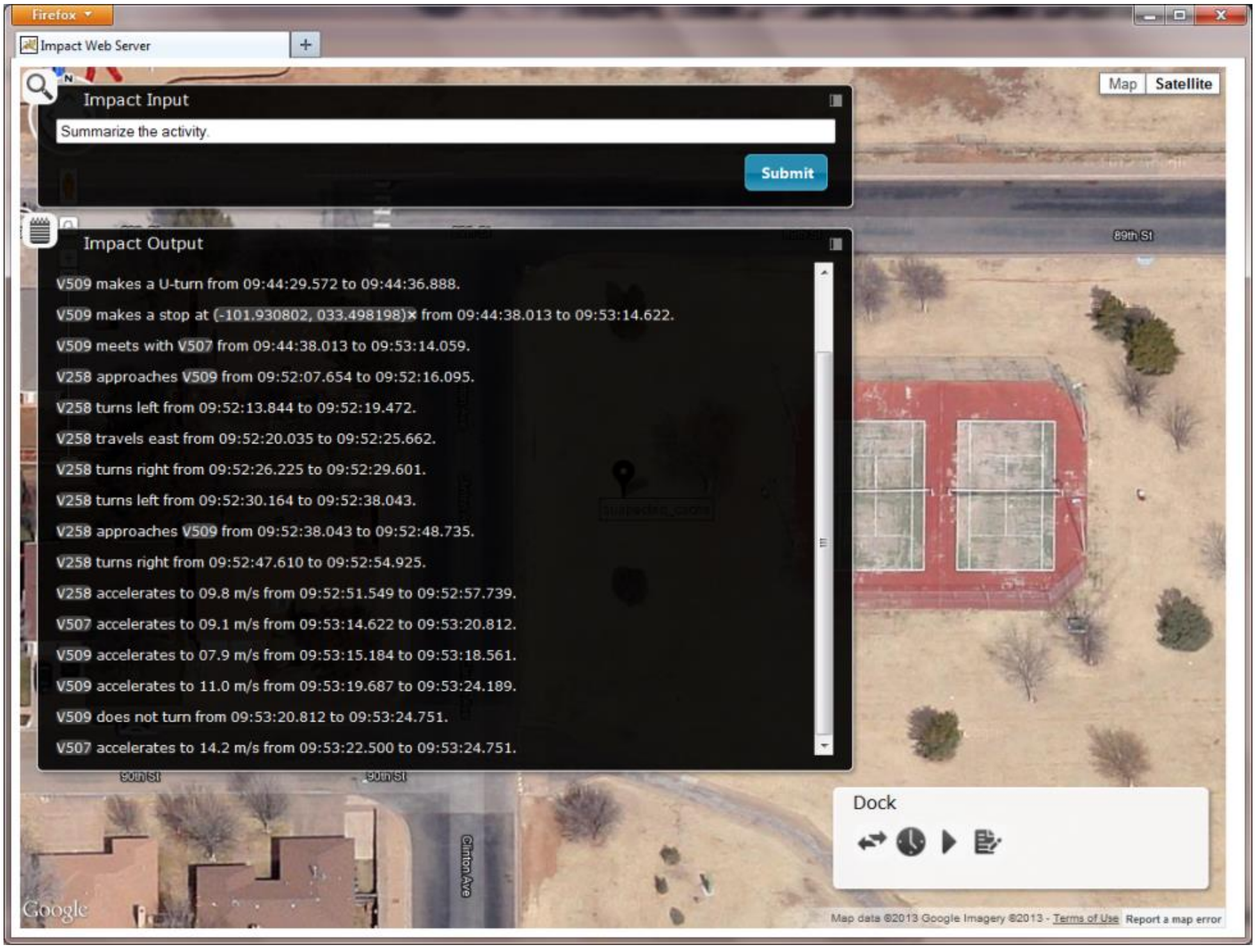
V258 turns right from 09:52:47.610 to 09:52:54.925.

V258 accelerates to 09.8 m/s from 09:52:51.549 to 09:52:57.739.

V507 accelerates to 09.1 m/s from 09:53:14.622 to 09:53:20.812.

V509 accelerates to 07.9 m/s from 09:53:15.184 to 09:53:18.561.

The 18 reported events involve three vehicles: V507, V509 and V258. Of particular interest are the occurrences of V507 and V509 each stopping for approximately 9 minutes at this location, overlapping temporally and spatially with one another to an extent that is sufficient for the system to conclude that the cars are "meeting". V258 travels through the focus region but does not interact with either V507 or V509 in a significant way.

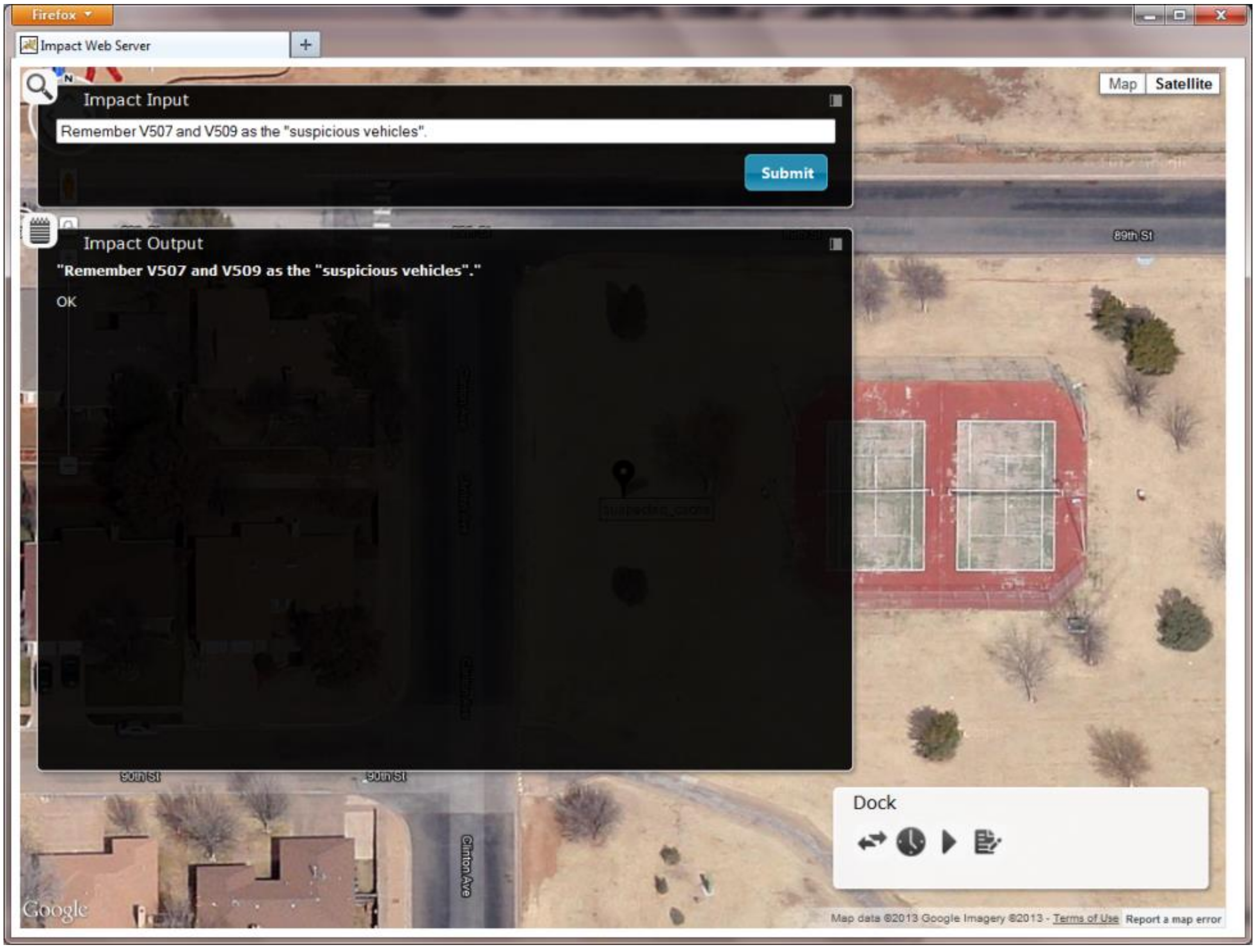


Summarize the activity.

Impact Output

```
V509 makes a U-turn from 09:44:29.572 to 09:44:36.888.
V509 makes a stop at (-101.930802, 033.498198)× from 09:44:38.013 to 09:53:14.622.
V509 meets with V507 from 09:44:38.013 to 09:53:14.059.
V258 approaches V509 from 09:52:07.654 to 09:52:16.095.
V258 turns left from 09:52:13.844 to 09:52:19.472.
V258 travels east from 09:52:20.035 to 09:52:25.662.
V258 turns right from 09:52:26.225 to 09:52:29.601.
V258 turns left from 09:52:30.164 to 09:52:38.043.
V258 approaches V509 from 09:52:38.043 to 09:52:48.735.
V258 turns right from 09:52:47.610 to 09:52:54.925.
V258 accelerates to 09.8 m/s from 09:52:51.549 to 09:52:57.739.
V507 accelerates to 09.1 m/s from 09:53:14.622 to 09:53:20.812.
V509 accelerates to 07.9 m/s from 09:53:15.184 to 09:53:18.561.
V509 accelerates to 11.0 m/s from 09:53:19.687 to 09:53:24.189.
V509 does not turn from 09:53:20.812 to 09:53:24.751.
V507 accelerates to 14.2 m/s from 09:53:22.500 to 09:53:24.751.
```

At this point, the user would like to be able to track the activities of both V509 and V507. The user assigns a name to these two vehicles: the “suspicious vehicles”.

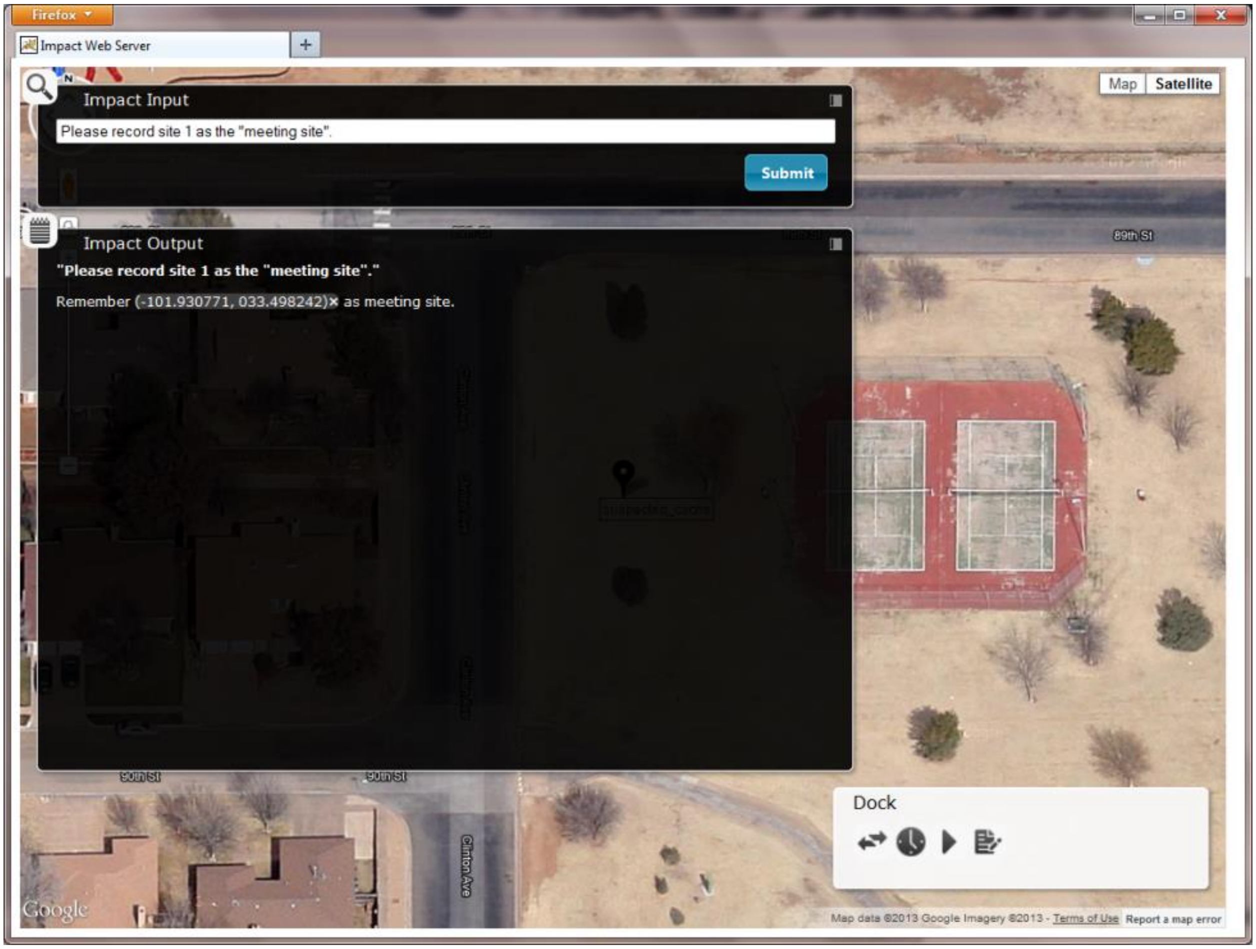


Remember V507 and V509 as the "suspicious vehicles".

"Remember V507 and V509 as the "suspicious vehicles"."

OK

The user also creates a more informative name for “site 1”, calling this position the “meeting site”.

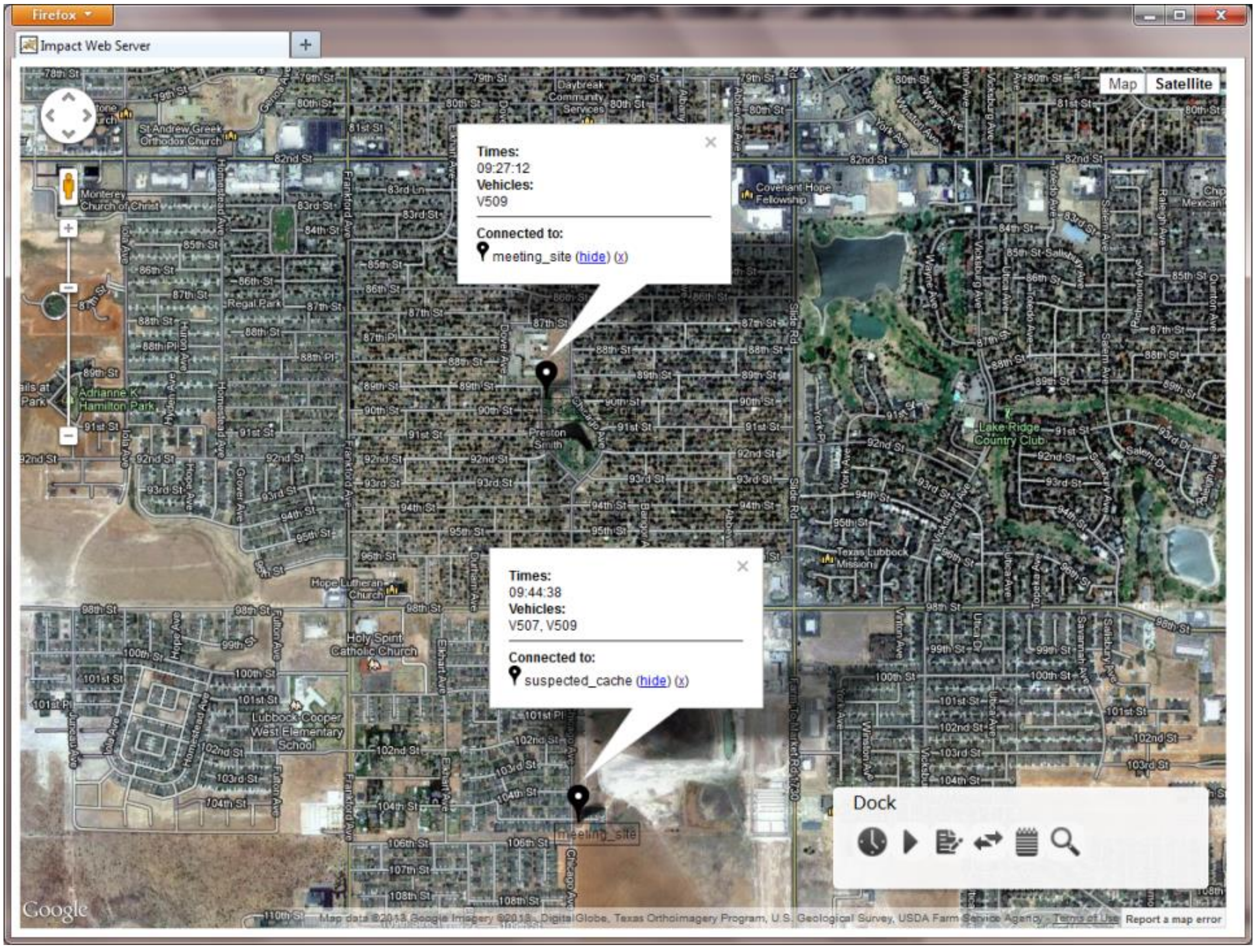


Please record site 1 as the "meeting site".

"Please record site 1 as the "meeting site"."

Remember (-101.930771, 033.498242)× as meeting site.

The user then submits requests, omitted here, instructing the system to associate the suspicious vehicles and the beginning time of their meeting with the "meeting site" and to associate the "suspected cache" with the "meeting site". The user then asks the system to display an updated version of its activity graph, containing the two locations and their associated vehicles and times.



Firefox
Impact Web Server
Map
Satellite
meeting_site
Dock
St Andrew Greek Orthodox Church
Monterey Church of Christ
Covenant Hope Fellowship
Daybreak Community Services
Adrianne K. Hamilton Park
Preston Smith
Lake Ridge Country Club
Texas Lubbock Mission
Hope Lutheran Church
Holy Spirit Catholic Church
Lubbock-Cooper West Elementary School
Frankford Ave
Slide Rd
Google
Map data ©2013 Google Imagery ©2013 , DigitalGlobe, Texas Orthoimagery Program, U.S. Geological Survey, USDA Farm Service Agency - Terms of Use Report a map error

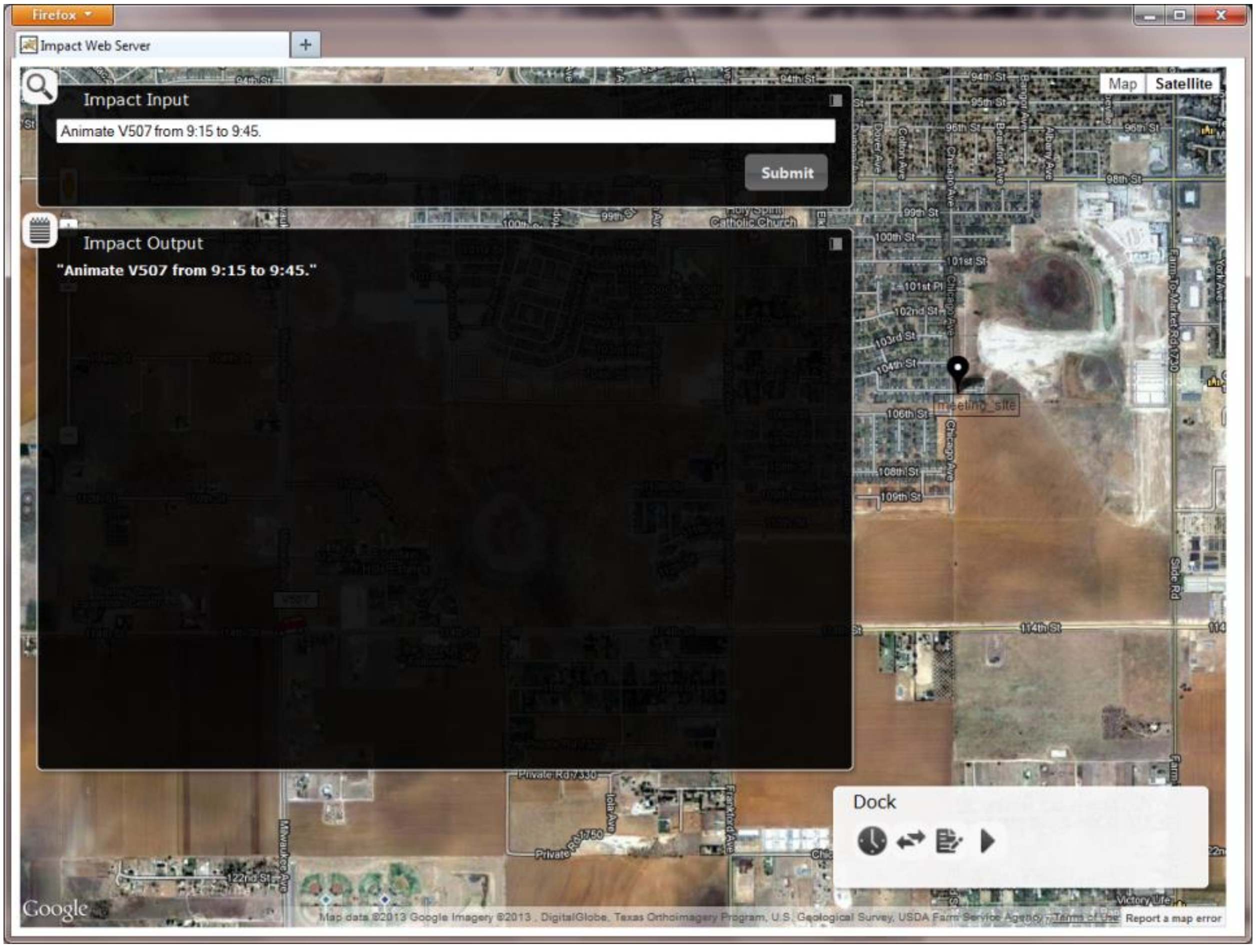
Times:
09:27:12
Vehicles:
V509

Connected to:
meeting_site (hide) (x)

Times:
09:44:38
Vehicles:
V507, V509

Connected to:
suspected_cache (hide) (x)

At this point, the user is aware that V507 and V509 have engaged in a meeting and that V509 has traveled to the meeting site from the suspected cache location. The user does not know what V507 has done prior to the meeting, however. The user could ask the system to focus on V507's activity over a time period preceding the meeting, then summarize events that the system has observed. Alternatively, the user can inspect the data stream directly by asking the system to display an animation of V507's positions over a period of time. The user does this, requesting an animation over a 30-minute time window ending just after V507 has arrived at the meeting site.



Impact Input

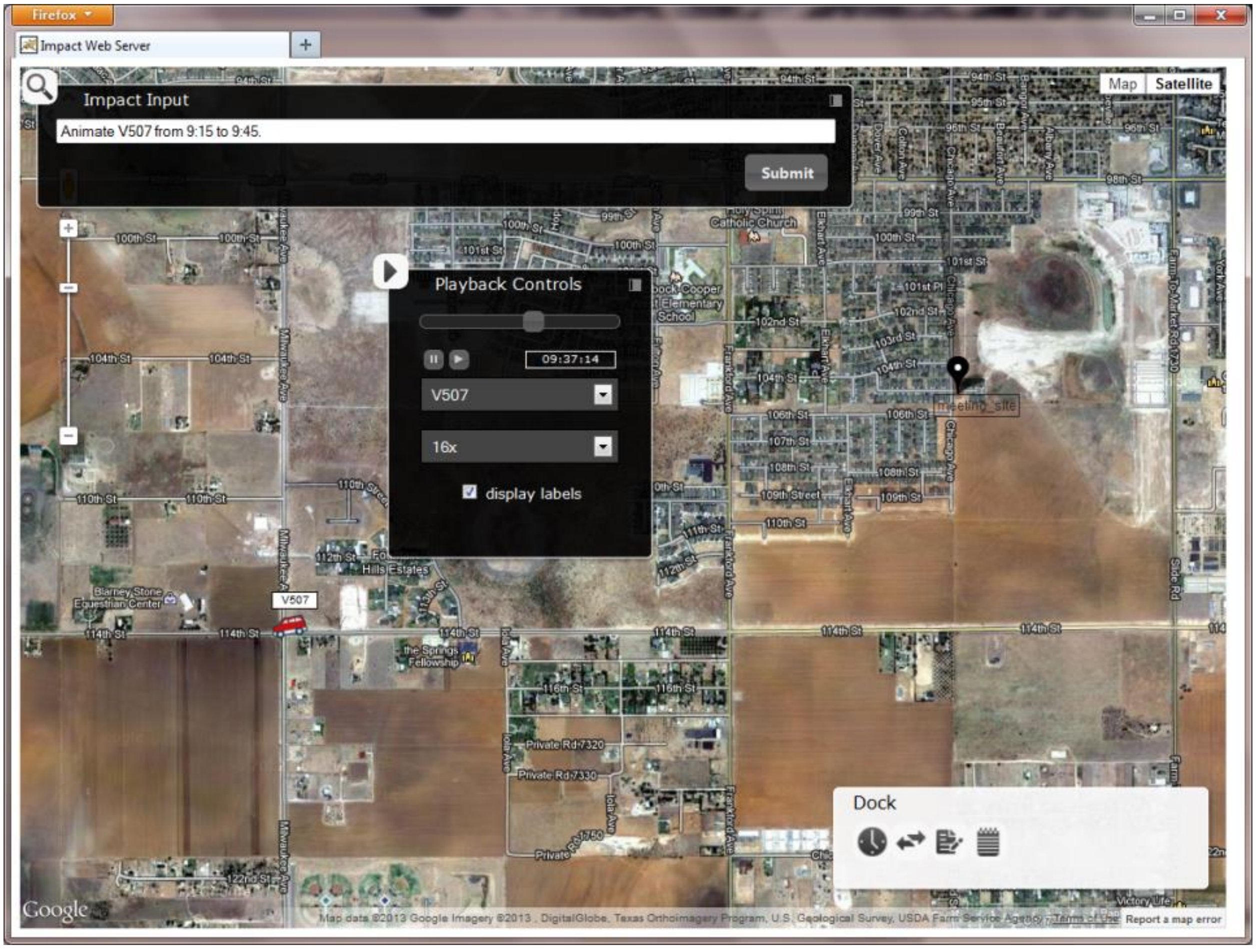
Animate V507 from 9:15 to 9:45.

Impact Output

"Animate V507 from 9:15 to 9:45."

Firefox
Impact Web Server
Map
Satellite
Submit
meeting_site
Dock
Google
Map data ©2013 Google Imagery ©2013, DigitalGlobe, Texas Orthoimagery Program, U.S. Geological Survey, USDA Farm Service Agency · Terms of Use
Report a map error

In the displayed animation, V507 is shown to be stationary at a location from the beginning of the recorded data segment, approximately 9:27 AM, to approximately 9:39 AM.



Impact Input

Animate V507 from 9:15 to 9:45.

Submit

Playback Controls

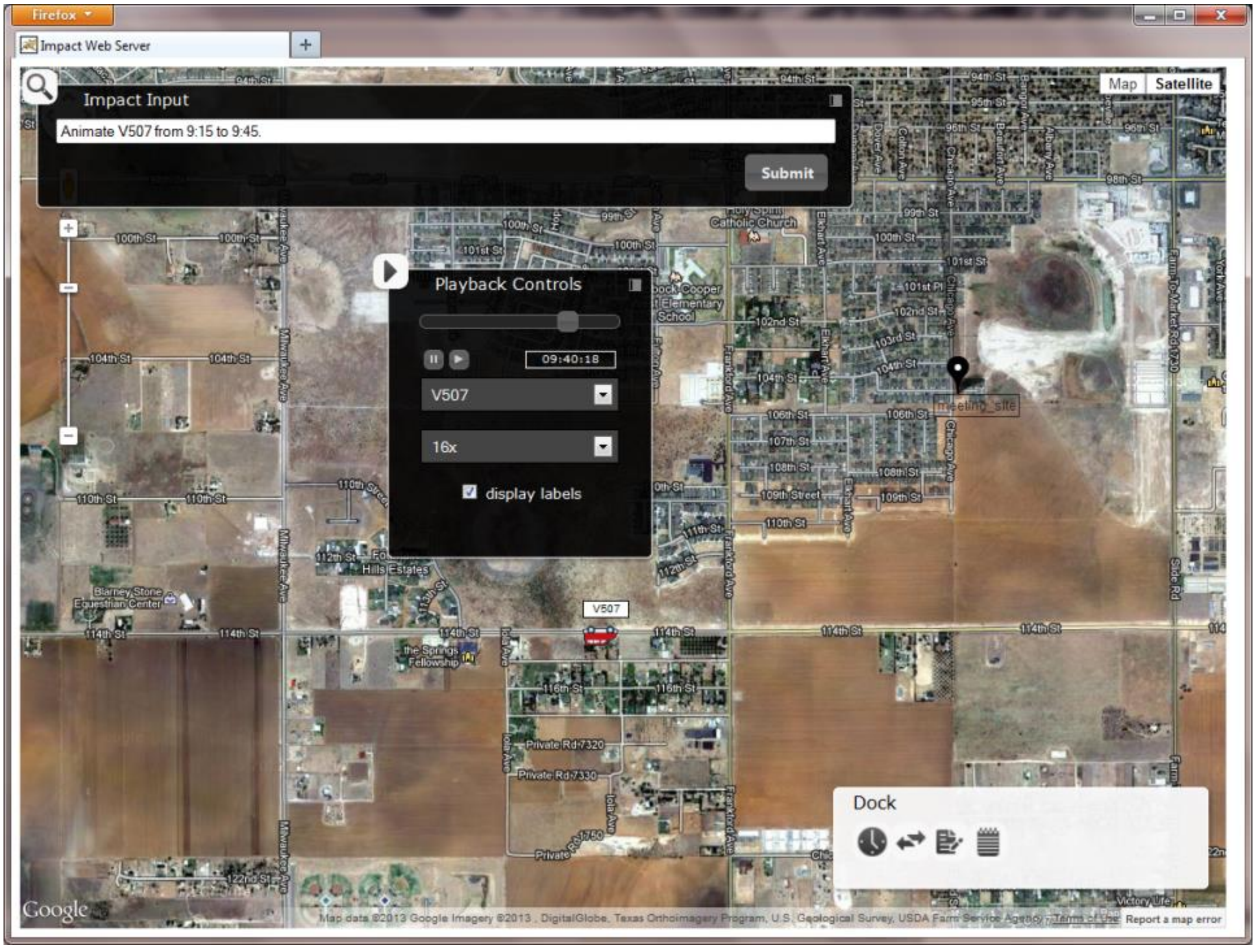
09:37:14

V507

16x

☑ display labels

V507 then proceeds eastbound on 114th Street, then maneuvers through several other streets before reaching the meeting site.



Impact Input

Animate V507 from 9:15 to 9:45.

Submit

Playback Controls

09:40:18

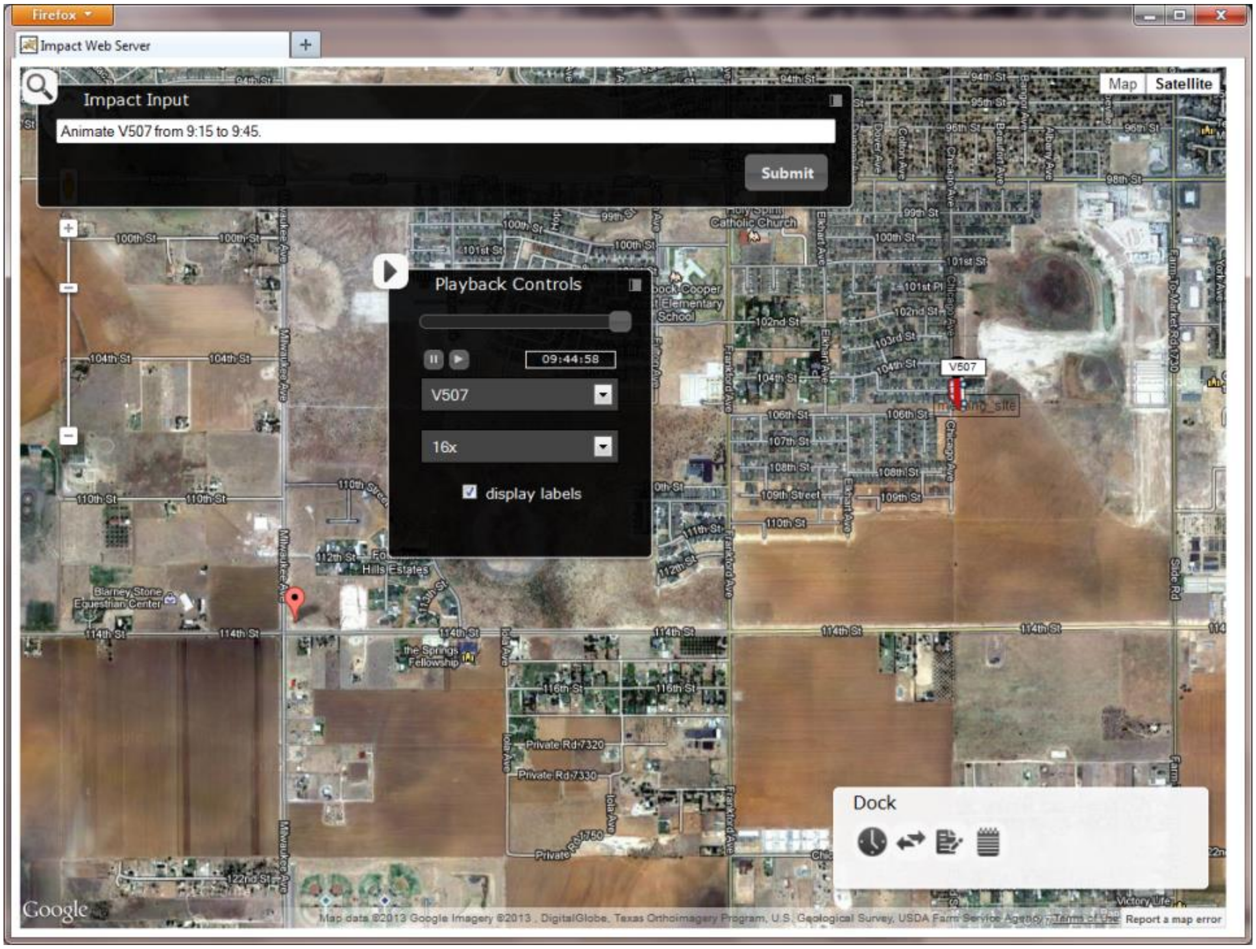
V507

16x

☑ display labels

Dock

The user clicks on V507's initial position in this animation, creating a graphical marker at this point and enabling the position to be referenced in natural language requests.



Impact Input

Animate V507 from 9:15 to 9:45.

Playback Controls

09:44:58

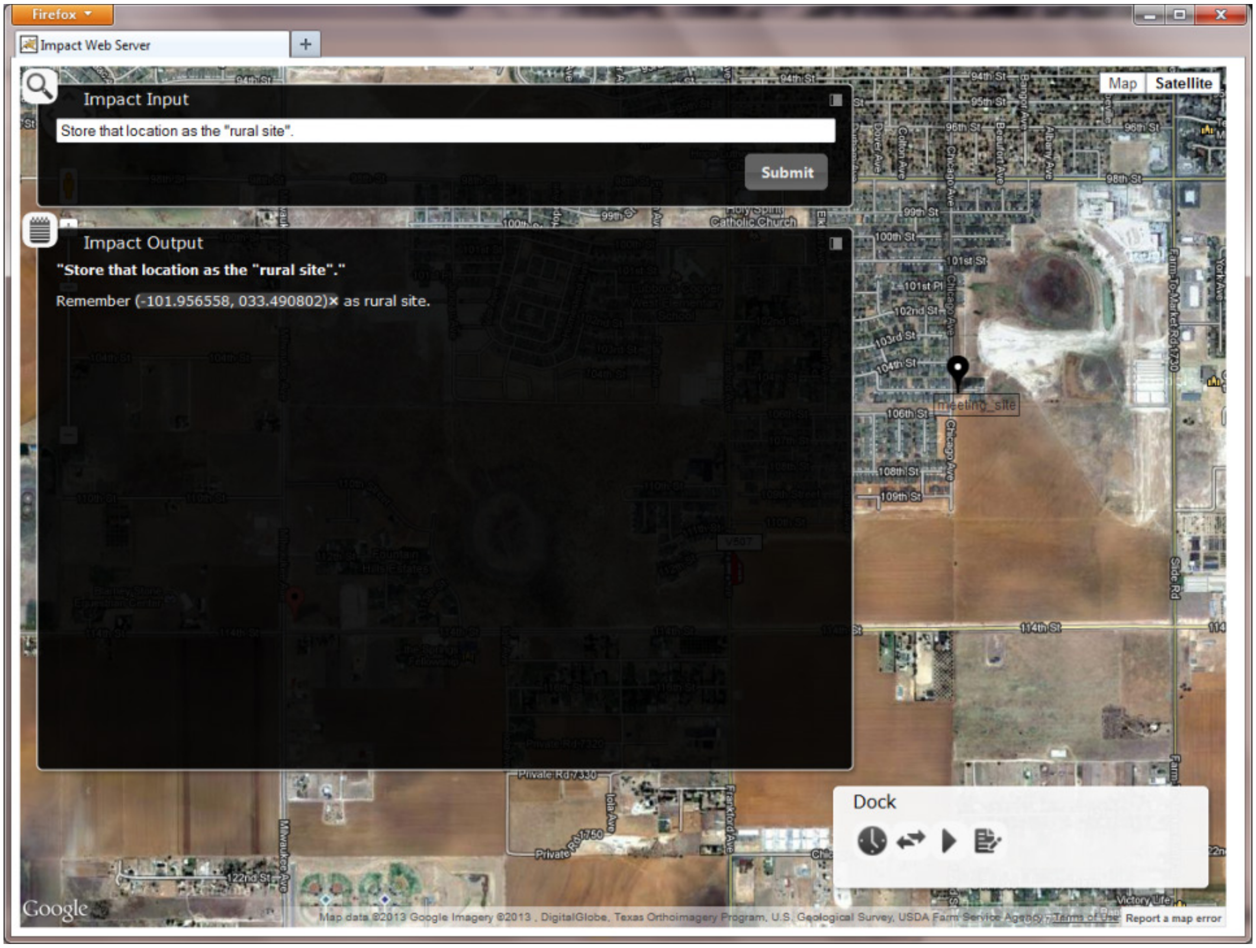
V507

16x

☑ display labels

Firefox
Impact Web Server
Map
Satellite
Submit
V507
Dock
Google
Map data ©2013 Google Imagery ©2013, DigitalGlobe, Texas Orthoimagery Program, U.S. Geological Survey, USDA Farm Service Agency - Terms of Use Report a map error

The user then asks the system to name this new location the “rural site”.



Impact Input

Store that location as the "rural site".

Impact Output

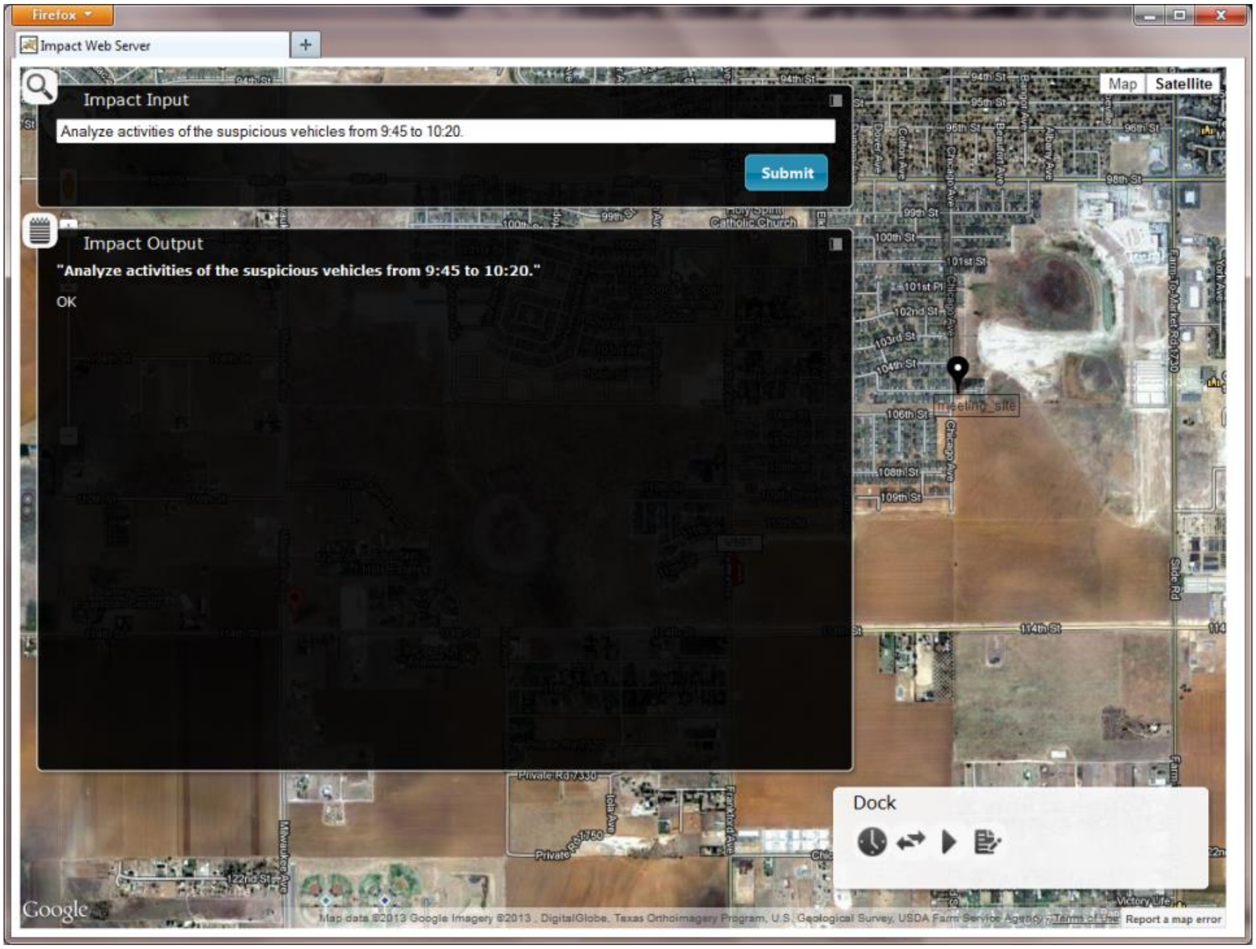
"Store that location as the "rural site"."

Remember (-101.956558, 033.490802)× as rural site.

Firefox
Impact Web Server
Submit
Map
Satellite
meeting_site
Dock
Google
Map data ©2013 Google Imagery ©2013 , DigitalGlobe, Texas Orthoimagery Program, U.S. Geological Survey, USDA Farm Service Agency - Terms of Use
Report a map error

The user then includes the rural site in the activity graph, associated with vehicle V507, the time of interest, and with the meeting site.

At this point, the user returns to an interpretation of unfolding events, asking the system to analyze the two suspicious vehicles, V509 and V507, over a period going forward from their encounter at the "meeting site", until 10:20 AM.

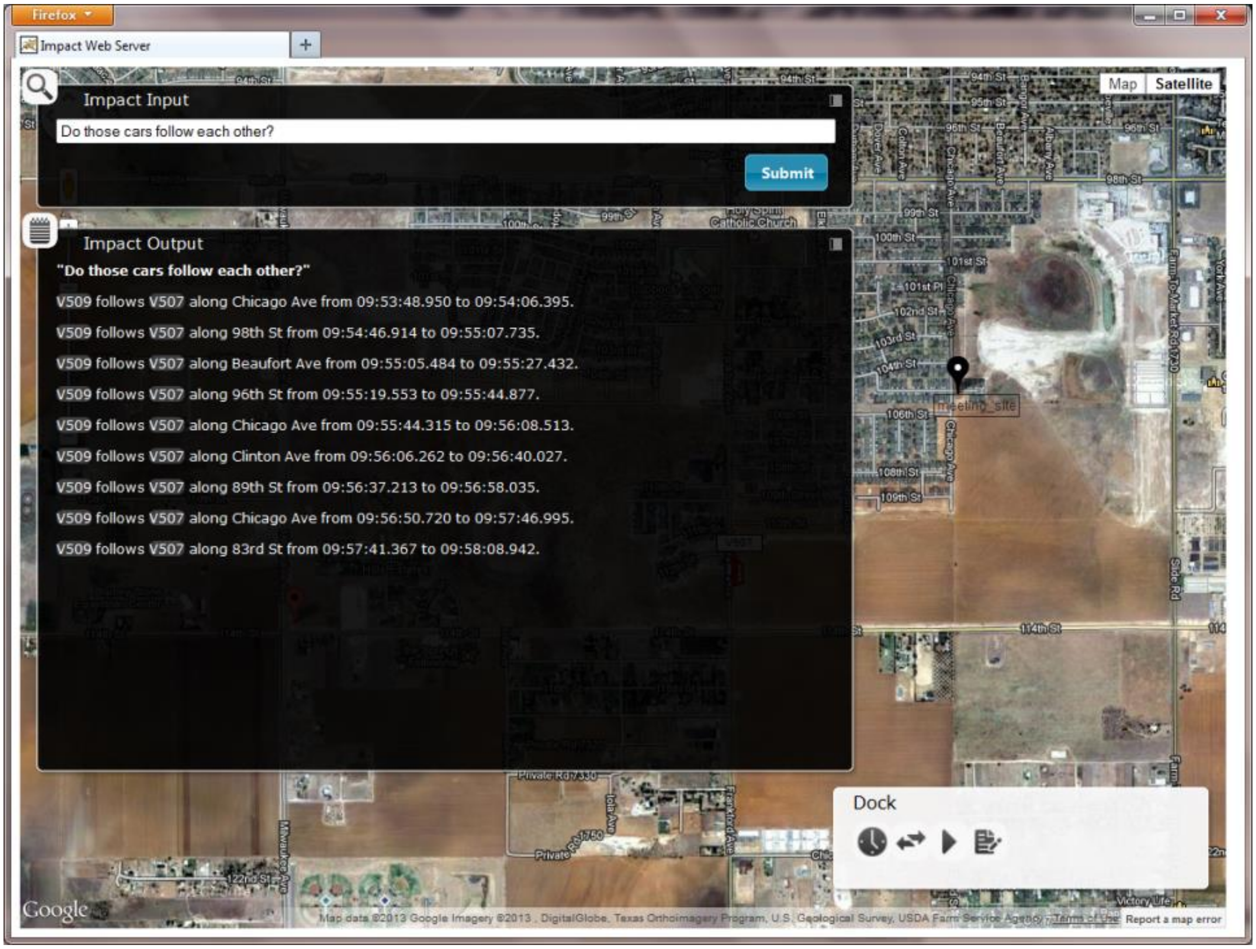


Analyze activities of the suspicious vehicles from 9:45 to 10:20.

"Analyze activities of the suspicious vehicles from 9:45 to 10:20."

OK

Once again, the user asks the system to recognize a specific, targeted type of event, rather than summarize all events. In this case, the user asks whether either of the suspicious cars follows the other car during the focus interval. The system responds that V509 follows V507 along a series of 9 street segments from 9:53 AM to 9:58 AM.

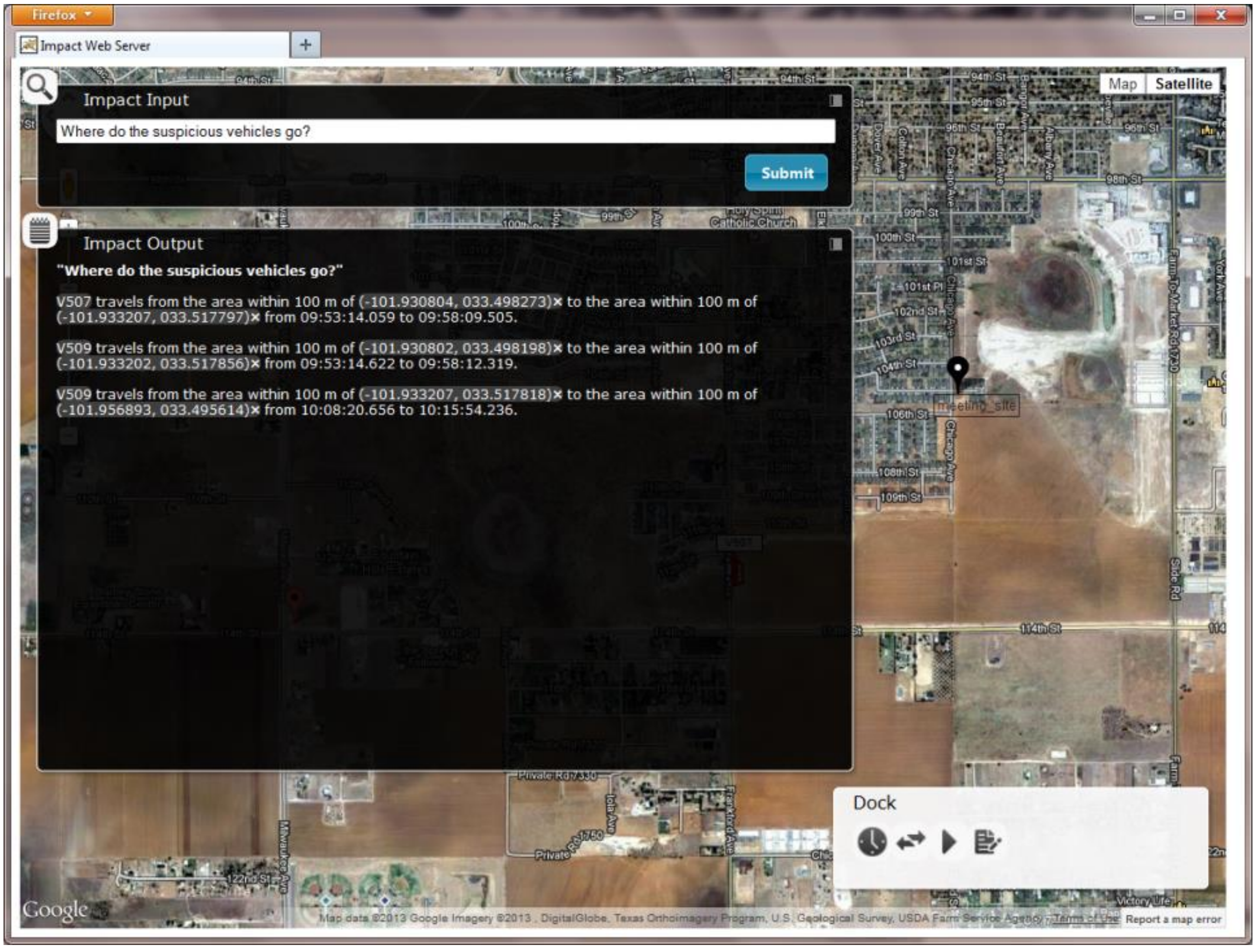


Do those cars follow each other?

"Do those cars follow each other?"

```
V509 follows V507 along Chicago Ave from 09:53:48.950 to 09:54:06.395.
V509 follows V507 along 98th St from 09:54:46.914 to 09:55:07.735.
V509 follows V507 along Beaufort Ave from 09:55:05.484 to 09:55:27.432.
V509 follows V507 along 96th St from 09:55:19.553 to 09:55:44.877.
V509 follows V507 along Chicago Ave from 09:55:44.315 to 09:56:08.513.
V509 follows V507 along Clinton Ave from 09:56:06.262 to 09:56:40.027.
V509 follows V507 along 89th St from 09:56:37.213 to 09:56:58.035.
V509 follows V507 along Chicago Ave from 09:56:50.720 to 09:57:46.995.
V509 follows V507 along 83rd St from 09:57:41.367 to 09:58:08.942.
```

The user then continues with a question about where the suspicious vehicles are going. The system responds that both vehicles travel from their positions near the meeting site to destinations close to one another, arriving around 9:58 AM, and that V509 subsequently travels onward from that location to another location, beginning at 10:08 AM and ending at about 10:16 AM.



Where do the suspicious vehicles go?

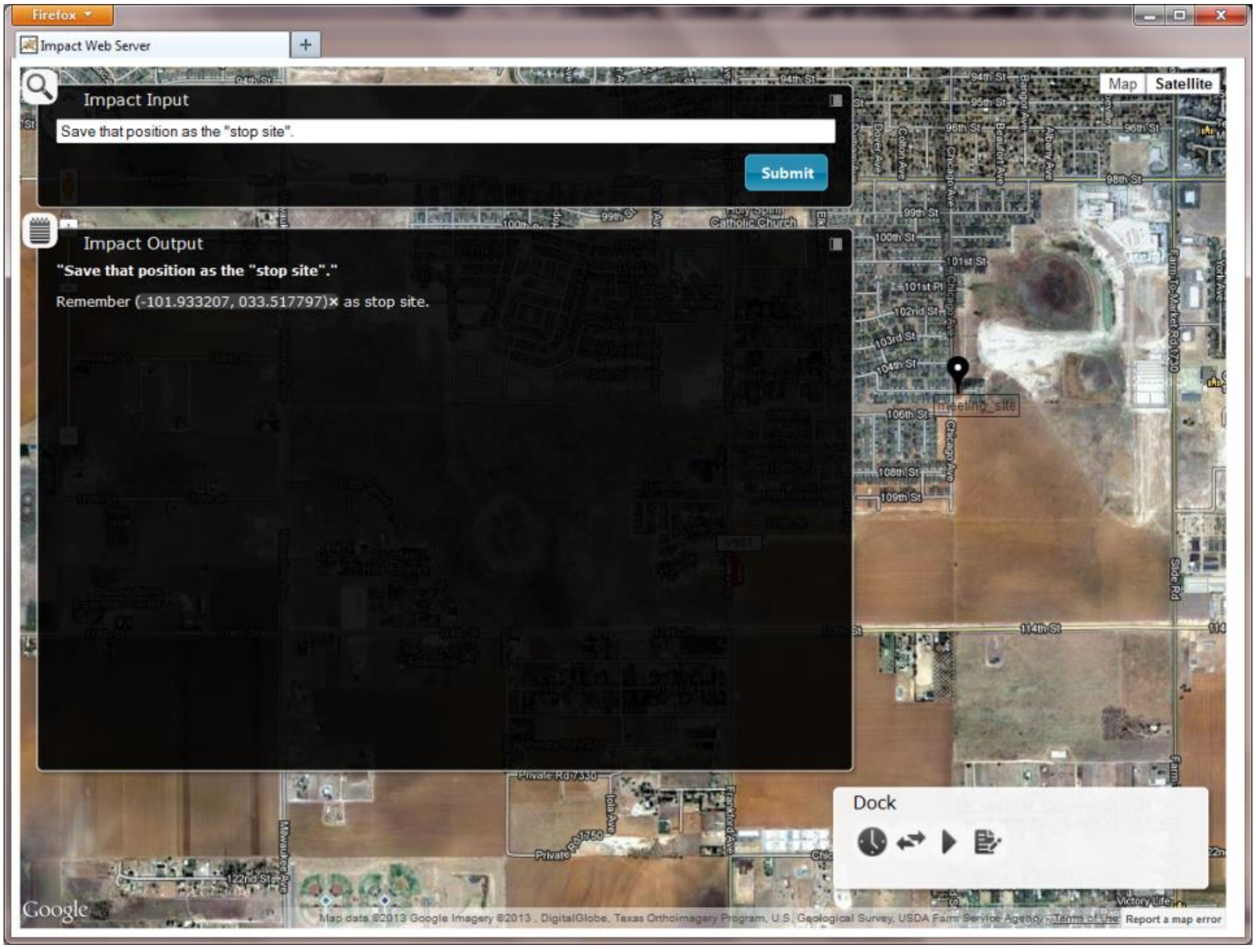
"Where do the suspicious vehicles go?"

V507 travels from the area within 100 m of (-101.930804, 033.498273)× to the area within 100 m of (-101.933207, 033.517797)× from 09:53:14.059 to 09:58:09.505.

V509 travels from the area within 100 m of (-101.930802, 033.498198)× to the area within 100 m of (-101.933202, 033.517856)× from 09:53:14.622 to 09:58:12.319.

V509 travels from the area within 100 m of (-101.933207, 033.517818)× to the area within 100 m of (-101.956893, 033.495614)× from 10:08:20.656 to 10:15:54.236.

The user then selects the destination of V507 during the joint V507–V509 transit and asks the system to remember this location as the “stop site”.



Impact Input

Save that position as the "stop site".

Submit

Impact Output

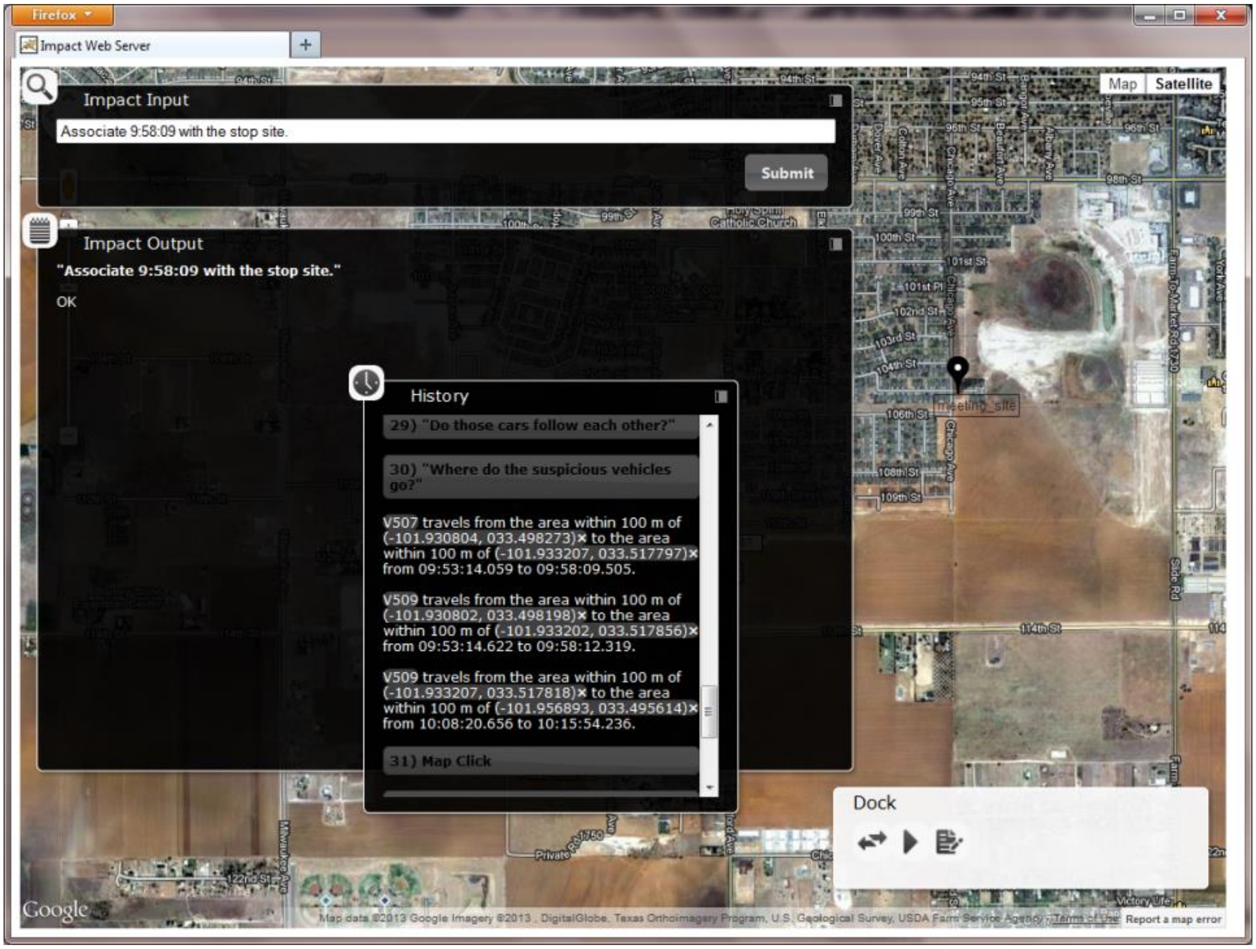
"Save that position as the "stop site"."

Remember (-101.933207, 033.517797)× as stop site.

Dock

The stop site is then linked to the meeting site and also associated with the two suspicious vehicles and their time of arrival at the stop site.

The user also wants to note the final destination of V509. To do this, the user opens a history pane in the user interface, navigates to the recent request–response pair concerning travel by V509 and V507, and selects the second destination of V509 from the recorded response.



Impact Input: Associate 9:58:09 with the stop site.

Impact Output:

"Associate 9:58:09 with the stop site."

OK

History:

29) "Do those cars follow each other?"

30) "Where do the suspicious vehicles go?"

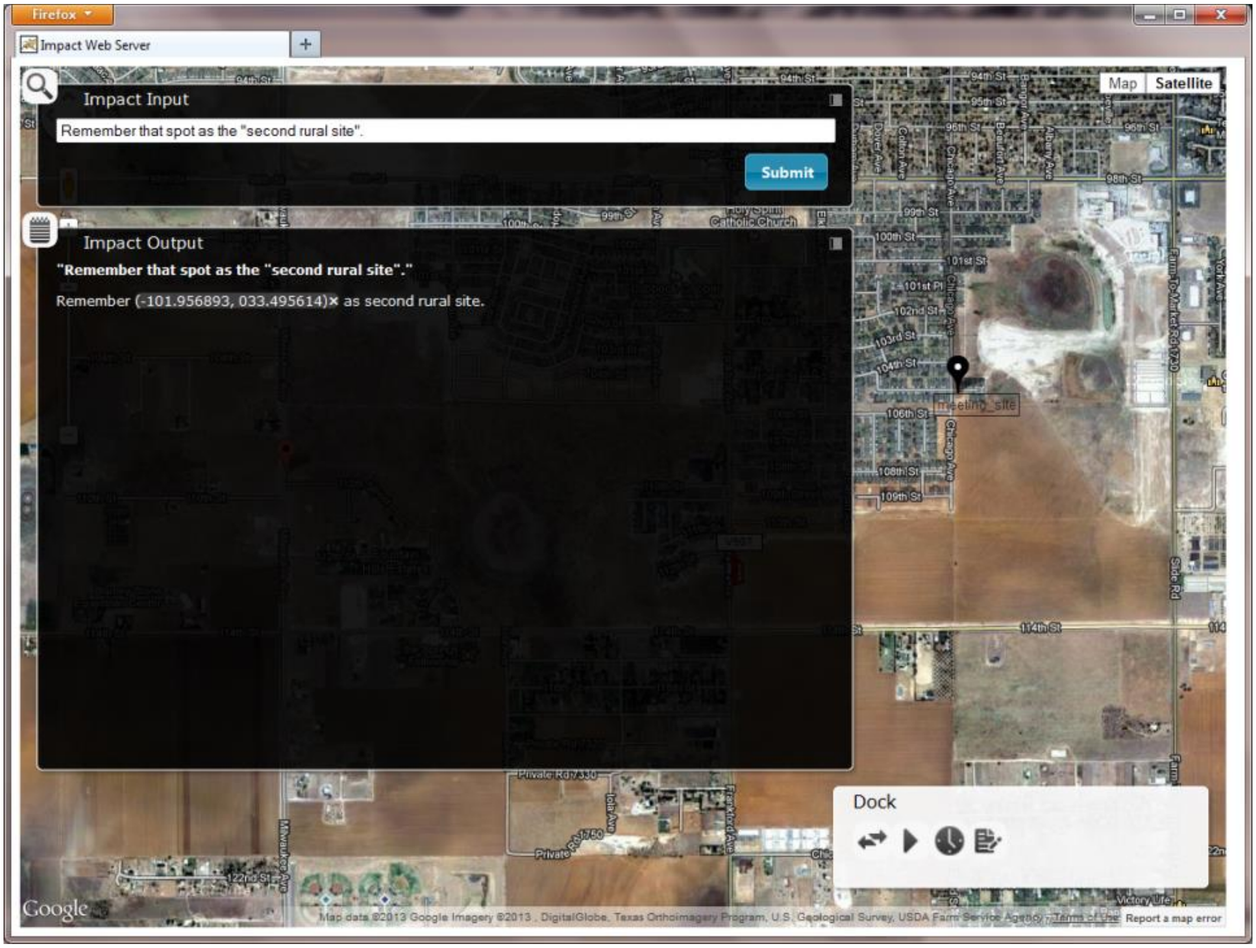
V507 travels from the area within 100 m of (-101.930804, 033.498273)× to the area within 100 m of (-101.933207, 033.517797)× from 09:53:14.059 to 09:58:09.505.

V509 travels from the area within 100 m of (-101.930802, 033.498198)× to the area within 100 m of (-101.933202, 033.517856)× from 09:53:14.622 to 09:58:12.319.

V509 travels from the area within 100 m of (-101.933207, 033.517818)× to the area within 100 m of (-101.956893, 033.495614)× from 10:08:20.656 to 10:15:54.236.

31) Map Click

The user then asks the system to remember this location as the “second rural site”.



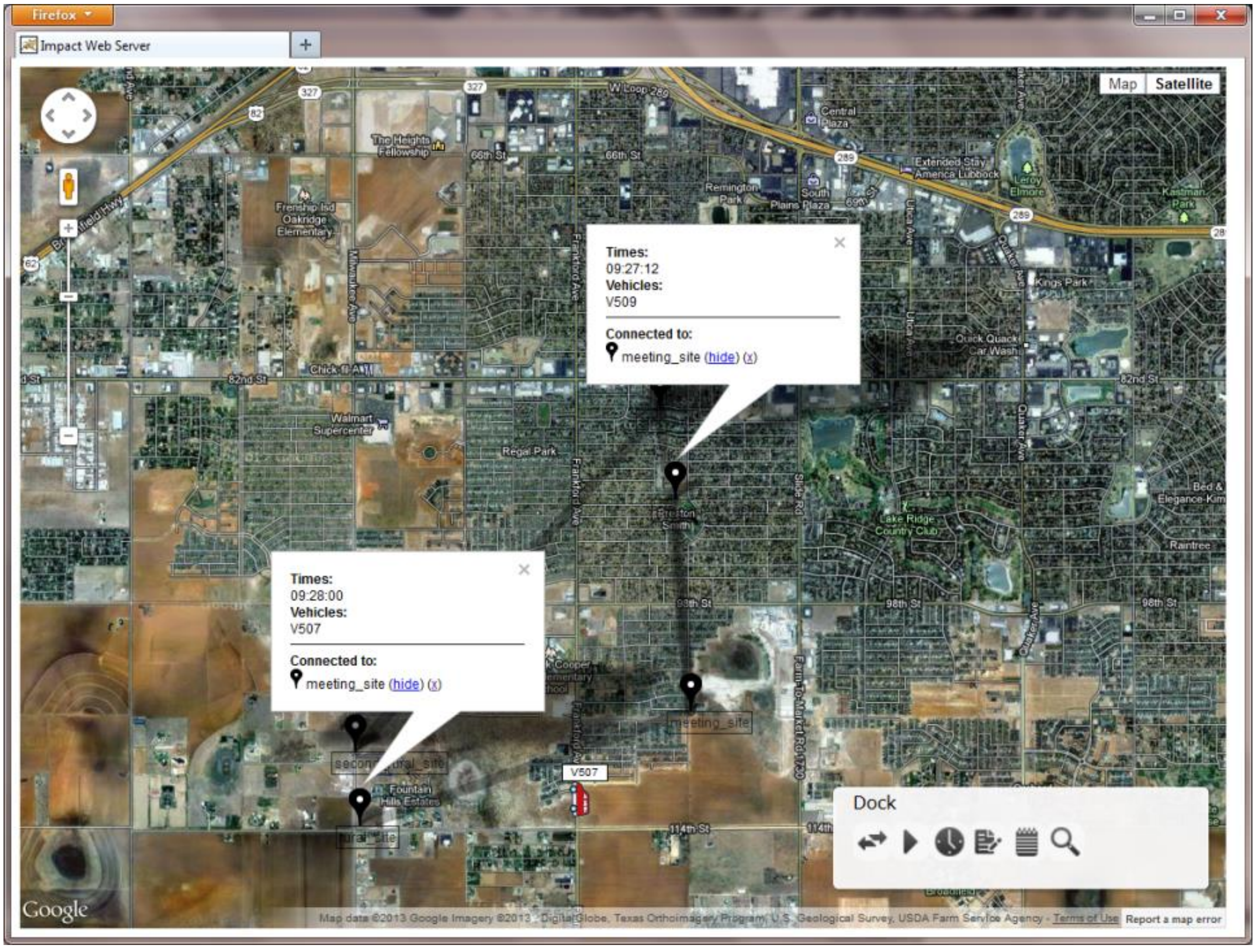
Remember that spot as the "second rural site".

"Remember that spot as the "second rural site"."

Remember (-101.956893, 033.495614)× as second rural site.

After submitting requests to link the second rural site to the stop site and also to associate the second rural site with V509 and the time of V509's arrival, the user asks the system to once again display its activity map.

Reconstructing the activity, the user observes that V507 and V509 begin at different locations, with V509 stopped near the possible cache of the original alert.


Firefox
Impact Web Server
Map
Satellite
meeting_site
secon ural_site
rural_site
V507
Dock
Google
Map data ©2013 Google Imagery ©2013 , DigitalGlobe, Texas Orthoimagery Program, U.S. Geological Survey, USDA Farm Service Agency - Terms of Use
Report a map error

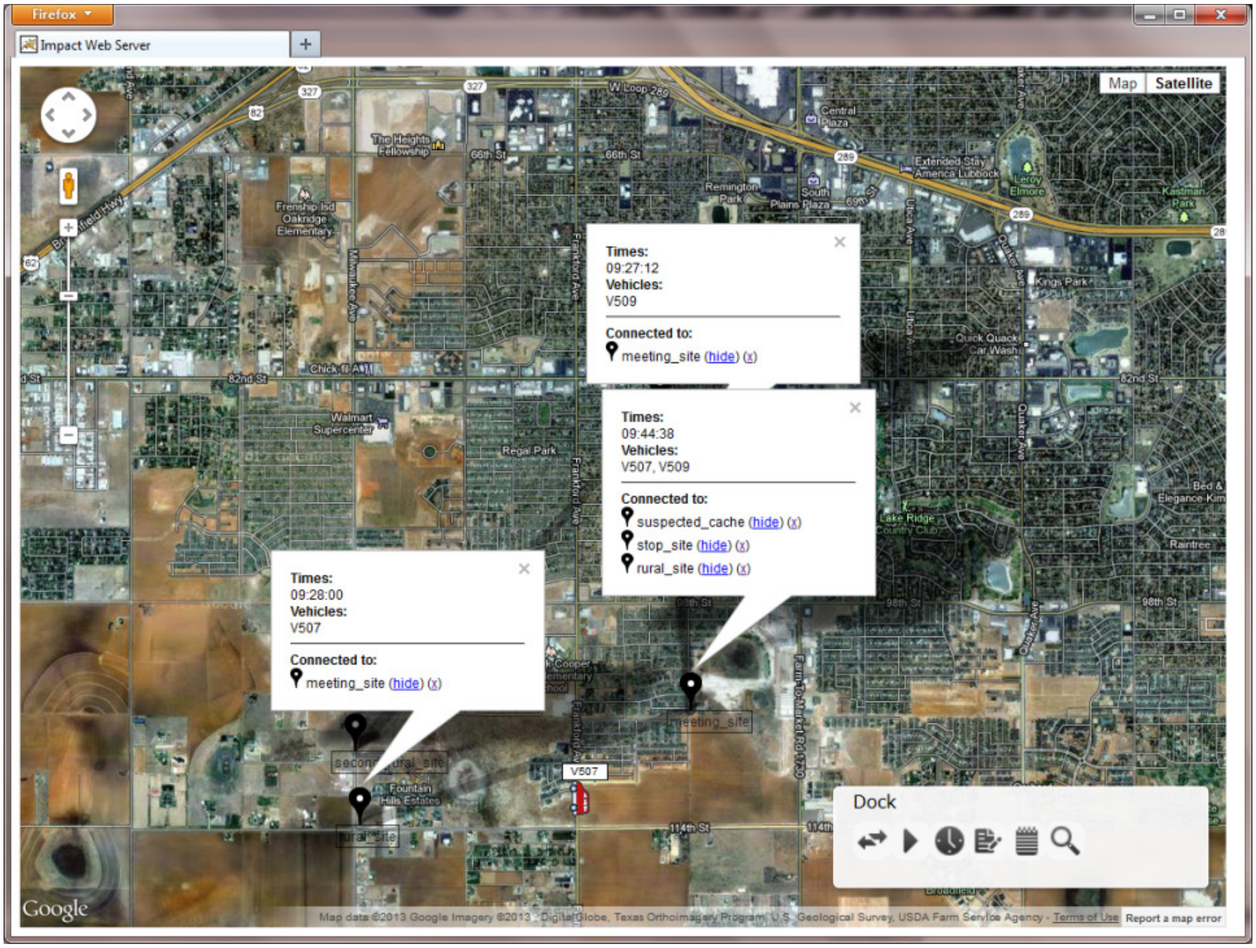
Times:
09:27:12
Vehicles:
V509

Connected to:
meeting_site (hide) (x)

Times:
09:28:00
Vehicles:
V507

Connected to:
meeting_site (hide) (x)

Both vehicles then proceed to the meeting site, arriving around 9:44 AM.



Times:
09:27:12
Vehicles:
V509

Connected to:
meeting_site (hide) (x)

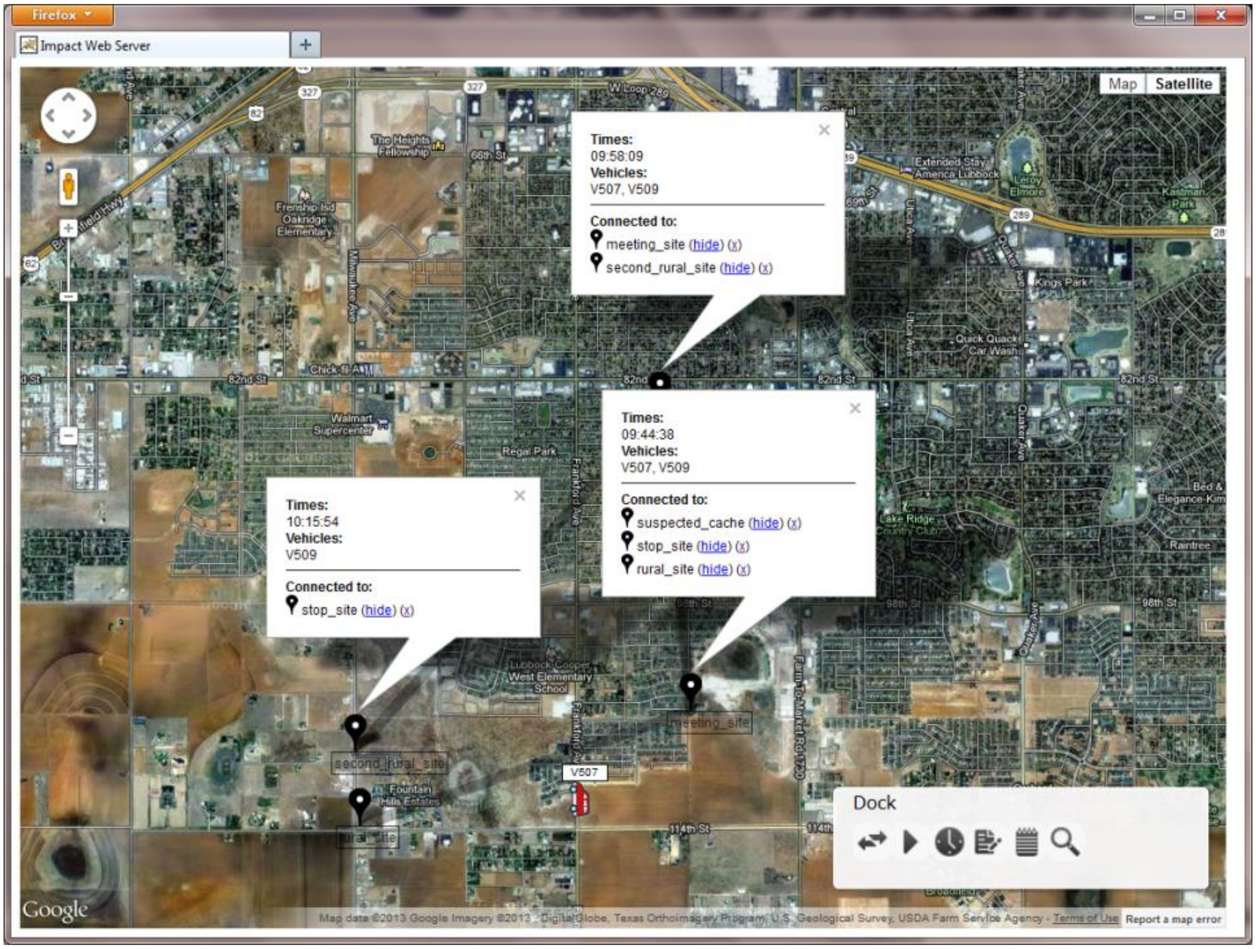
Times:
09:44:38
Vehicles:
V507, V509

Connected to:
suspected_cache (hide) (x)
stop_site (hide) (x)
rural_site (hide) (x)

Times:
09:28:00
Vehicles:
V507

Connected to:
meeting_site (hide) (x)

From the meeting site, V509 follows V507 to a second site, the “stop site”, arriving at about 9:58 AM. From the stop site, V509 progresses to the second rural site, arriving approximately 10:16 AM.



Times:
09:58:09
Vehicles:
V507, V509

Connected to:
meeting_site (hide) (x)
second_rural_site (hide) (x)

Times:
09:44:38
Vehicles:
V507, V509

Connected to:
suspected_cache (hide) (x)
stop_site (hide) (x)
rural_site (hide) (x)

Times:
10:15:54
Vehicles:
V509

Connected to:
stop_site (hide) (x)

At this stage in the analysis, the user has formed an interpretation of the overall activity in which V509 may have retrieved an Improvised Explosive Device or explosive-making materials at the cache, met with V507, progressed to a new site, then progressed to a site likely to be the location of IED emplacement, given that V507 has recently lingered at a site close to that position, prior to meeting with V509. The user may then issue an alert so that the potential IED attack can be averted before it can be carried through to completion.

References

Borchardt, G. C. (1992). “Understanding Causal Descriptions of Physical Systems.” In *Proceedings of the AAAI Tenth National Conference on Artificial Intelligence*, 2–8.

Borchardt, G. C. (1994). *Thinking between the Lines: Computers and the Comprehension of Causal Descriptions*, MIT Press.

Borchardt, G. C. (2014). *Möbius Language Reference, Version 1.2*, Technical Report MIT-CSAIL-TR-2014-005, MIT Computer Science and Artificial Intelligence Laboratory.

Burns, J. B., Connolly, C. I., Thomere, J. F., and Wolverton, M. J. (2006). "Event Recognition in Airborne Motion Imagery." In *Capturing and Using Patterns for Evidence Detection, Papers from the AAAI Fall Symposium*, AAAI Press.

Buxton, H. (2003). "Learning and Understanding Dynamic Scene Activity: A Review." In *Image and Vision Computing*, Vol. 21, 125–136.

Grice, H. P. (1975). "Logic and Conversation." In Cole, P. and Morgan, J. L. (eds.), *Syntax and Semantics, Volume 3: Speech Acts*, Academic Press, 41–58.

Hu, W., Tan, T., Wang, L., and Maybank, S. (2004). "A Survey on Visual Surveillance of Object Motion and Behaviors." In *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 34, No. 3, 334–352.

Katz, B. (1990). "Using English for Indexing and Retrieving." In *Artificial Intelligence at MIT: Expanding Frontiers, Vol. 1*, Cambridge, Massachusetts, 134–165.

Katz, B. (1997). "Annotating the World Wide Web Using Natural Language." In *Proceedings of the 5th RIAO Conference on Computer Assisted Information Searching on the Internet (RIAO '97)*, Montreal, Canada, 136–155.

Katz, B., Borchardt, G., and Felshin, S. (2005). "Syntactic and Semantic Decomposition Strategies for Question Answering from Multiple Resources." In *Proceedings of the AAAI 2005 Workshop on Inference for Textual Question Answering*, 35–41.

Katz, B., Borchardt, G., and Felshin, S. (2006). "Natural Language Annotations for Question Answering." In *Proceedings of the 19th International FLAIRS Conference (FLAIRS 2006)*, Melbourne Beach, Florida, 303–306.

Katz, B., Borchardt, G., Felshin, S., and Mora, F. (2007). "Harnessing Language in Mobile Environments." In *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)*, 421–428.

Katz, B. and Levin, B. (1988). "Exploiting Lexical Regularities in Designing Natural Language Systems." In *Proceedings of the 12th International Conference on Computational Linguistics (COLING '88)*, Budapest, Hungary.

Lavee, G., Rivlin, E., and Rudzsky, M. (2009). "Understanding Video Events: A Survey of Methods for Automatic Interpretation of Semantic Occurrences in Video." In *IEEE Transactions on Systems, Man, and Cybernetics — Part C: Applications and Reviews*, Vol. 39, No. 5, 489–504.

Miller, G. A. and Johnson-Laird, P. N. (1976). *Language and Perception*, Harvard University Press.

Narayanaswamy, S., Barbu, A., and Siskind, J. M. (2014). "Seeing What You're Told: Sentence-Guided Activity Recognition in Video." In *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)*, to appear.

Porter, R., Ruggiero, C., and Morrison, J.D. (2009). "A Framework for Activity Detection in Wide-Area Motion Imagery." In Rahman, Z.-U., Reichenbach, S. E., and Neifeld, M. A. (eds.), *Proceedings Vol. 7341, Visual Information Processing XVIII*.

Siskind, J. M. (2003). "Reconstructing Force-Dynamic Models from Video Sequences." In *Artificial Intelligence*, 151, 91–154.

Turaga, P., Chellappa, R., Subrahmanian, V.S., and Udrea, O. (2008). "Machine Recognition of Human Activities: A Survey." In *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 18, No. 11, 1473–1488.