Using Semantic Technologies to Improve Information Exploitation in Military and Civilian Application Contexts

Paul R. Smart[1*], Alistair Russell[1], Shao Fen Liang[1], Nigel R. Shadbolt[1], Chris Booth[2], Neil Briscombe[2] and Andrew Rankin[2]

[1]*School of Electronics and Computing Science, University of Southampton, Southampton, SO17 1BJ, UK*
[2]*Malvern Technology Centre, St. Andrews Road, Malvern, Worcestershire, WR14 3PS, UK*

[*]*Corresponding author: ps02v@ecs.soton.ac.uk*

Abstract—*Military and civilian agencies confront a number of challenges in harnessing the power of modern large-scale information networks. These challenges include the need to identify relevant information and exploit that information in the context of specific knowledge processing tasks. Our work in the SEMIOTIKS component of the UK Data and Information Fusion Defence Technology Centre (DIFDTC) research programme aims to support military and civilian agencies in leveraging the latent potential of large-scale information networks to enhance situation awareness and improve knowledge processing efficiency. The SEMIOTIKS system, described in this paper, comprises a number of technology components that support the user in a variety of knowledge-based activities. These include the discovery and organization of information resources; the extraction of task-relevant information using information harvesting and knowledge editing techniques; the retrieval, integration and transformation of domain-relevant knowledge using semantic query capabilities; the monitoring of large-scale information repositories for specific events and contingencies; and the use of semantic reasoning techniques to support domain-relevant decision-making. In addition to a technical description of SEMIOTIKS system components, this paper describes a demonstration scenario that highlights the way in which SEMIOTIKS technologies might be used to support advanced modes of information exploitation in multi-agency collaborative situations.*

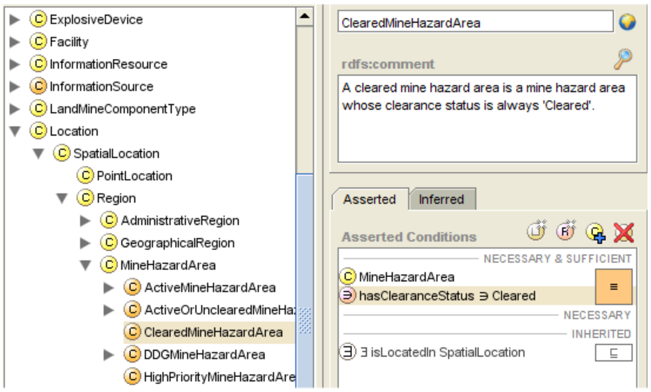
1. INTRODUCTION

Military and civilian agencies confront a number of challenges in harnessing the power of modern large-scale information networks. One challenge relates to the need to identify relevant information resources in large-scale (and often distributed) information repositories. Another challenge relates to the need to rapidly retrieve task-relevant information in a manner that aligns itself with the knowledge processing objectives of the end user. This often requires the extraction and restructuring of information content, perhaps from an unstructured textual format to a machine-processable format, such as RDF/XML. Finally, military and civilian agencies need to be able to utilize large quantities of information, typically as part of some ongoing problem-solving activity, e.g. military planning. The challenge here relates to the use of technologies that can efficiently process large quantities of information in order to support the realization of operationally-effective decision outcomes.

Our work in the SEMIOTIKS component of the UK DIFDTC research programme aims to use a number of techniques and technologies to support end users with respect to the exploitation of large-scale information repositories. Our approach is grounded in the development of an integrated suite of technology components that assists users with the discovery, retrieval and processing of task-relevant information. The technology components include a Resource Discovery System (see Section 3.1), which identifies resources of potential relevance to an end user using resource processing techniques; a Resource Browsing System (see Section 3.2), which enables users to organize and navigate the contents of a resource repository; a Knowledge Acquisition System (see Section 3.3), which extracts the information content of resources; an Information Retrieval System (see Section 3.4), which supports information retrieval via a graphical semantic query interface; a Knowledge Monitoring System (see Section 3.5), which alerts end users to the occurrence of specific events and contingencies in a domain of interest; and finally, a Knowledge Processing System (see Section 3.6), which provides task-specific decision support.

This paper provides an overview of recent scientific and technical progress within the SEMIOTIKS project. Section 2 provides an overview of the domain ontologies which were developed as part of the project. The core ontology is a humanitarian demining ontology, which was developed to support the discovery, analysis and utilization of information relevant to the humanitarian demining domain. Section 3 describes the functionality of the SEMIOTIKS system components (e.g. the Resource Discovery System, Knowledge Monitoring System, and so on). Section 4 describes a scenario that was developed to support the demonstration of SEMIOTIKS system components. The scenario features the collaboration of military (US Air Force and UK Armed Forces) and civilian (humanitarian demining) agencies in analyzing, integrating and exploiting information for the purposes of improved decision-making.

Finally, Section 5 provides a summary of current progress and a description of future work



ClearedMineHazardArea

A cleared mine hazard area is a mine hazard area whose clearance status is always 'Cleared'.

MineHazardArea
hasClearanceStatus ∋ Cleared
∃ isLocatedIn SpatialLocation

Figure 1. SEMIOTIKS humanitarian demining ontology loaded into the Protégé ontology editor.

2. DOMAIN ONTOLOGIES

A number of ontologies were developed as part of the SEMIOTIKS project. These ontologies were developed using the description logic variant of the Web Ontology Language (OWL) [1] – a language that was originally developed to support the representation of knowledge in large-scale distributed network environments. The ontologies developed as part of the SEMIOTIKS project were designed to support the demonstration of SEMIOTIKS technologies with respect to a specific demonstration scenario, namely a scenario featuring the exchange, analysis and integration of information in the humanitarian demining domain (see Section 4). In light of this, the core ontology that was developed as part of the SEMIOTIKS project is a humanitarian demining ontology that describes concepts relevant to the humanitarian demining domain. These concepts include explosive devices (and their associated characteristics), demining organizations, spatial regions, buildings, mine action activities, mine incidents, and so on. Figure 1 illustrates a small part of the humanitarian demining ontology (hereafter referred to as the SEMIOTIKS Mine Action Ontology, or SMAO) as it appears in the Protégé-OWL ontology editor [2]. The most detailed parts of the ontology are those concerned with the enumeration and characterization of explosive device types.

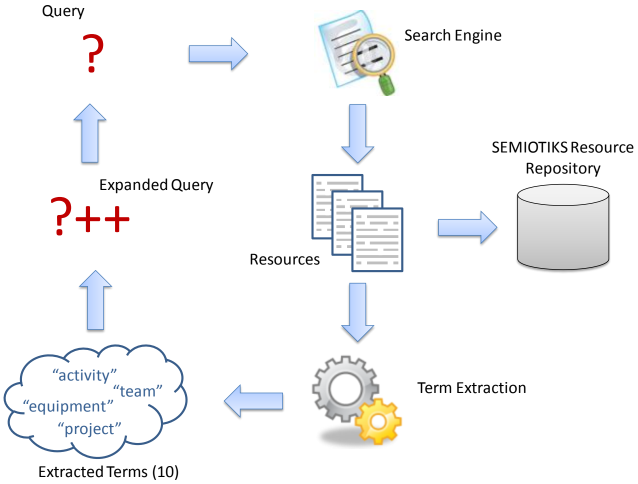


Figure 2. Windmill technique for discovering relevant information resources.

Among the other ontologies that were developed for the SEMIOTIKS project, two ontologies are of notable significance to the demonstration scenario described in Section 4. One of these serves as a semantic wrapper for a (contrived) database describing ordnance drops by the U.S. Air Force (USAF). The ontology enables the information contents of the database to be shared with other agencies (e.g. UK Armed Forces and the United Nations Mine Action Centre of Afghanistan – UNMACA) in a manner that preserves that semantics of the original dataset. Another ontology that was developed to support the scenario in Section 4 is an ontology of UK military ground operations. This ontology describes activities such as tactical targeting operations and field artillery strikes by the British Army.

3. SYSTEM COMPONENTS

The SEMIOTIKS system comprises a number of technology components that support end users with respect to the discovery, extraction and exploitation of domain-relevant information. This section provides an overview of the six main systems that are currently available for demonstration as part of the SEMIOTIKS project.

3.1. Resource Discovery System

The SEMIOTIKS Resource Discovery System aims to improve the discovery of information resources based on a user's task-specific needs and concerns. It uses a technique, called the Windmill technique, for identifying relevant resources from large scale information repositories. Figure 2 illustrates the main processing steps associated with the Windmill technique.

The Windmill technique uses a search engine to return an initial set of resources based on a user-selected term or phrase[1]. These resources are then subjected to a term extraction procedure that automatically extracts the most frequent successor words to the query term, after lexical elements such as stop words and HTML elements have been removed. Once the successor words have been extracted, the top 10 most frequent successor words are then used to expand the initial query (i.e. initial term + successor term). The process is then repeated. For each processing cycle, the resources that are selected for term extraction are cached in the SEMIOTIKS Resource Repository – a filestore containing documents, datasets, images and other resources that are relevant to the target application domain. Obviously not all of the resources that are returned by the search engine are submitted to the term extraction component. Only resources whose textual contents can be extracted are actually processed; all other resource types are rejected. At

---

[1] The system has been tested with the Google™ and Lucene search engines; however, any search engine can be used with the system

the present time, the system can support the analysis of Word, PDF and HTML resources.

The term extraction procedure selects the top 10 most frequent successor words in the resources that are returned from the search engine. (In this case, a successor word is defined as a sequence of alphanumeric characters that is terminated by a whitespace character.) Once the terms have been extracted from the resources, they are appended to the term(s) comprising the original query. So, if the original query was the term 'demining', and the extracted terms were 'equipment', 'organization', 'team', etc., then the expanded queries would be 'demining equipment', 'demining organization', 'demining team', and so on. These expanded queries would then be used to retrieve yet more resources, and these would then be used to extract yet more terms. Because only unique resources are cached in the Resource Repository, the number of resources in the Resource Repository may be significantly less than the number of resources that are actually processed for term extraction.

Our initial performance benchmarks suggest that it takes, on average, 0.0013 seconds (standard deviation: 0.0004 seconds) to process a resource once the resource has been retrieved from a filestore or network location (results obtained using a machine running Windows Vista™ with a 2.13 GHz Intel™ x64 processor and 2 GB of RAM).

3.2. Resource Browsing System

The purpose of the SEMIOTIKS Resource Browsing System is to enable users to browse a large-scale information repository and organize information resources with respect to meaningful categories or dimensions of interest. One of the main functions of the Resource Browsing System is to support a form of resource classification whereby resources can be categorized with respect to the elements of a domain ontology. Because ontologies provide semantically-enriched representations of the target domain, their association with resources in the Resource Repository supports forms of information search and retrieval that are potentially more powerful than those based on conventional keyword-driven searches. For example, if a resource provides information about demining operations in Ghazni province, we might expect the resource to be associated with representations of 'demining operations' and 'Ghazni' in the demining ontology (see Section 2). This association facilitates search and retrieval because a reasoner can infer that a resource that describes demining operations in Ghazni also describes demining operations in Afghanistan. This is because the reasoner 'knows', by virtue of the geospatial relationship between Ghazni and Afghanistan in the ontology, that Ghazni is located in Afghanistan and that anything that occurs in Ghazni must therefore also occur in Afghanistan. As such, if we are looking for resources about demining operations in Afghanistan, rather than Ghazni, our search is still likely to prove fruitful, even if the term Ghazni does not appear as part of our search string and the term Afghanistan does not appear in the source text of any of the target resources. Either of these cases would, of course, confound a conventional search engine based on keyword searches.

At the present time, the main mechanism for associating specific resources with elements (i.e. class, properties and individuals) in the domain ontology is a form of text analysis that relies on the execution of lexical classification rules[2]. The user is able to specify lexicalizations of each concept using custom OWL annotation properties, and these lexicalizations are subsequently used to categorize resources with respect to ontology elements. There are two main types of annotation properties that may be used for the purpose of specifying lexicalizations of ontology elements. These are term annotation properties, which specify a particular term, phrase, or text string, and regular expression properties, which specify a regular expression. When the resource categorization process is invoked from within the Resource Browsing System, the lexicalizations corresponding to each ontology element are matched against the resources in the Resource Repository. If the lexicalization is found in the resource then the resource is automatically associated with the ontology element(s) corresponding to the lexicalization.

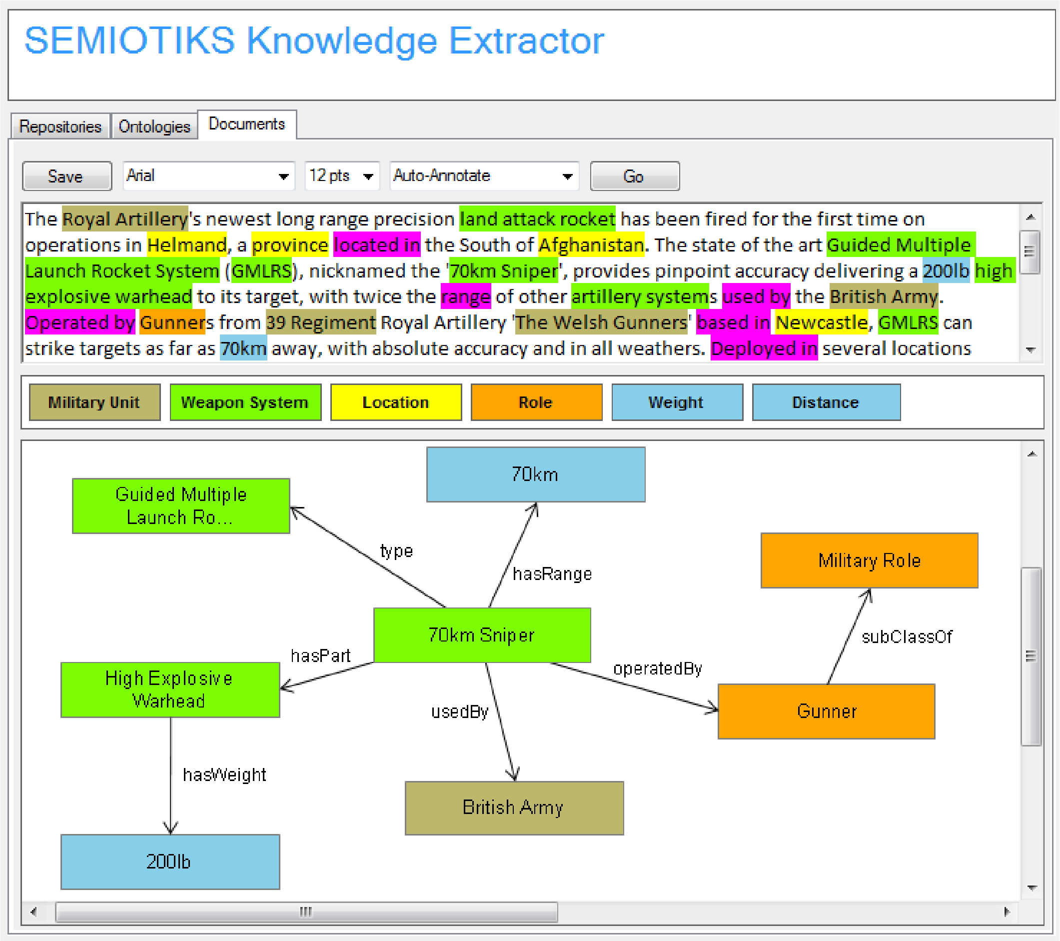
In addition to this categorization process, the organization of resources in the SEMIOTIKS Resource Repository is supported by the operation of the Resource Discovery System (see Section 3.1). The Resource Discovery System caches resources in the SEMIOTIKS Resource Repository whenever a unique resource is encountered by the term extraction component (see Figure 2). Whenever a resource is cached, the query terms that contributed to the discovery of that resource are automatically associated with the resource itself. This means that the query terms can be used to rapidly retrieve resources from the Resource Repository on subsequent occasions. Moreover, to the extent that query terms can be mapped to ontology elements (for example, the query term 'demining organization' could be mapped to the ontology class 'DeminingOrganization'), the query terms provide a relatively cheap, albeit perhaps somewhat inaccurate, means of classifying resources with respect to a domain ontology. One issue here, of course, concerns the extent to which a query term, such as 'demining organization' targets the same set of resources that might be associated with an ontology element (e.g. 'DeminingOrganization') via other means (e.g. manual annotation or the output of the aforementioned text analysis component). This is a question that we hope to address in the context of future research.

3.3. Knowledge Acquisition System

The Knowledge Acquisition System aims to provide a partially automated knowledge acquisition capability for SEMIOTIKS. Its main purpose is to distil the knowledge content from a resource in a form that is suitable for machine-based processing. One way in which this can be accomplished is by using screen-scraping techniques. Some

---

[2] The association between ontology elements and resources can also be established using manual editing techniques.



The Royal Artillery's newest long range precision land attack rocket has been fired for the first time on operations in Helmand, a province located in the South of Afghanistan. The state of the art Guided Multiple Launch Rocket System (GMLRS), nicknamed the '70km Sniper', provides pinpoint accuracy delivering a 200lb high explosive warhead to its target, with twice the range of other artillery systems used by the British Army. Operated by Gunners from 39 Regiment Royal Artillery 'The Welsh Gunners' based in Newcastle, GMLRS can strike targets as far as 70km away, with absolute accuracy and in all weathers. Deployed in several locations

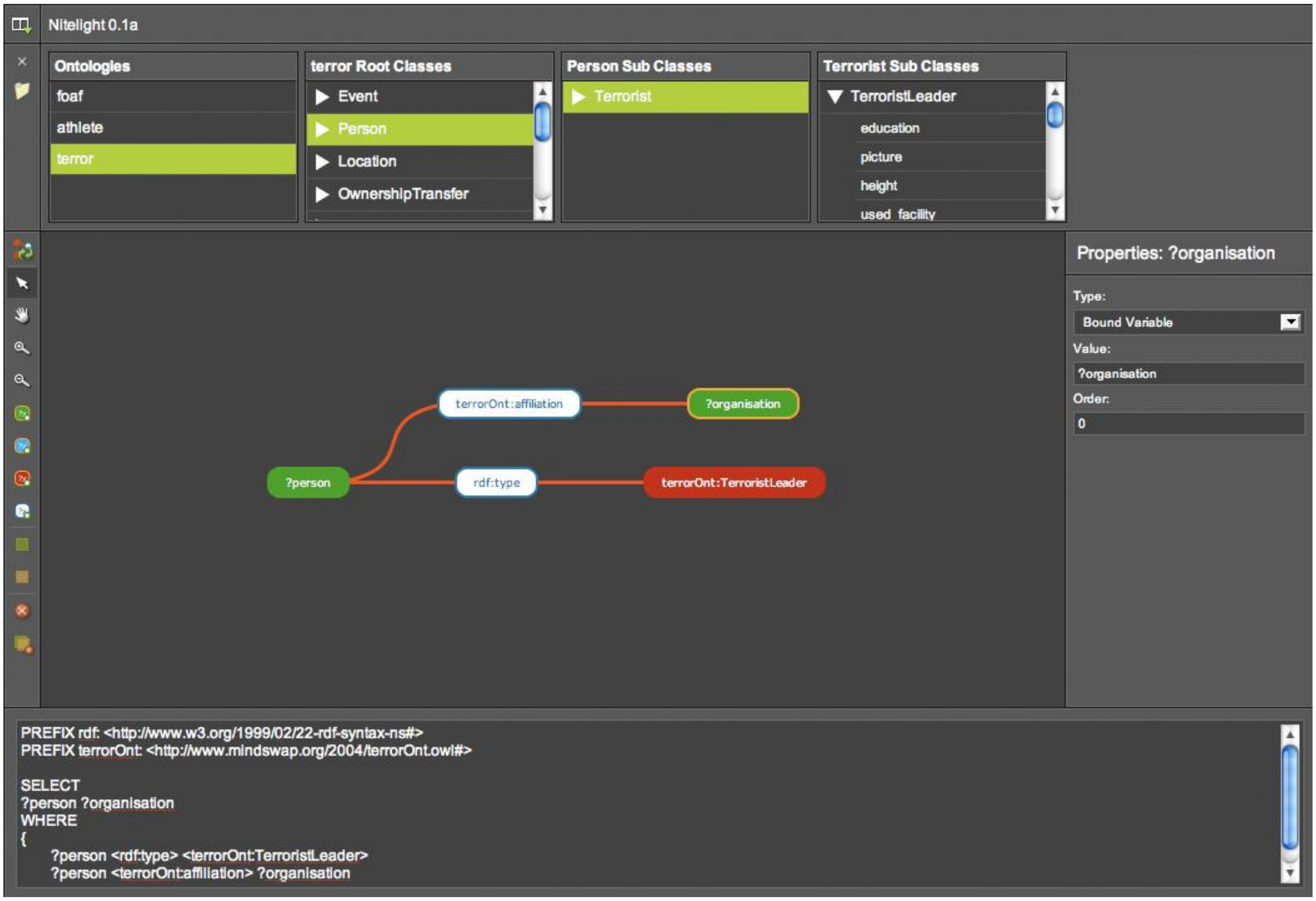
Figure 3. SEMIOTIKS Knowledge Extraction Tool.

resources, such as the ORDATA system [3], provide access to rich sources of semi-structured information, and this can be harvested for the purposes of populating a domain ontology. Using screen scraping techniques, the SEMIOTIKS Knowledge Acquisition System was able to extract information about 953 types of landmine and assert this information into the SMAO. Information extracted by screen scraping techniques included the country of use, fragmentation range, detectability and physical dimensions of each landmine.

The primary disadvantage of screen scrapers as a method of knowledge acquisition is that a significant amount of manual effort is required to develop and maintain them. Any change in the way in which information is presented in the source document tends to require significant adaptation of the screen scraping component. Alternative techniques, such as those based on machine learning and information extraction techniques [4], seem to afford much greater promise when it comes to harvesting information from Web-accessible resources. We have not, however, examined the use of such techniques in SEMIOTIKS at the present time.

Another method for extracting knowledge content from textual resources entails the use of semantic annotation and knowledge editing tools. One such tool (the SEMIOTIKS Knowledge Extraction Tool) has been developed within SEMIOTIKS to assist users with the process of extracting knowledge from unstructured textual resources. Figure 3 shows the main interface of the Knowledge Extraction Tool. The textual content of a selected resource is displayed in a text pane in the upper half of the tool window. The user can highlight fragments of this text using a number of ‘marker pens’, each of which denotes a particular type of ontology element. For example, a user may use a marker pen that is associated with a particular class in the ontology, e.g. ‘Province’, to mark-up a fragment of text that reads ‘Helmand Province’. Similarly, they could use a marker pen associated with the ontology property ‘isLocatedIn’ to highlight text fragments such as ‘located in’. This process of manual semantic annotation is a common feature of many knowledge acquisition tools. It features as part of the knowledge acquisition and modeling toolkit, PCPACK [5], and the technique has even been applied to machine learning scenarios in which a system progressively learns to annotate text fragments by continuously monitoring user input [6, 7].

The bottom pane of the Knowledge Extraction Tool (see Figure 3) is used to structure the knowledge content identified by the mark-up process. This is accomplished using an interactive graph-based visualization approach. The



Properties: ?organisation

Type: Bound Variable

Value: ?organisation

Order: 0

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX terrorOnt: <http://www.mindswap.org/2004/terrorOnt.owl#>

SELECT
?person ?organisation
WHERE
{
      ?person <rdf:type> <terrorOnt:TerroristLeader>
      ?person <terrorOnt:affiliation> ?organisation
```

Figure 4. NITELIGHT graphical query designer.

basic idea is that knowledge objects (represented as nodes in the graph visualization) can be connected to one another using relationships (represented as arcs). For example, the knowledge object ‘Helmand Province’ may be associated with another knowledge object, e.g. ‘Afghanistan’, using the relationship ‘isLocatedIn’. Eventually, as more and more knowledge objects in the text are associated with one another using relationships (or properties), a semantically-enriched representation of the knowledge contained in the resource is created. This representation can be subsequently converted to OWL and stored in the Knowledge Repository. As more and more resources are processed, an increasingly rich and detailed picture is established of the situations, phenomena or objects described by multiple textual resources.

A number of services support the user in extracting knowledge content using the Knowledge Extraction Tool. Firstly, a named entity resolution service facilitates the automatic identification of common entities, such as places, people, agencies and organizations. In addition, a date/time resolution service assists users with the identification and instantiation of date/time values in the ontology. Another annotation service that is currently under development is the Concept Resolution Service (CRS). This service uses semantic lexicons, such as WordNET [8], to improve the recognition of ontology elements in textual resources. Such services support the user in identifying and extracting knowledge content, but they do not necessarily support the knowledge structuring process, i.e. the process that uses relationships to connect knowledge objects into a semantic network. Ideally, what is required is a system that is able to automatically process textual resources and extract detailed *relational* information, i.e. the information that connects entities mentioned at different locations in a single source document (or even in multiple, separate documents). Very few systems can extract information of this kind, although Alani et al [9] present a system that seems capable of at least some forms of relational information extraction.

3.4. Information Retrieval System

The SEMIOTIKS Information Retrieval System consists of a graphical query editor that (among other things) supports the retrieval of task-relevant information. The editor is based on the NITELIGHT tool (see Figure 4), which is a Web-based graphical query editor designed to support the creation and editing of semantic queries[3] [11-13]. The tool combines a number of features to support end-users with respect to the creation of semantic queries. These include a columnar ontology browser, an interactive graphical design surface, a SPARQL-compliant visual query language and a

---

[3] NITELIGHT supports the creation of semantic queries using the SPARQL query language [10].

SPARQL syntax viewer. The tool also allows users to execute SPARQL queries against a specified knowledge base and to view the results of query execution using an integrated query results browser. Because NITELIGHT has been the focus of a number of previous publications [11-13], we refrain from a detailed description of the capabilities in the current paper; the interested reader is referred to Smart et al [12] for more information.

Besides its role in information retrieval, the NITELIGHT tool also supports a number of query-related capabilities. These include rule creation, rule processing and semantically-mediated information integration. Regarding the rule creation capability, NITELIGHT capitalizes on the availability of the SPARQL CONSTRUCT query form [see 10] to contingently modify and extend an ontology based on the presence of information contained in multiple, distributed information repositories. An end user can use NITELIGHT to create a SPARQL CONSTRUCT query that is functionally equivalent (in terms of the contingent assertion of RDF triples) to a rule represented in a more conventional language, such as SWRL [14]. Figure 5 illustrates a SWRL rule that is used to categorize mine incidents based on the number of casualties associated with the incident. This rule could be executed as part of the SEMIOTIKS Knowledge Processing System (see Section 3.6), but the ability to create CONSTRUCT queries using the NITELIGHT editor enables us to realize the same capability within the context of the Information Retrieval System. In this case, the NITELIGHT tool could be used to create the CONSTRUCT query illustrated in Figure 6. The execution of this query would support the same kind of rule-based categorization function as is achieved with the SWRL rule in Figure 5.

```
MineIncident(?x) ∧
hasCasualityNumber(?x, ?y) ∧
swrlb:greaterThan(?y, 20)
→
MajorMineIncident(?x)
```

Figure 5. A SWRL rule representing the fact that mine incidents with more than 20 causalities are members of the 'MajorMineIncident' class.

The role of the NITELIGHT tool in effecting semantic information integration is based on the same principle as that seen in the case of rule processing. Here a SPARQL CONSTRUCT query effectively implements an information exchange or information transfer solution that is grounded in the ontology alignments derived using manual and/or automatic methods. Figure 7 illustrates a CONSTRUCT query that exemplifies the approach. The query translates information about suicide bomb attacks from one ontology ('ito') to another ontology ('edto').

The Information Retrieval System also supports the direct entry of SPARQL queries using a Semantic Query Editor component. This is a simple text editing facility that enables users to create a semantic query using the SPARQL query language. This technique is clearly less sophisticated than the use of graphical query editing techniques, but it is often more straightforward for users who are already familiar with the syntactic idiosyncrasies of the SPARQL query language.

```
CONSTRUCT
{
   _:t rdf:type sem: MajorMineIncident .
}
WHERE
{
   ?x rdf:type sem:MineIncident .
   ?x sem:hasCasualityNumber ?y .
   FILTER (?y > 20)
}
```

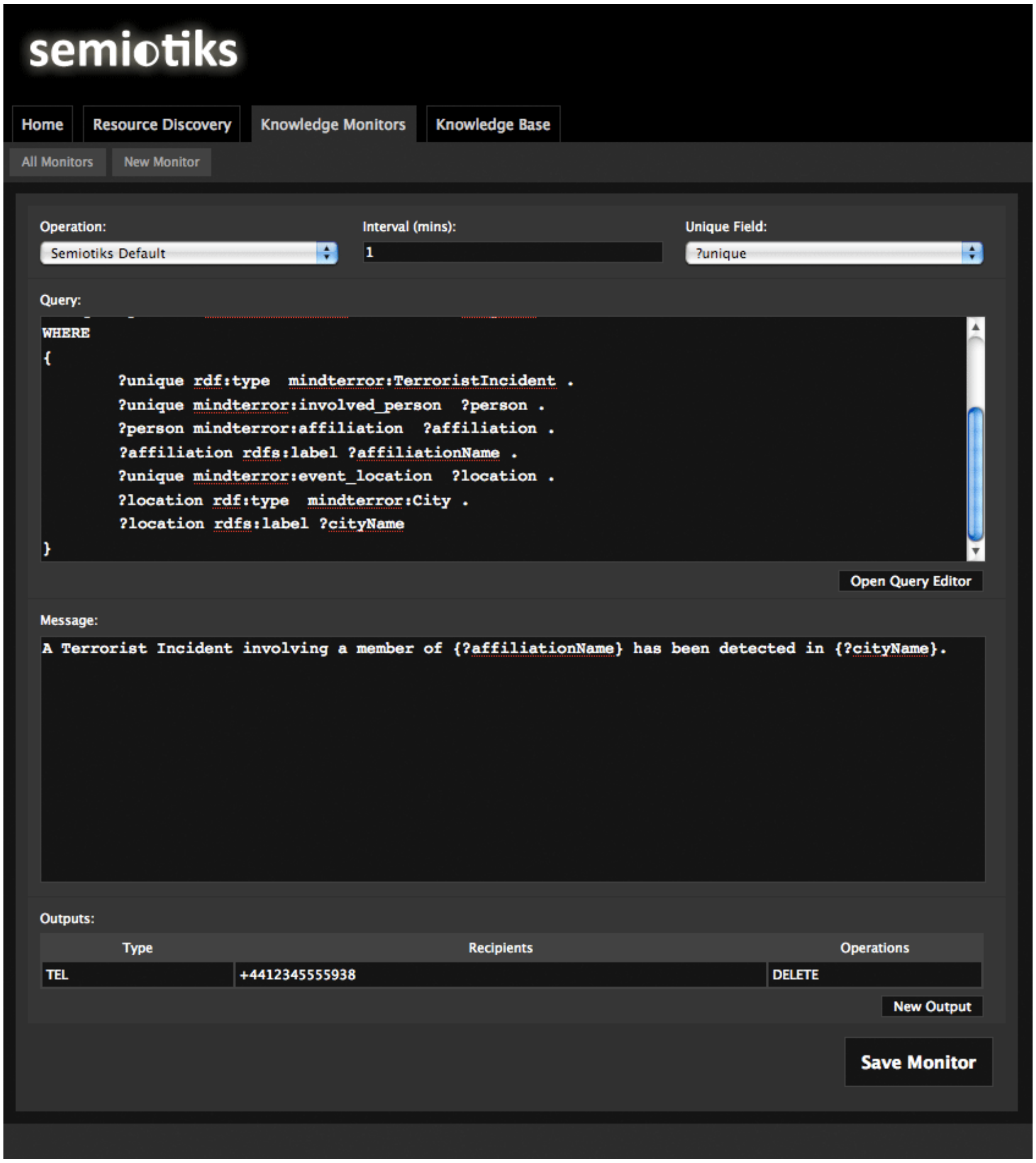
Figure 6. A SPARQL CONSTRUCT query that is functionally equivalent to the SWRL rule in Figure 5.

3.5. Knowledge Monitoring System

The notion of a semantically-mediated knowledge monitoring system was introduced in Smart et al [15] as a means of supporting enhanced awareness about the occurrence of specific events or contingencies in a domain of interest. In the SEMIOTIKS Knowledge Monitoring System, knowledge monitors are implemented as semantic queries that periodically execute against a specified SPARQL endpoint. In creating the semantic query component of a knowledge monitor, a user can exploit the NITELIGHT tool (see Section 3.4) to graphically edit the required semantic query. They can also edit the value of a number of settings that determine things such as the frequency of query execution, the location of SPARQL endpoints and the type of alerting mechanism to use whenever the semantic query returns a new query resultset. This last feature enables a user to customize the way they are notified of changes to an information repository. For example, one type of alerting mechanism enables users to edit a voicemail message that can be sent to the voice inbox of a mobile phone. The message can include a combination of plain text strings and query variable bindings, so that the

```
CONSTRUCT
{
   _:t rdf:type edto:TerroristAttack .
   _:t edto:isSuicideAttack xsd:true .
   _:d ref:type edto:ExplosiveDevice .
   _:t edto:uses edto _:d
}
WHERE
{
   ?x rdf:type ito:TerroristIncident .
   ?x ito:hasType ito:Bombing .
   ?x ito:involvesWeapon ito:Explosive .
   ?x ito:hasVictim ?victim .
   ?victim ito:isFatality xsd:true .
   ?victim rdf:type ito:Terrorist .
   ?x ito:perpetratedBy ?victim
}
```

Figure 7. SPARQL CONSTRUCT query implementing an information exchange solution for two ontologies. The query translates information about suicide bomb attacks from one ontology to the other.



Operation: Semiotiks Default

Interval (mins): 1

Unique Field: ?unique

Query:

```
WHERE
{
        ?unique rdf:type  mindterror:TerroristIncident .
        ?unique mindterror:involved_person  ?person .
        ?person mindterror:affiliation  ?affiliation .
        ?affiliation rdfs:label ?affiliationName .
        ?unique mindterror:event_location  ?location .
        ?location rdf:type  mindterror:City .
        ?location rdfs:label ?cityName
}
```

Message:

```
A Terrorist Incident involving a member of {?affiliationName} has been detected in {?cityName}.
```

Outputs:

Type	Recipients	Operations
TEL	+4412345555938	DELETE

Figure 8. SEMIOTIKS Knowledge Monitor Editor component.

results of query execution can be directly incorporated into the voicemail message (see Figure 8). Whenever the knowledge monitor is activated (i.e. the semantic query associated with the knowledge monitor returns a new set of query results), the value of variable bindings are incorporated into the text message and the message is converted into a voicemail message using the Microsoft Speech Synthesis API. The voicemail message is finally distributed to its target destination (i.e. a specific telephone number) using the Skype4COM API[4].

Other alerting mechanisms that are currently supported by the Knowledge Monitoring System include email, Simple Message Service (SMS) and RSS feeds. Figure 8 illustrates the user interface of the Knowledge Monitor Editor component. This component supports users in creating and configuring knowledge monitors.

3.6. Knowledge Processing System

The SEMIOTIKS Knowledge Processing System provides the functionality necessary to create, load and execute rules that support task-relevant decision-making. At the heart of the Knowledge Processing System are two third-party reasoning engines that provide support for the two types of reasoning encountered in the SEMIOTIKS Knowledge Processing System[5]:

1) **Description Logic (DL) Reasoning.** DL reasoning relies on the semantics of the knowledge representation language in order to infer information about the domain of discourse. This is the type of reasoning that is most commonly encountered in Semantic Web

---

[4] See https://developer.skype.com/Docs/Skype4COM.

[5] This distinction is somewhat arbitrary. DL or subsumption reasoning can easily be supported using conventional rules-based systems, such as JESS or CLIPS. Meditskos and Bassiliades [16] provide a demonstration of this capability using the CLIPS expert system shell.

application contexts. It can be easily accomplished by relying on ontology representation languages, such as OWL, and a DL reasoner, such as Pellet [17].

2) **Rule-Based Reasoning.** Rule-based reasoning involves the use of rules that capture knowledge-rich contingencies that inhere in the domain of discourse. The rules are represented using the Semantic Web Rule Language (SWRL) [14], and reasoning is performed using a rule engine such as the Java Expert System Shell (JESS) [18].

The SEMIOTIKS Knowledge Processing System delegates DL Reasoning to Pellet, an open-source Java-based OWL DL reasoner [17]. Rule-based reasoning is, however, more complicated. In this case, the rules are represented using an extended[6] version of the rule language SWRL [14], and reasoning is performed using the JESS rule engine. SWRL is a candidate rules language for the Semantic Web that extends the OWL specification to include support for rules. The rules take the form of an implication between an antecedent (body) and a consequent (head), and the intended meaning of the rule can be read as: whenever the conditions specified in the antecedent hold, then the conditions specified in the consequent must also hold. Figure 5 illustrates a simple SWRL rule for the domain of humanitarian demining. As was described in Section 3.4, this rule is being used to categorize mine incidents based on the number of casualties that are associated with the mine incident. This categorization is not possible with the DL reasoner Pellet because the ranges of OWL Datatype properties (for example, string and integer values) are not processed by the reasoner. Figure 9 presents a more complex SWRL rule that is defined as part of the SEMIOTIKS demonstration scenario (see Section 4). The purpose of this rule is to support the identification of high priority mine hazard areas; i.e. mine hazard areas that are regarded as high priority targets for humanitarian demining interventions.

The SEMIOTIKS Knowledge Processing System is built from a number of third-party components, including the aforementioned Pellet and JESS rule engines, as well as the Protégé-based SWRLTab editor [19]. DL reasoning is supported by the DL Implementation Group (DIG) interface provided by the Protégé ontology editor, while SWRL rule execution is supported via a SWRL-JESS bridge component that instantiates SWRL rules within the JESS rule engine environment [see 19].

Although, complex rules can be created and executed in the SEMIOTIKS Knowledge Processing System, it should be remembered that, at least in some cases, the NITELIGHT component of the SEMIOTIKS Information Retrieval System can also be used for rule-based processing. As we discussed in Section 3.4, the use of the SPARQL CONSTRUCT query form, enables the NITELIGHT editor to serve as a simple rule editor component. In some cases, this can be used to support rule execution, although more complex forms of rule execution (e.g. those requiring the definition and invocation of SWRL built-in or custom functions) still require the SEMIOTIKS Knowledge Processing System.

```
MineIncidentIn2007(?incident) ∧
hasCasualityNumber(?incident, ?casualties) ∧
swrlb:greaterThan(?casualties, 2) ∧
hasSpatialLocation(?incident, ?incidentLoc) ∧
hasLatitude(?incidentLoc, ?incidentLat) ∧
hasLongitude(?incidentLoc, ?incidentLon) ∧
Settlement(?settlement) ∧
hasSpatialLocation(?settlement, ?setLoc) ∧
hasLatitude(?setLoc, ?setLat) ∧
hasLongitude(?setLoc, ?setLon) ∧
distanceFromLocationLessThan(?incidentLat,
?incidentLon, ?setLat, ?setLon, 20) ∧
hasPopulation(?settlement, ?population) ∧
swrlb:greaterThanOrEqual(?population, 500) ∧
MineHazardArea(?area) ∧
contains(?area, ?incident)
  →
HighPriorityMineHazardArea(?area)
```

Figure 9. A SWRL rule used to categorize mine hazard areas according to their relative priority for mine clearance operations. This particular rule is used to detect 'high-priority' mine hazard areas.

4. DEMONSTRATION SCENARIO

In order to demonstrate the use of SEMIOTIKS system components in a multi-agency, coalition context, we developed a scenario that features the involvement of both military and civilian agencies in the analysis, exchange and utilization of information from a variety of sources. The scenario shows how a particular coordinating agency (in our case, UNMACA) could exploit SEMIOTIKS technology components for the purposes of advanced knowledge processing. It also shows, by virtue of the SEMIOTIKS knowledge monitoring capabilities, how military agents can be apprised of the decisions and actions of civilian agencies in order to shape the profile of both current and future military engagements.

The scenario is set in Afghanistan in late 2008. It assumes that UNMACA (the agency responsible for the coordination of mine action activities in Afghanistan) is charged with the task of systematically assessing and prioritizing areas for humanitarian demining in the southwestern region of Afghanistan, a region that subsumes the provinces of Helmand, Kandahar, Nimruz, Orūzgān and Zabul. As a result of the recent, and ongoing, conflict between neo-Taleban insurgent fighters and military coalition forces, the area of interest has been extensively contaminated with Explosive Remnants of War (ERW). This means that UNMACA needs to update its mine hazard area databases, and reassess the requirements for humanitarian demining in the affected region. In order to accomplish this, UNMACA requests information about coalition military activities for the period 2003-2008. Military activities of interest to

---

[6] The form of extension here relates to the inclusion of custom functions to support the evaluation of specific contingencies, e.g. functions to support distance and bearing calculations between two geographic points.

UNMACA include precision air strikes against Taleban positions (undertaken predominantly by the US Air Force) and offensive ground operations (led exclusively by British Armed Forces). The scenario comprises the following sequence of demonstration steps:

1) **Transfer of information about USAF air strikes.** USAF authorizes the release of declassified data concerning ordnance drops against Taleban insurgents for the period 2003-2008. The transfer is accomplished using a simple export of a USAF Air Tasking Orders (ATO) database to RDF/XML. The RDF data is supplied to UNMACA, along with a simple ontology describing the semantics of the dataset.
2) **Transfer of resources detailing UK ground operations.** For the purposes of demonstration, we assume that the only source of information about UK ground operations is a series of unstructured and semi-structured textual documents. These textual documents are derived from a variety of sources including military plans, SITREPS and field reports. The resources are made available to UNMACA as is; i.e. no attempt is made to analyze the information content of the resources, other than to remove classified or sensitive information.
3) **Integration of USAF air strike data with the mine action ontology.** In order to support the analysis of information about US air strikes, analysts at UNMACA align and integrate information about the USAF ordnance drops with their mine action ontology (i.e. the SEMIOTIKS Mine Action Ontology described in Section 2). The ontology mappings resulting from the alignment process[7] are represented as SPARQL CONSTRUCT queries using the NITELIGHT tool (see Section 3.4), and information transfer (from the USAF ordnance drop ontology to the SMAO) is effected using a SPARQL query processor.
4) **Selection of relevant textual reports concerning UK ground operations.** As mentioned previously, information about UK ground operations is made available to UNMACA as a large set of unstructured and semi-structured textual resources. In order to make the information content of these resources accessible to automated forms of knowledge processing and analysis, UNMACA uses the SEMIOTIKS Resource Browsing System (see Section 3.2) and the SEMIOTIKS Resource Discovery System (see Section 3.1) to organize and filter the information resources provided by the British Army. The Resource Discovery System is used to analyse the frequency of specific terms and phrases in the resources, and this information is subsequently used to guide the specification of lexicalizations for relevant elements of the SMAO[8]. Once specified, the lexicalizations are used by the Resource Browsing system to effect the categorization of resources with respect to the elements (classes, properties and individuals) of the SMAO.
5) **Detailed analysis of textual reports concerning UK military operations.** Having selected a subset of relevant textual reports using the SEMIOTIKS Resource Discovery System (see Section 3.1) and the Resource Browsing System (see Section 3.2), UNMACA analysts extract knowledge from the text documents using the SEMIOTIKS Knowledge Extraction Tool (see Section 3.3). The result of the knowledge extraction process is a set of resource-related knowledge statements (RDF triples) that can be added as instance data to the SMAO.
6) **Analysis of integrated mine action ontology using semantic query and rule processing capabilities.** Now that all relevant military data has been assimilated into the SMAO, UNMACA analysts use the SEMIOTIKS Information Retrieval System (see Section 3.4), as well as the SEMIOTIKS Knowledge Processing System (see Section 3.6) to analyse the integrated dataset and make decisions about the relative priority of mine hazard areas. As part of their decision-making routines they invoke SWRL rules that categorize mine hazard areas into 'low', 'medium' and 'high' priority mine hazard areas. They also assess the ability of different demining organizations to clear specific areas. This latter assessment capability is important, because not all demining organizations are willing to operate in areas of ongoing military conflict. In addition, specific demining organizations may be differentially suited to deal with the particular kinds of ERW likely to be encountered in different mine hazard areas (this might be because of differences in experience, training or access to particular types of equipment).
7) **Creation of knowledge monitors to support situation awareness about humanitarian demining operations.** In the final stage of the scenario, we assume that UNMACA has generated a prioritized list of mine hazard areas and has agreed contracts for ERW clearance operations with specific humanitarian demining organizations. In order to support an awareness of Who is Doing What Where (WDWW), UNMACA maintains a database of active humanitarian demining activities. The contents of this database are made available to coalition military forces via a SPARQL endpoint that executes semantic queries against an ontology-based interface to the underlying WDWW database. US and UK coalition forces use this SPARQL endpoint, in conjunction with the SEMIOTIKS Knowledge Monitoring System (see Section 3.5), to establish knowledge monitors that provide them with up-to-date information about the location and status of humanitarian demining activities within specific Areas of Interest (AoI). The mechanisms used for the knowledge monitor alerts include email and mobile phone (voicemail) alerts. In

---

[7] The ontology alignment process itself does not form part of the demonstration.

[8] In the context of the demonstration scenario, relevant ontology elements include those representing particular provinces (e.g. 'sem:HelmandProvince') and those representing military activities entailing the use of particular weapons systems (e.g. ground operations involving the use of AS90 artillery systems).

addition, the outputs of the knowledge monitors are integrated into electronic map products in order to guide coalition military planning activities.

5. CONCLUSION

This paper has described some of the technologies developed as part of the DIFDTC SEMIOTIKS project to support the discovery and utilization of information from large-scale information repositories. The components that are currently available for demonstration include resource discovery and resource classification components, software-assisted knowledge extraction tools, graphical query interfaces, knowledge monitoring capabilities and a knowledge processing system that supports decision-making in a specific task context, namely, mine hazard area assessment and prioritization.

Future work in this area will aim to extend the knowledge extraction capabilities of the system by investigating whether Natural Language Processing (NLP) technologies can extract *relational* information (i.e. information about the relationship between particular entities) from unstructured textual resources. We also aim to extend the resource classification capability, described in Section 3.2, by exploiting Machine Learning capabilities. In particular, we aim to develop a system that learns to classify resources using a Support Vector Machine (SVM) learning algorithm. A final focus area for research concerns the development of a natural language query system to complement the graphical query system described in Section 3.4. Our specific aim in this respect is to develop a natural language question-answering capability similar to that provided by the QUESTIO system [20].

In conclusion, this paper has presented the various technology components developed as part of the DIFDTC SEMIOTIKS project. These components can be used to enhance the discovery, retrieval and utilization of task-relevant information from large-scale, distributed information repositories. Such capabilities are of generic relevance to military and civilian agencies who want to leverage the latent potential of information networks to enhance situation awareness and improve knowledge processing efficiency.

ACKNOWLEDGMENT

This work was supported by a contract awarded to QinetiQ and the University of Southampton as part of the UK DIFDTC initiative. QinetiQ act as cluster lead for the SEMIOTIKS initiative and sponsor the research undertaken therein. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the UK Ministry of Defence, or the UK Government.

REFERENCES

[1] D. L. McGuiness and F. van Harmelen, "OWL Web Ontology Language Overview," 2004, http://www.w3.org/TR/owl-features/.

[2] K. Holger, R. W. Ferguson, N. F. Noy, and M. A. Musen, "The Protege OWL Plugin: An Open Development Environment for Semantic Web Applications," in *3rd International Semantic Web Conference (ISWC'04)*, Hiroshima, Japan, 2004.

[3] "ORDATA Online: International Deminer's Guide to UXO Identification, Recovery and Disposal," http://maic.jmu.edu/ordata/Mission.asp.

[4] F. Ciravegna, S. Chapman, A. Dingli, and Y. Wilks, "Learning to harvest information for the semantic web," in *The Semantic Web: Research and Applications (First European Semantic Web Symposium, ESWS 2004)*, C. Bussler, J. Davies, D. Fensel, and R. Studer, Eds. London: Springer Verlag, 2004, pp. 312-326.

[5] "Epistemics Website," http://www.epistemics.co.uk/.

[6] M. Vargas-Vera, E. Motta, J. Domingue, M. Lanzoni, A. Stutt, and F. Ciravegna, "MnM: ontology driven semi-automatic or automatic support for semantic markup," in *13th International Conference on Knowledge Engineering and Knowledge Management*, Siguenza, Spain, 2002.

[7] F. Ciravegna and Y. Wilks, "Designing adaptive information extraction for the Semantic Web in Amilcare," in *Annotation for the Semantic Web*, S. Handschuh and S. Staab, Eds. Amsterdam: IOS Press, 2003.

[8] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to WordNet: An On-line Lexical Database," *International Journal of Lexicography,* vol. 3, pp. 235-244, 2004.

[9] H. Alani, S. Kim, D. E. Millard, M. J. Weal, W. Hall, P. Lewis, and N. R. Shadbolt, "Automatic Ontology-Based Knowledge Extraction from Web Documents," *IEEE Intelligent Systems,* vol. 18, pp. 14-21, 2003.

[10] E. Prud'hommeaux and A. Seaborne, "SPARQL Query Language for RDF," 2005, http://www.w3.org/2001/sw/DataAccess/rq23/.

[11] A. Russell, P. R. Smart, D. Braines, and N. R. Shadbolt, "NITELIGHT: A Graphical Tool for Semantic Query Construction," in *Semantic Web User Interaction Workshop (SWUI'08) hosted by the 26th CHI Conference (CHI'08)*, Florence, Italy, 2008.

[12] P. R. Smart, A. Russell, D. Braines, Y. Kalfoglou, J. Bao, and N. R. Shadbolt, "A Visual Approach to Semantic Query Design Using a Web-Based Graphical Query Designer," in *16th International Conference on Knowledge Engineering and Knowledge Management (EKAW'08)*, Acitrezza, Catania, Italy, 2008.

[13] A. Russell and P. R. Smart, "NITELIGHT: A Graphical Editor for SPARQL Queries," in *7th International Semantic Web Conference (ISWC'08)*, Karlsruhe, Germany, 2008.

[14] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, G. Benjamin, and M. Dean, "SWRL: A Semantic Web Rule Language Combining OWL and RuleML," 2004, http://www.w3.org/Submission/SWRL/.

[15] P. R. Smart, A. Russell, N. Shadbolt, m. c. schraefel, and L. A. Carr, "AKTiveSA: A Technical Demonstrator System for Enhanced Situation Awareness in Military Operations Other Than War," *The Computer Journal,* vol. 50, pp. 703-716, 2007.

[16] G. Meditskos and N. Bassiliades, "A Rule-based Object-Oriented OWL Reasoner," *IEEE Transactions on Knowledge and Data Engineering,* vol. 20, pp. 397-410, 2008.

[17] B. Parsia and E. Sirin, "Pellet: An OWL DL Reasoner," in *International Workshop on Description Logics*, British Columbia, Canada, 2004.

[18] "JESS: The Rule Engine for the Java Platform," http://herzberg.ca.sandia.gov/jess/.

[19] M. J. O'Connor and A. Das, "The SWRLTab: An Extensible Environment for working with SWRL Rules in Protégé-OWL," in *RuleML*, Athens, Georgia, USA, 2006.

[20] V. Tablan, D. Damljanovic, and K. Bontcheva, "A Natural Language Query Interface to Structured Information," in *5th European Semantic Web Conference (ESWC'08)*, Tenerife, Spain, 2008.