2006 CCRTS

THE STATE OF THE ART AND THE STATE OF THE PRACTICE

Taxonomic and Faceted Classification for Intelligent Tagging and Discovery in Net-Centric Command and Control

The publication of this paper does not indicate endorsement by the Department of Defense or IDA, nor should the contents be construed as reflecting the official positions of those organizations.

Topics: Cognitive Domain Issues, Social Domain Issues, C2 Experimentation

Authors: Dale E. Lichtblau, Andrew W. Trice, Steven P. Wartik

Point of Contact: Dale E. Lichtblau

Institute for Defense Analyses

4850 Mark Center Dr.

Alexandria, VA 22311

703-845-6683

del@ida.org

Abstract

The success of Net-Centric Operations and Warfare (NCOW) depends upon the ability of net-centric environment (NCE) users—both human and automated—to readily discover useful information and services. Effective discovery requires, in turn, good semantic metadata "tagging" (i.e., indexing the functions of the services). Good tagging reflects the contextual relationships among the discoverable artifacts. It derives its value from the soundness—and intuitiveness—of its underlying approach to information and services classification. Unfortunately, classification "soundness" is mostly in the eye of the beholder, particularly for services that can be deployed for many different purposes, and not all necessarily foreseen by their initial developers. Ultimately, therefore, what is needed for more rapid and effective tagging and discovery is a services classification approach that accommodates multiple perspectives as to what constitutes a natural and intuitive characterization, plus tools that enable NCE users to take advantage of these capabilities without being overwhelmed by the sheer multiplicity of different classification perspectives. This paper presents a proposed structure for the semantic metadata that we believe will facilitate service and information discovery in the NCE, and will easily accommodate use by intelligent agents.

Taxonomic and Faceted Classification for Intelligent Tagging and Discovery in Net-Centric Command and Control

Dale E. Lichtblau, Andrew W. Trice, Steven P. Wartik
Information Technology and Systems Division
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311
{del,atrice,swartik}@ida.org

1. Introduction

It has been argued before in [1] "that the envisaged net-centricity in future warfare (command and control), business operations, and enterprise management is dependent upon a robust intelligent assistance capability based on the profuse use of intelligent agents throughout the [Global Information Grid's Net-Centric Information Environment]." This paper takes that argument one step further, arguing that the success of Net-Centric Operations and Warfare (NCOW) depends upon the ability of net-centric environment (NCE) users—both human and automated—to readily discover useful information and services. It also presents a proposed structure for the semantic metadata that we believe will facilitate service and information discovery in the NCE. Effective discovery requires, in turn, effective metadata "semantic tagging" (i.e., tagging that describes the function and meaning of services).[1] *Discovery* and *tagging* can be considered opposite sides of the same coin: successful discovery depends on good tagging; good tagging reflects the contextual relationships among the discoverable artifacts. This coin thus derives its value from the soundness—and intuitiveness—of its underlying approach to information and services classification. Unfortunately, classification "soundness" is often in the eye of the beholder[2], particularly for services that can be deployed for many different

[1] We use the term "semantic tagging" in contrast to "syntactic tagging". Syntactic tagging, as implemented by standards such as WSDL [2], explains how to invoke the service or the service point-of-contact, and so forth. The recent draft version of Department of Defense's Enterprise Services Strategy (ESS), [3], underscores the importance of meta-data tagging for NCE services, but it does not offer a practical approach to effectively "tagging" *what* a tagged service is suppose to do. This paper attempts, in part, to address that lack.

[2] A simple example will illustrate the essence of the problem. There are two obvious candidates for the top-most layer of a services taxonomy for DoD services, one based on the four core components of warfare (strike, maneuver, logistics, and force protection) and one based on the high-level functional characteriza-

purposes, and not all necessarily foreseen by their initial developers. Ultimately, therefore, what is needed for more rapid and effective tagging and discovery is a services classification approach that accommodates multiple perspectives as to what constitutes a natural and intuitive characterization, plus tools that enable NCE users to take advantage of these capabilities without being overwhelmed by the sheer multiplicity of different classification perspectives.

This paper describes a novel approach to the classification of net-centric services and a prototype of a services[3] discovery and tagging tool that implements this classification scheme. The prototype demonstrates the feasibility of the approach and illustrates its value in delivering the promise of a services-oriented architecture (SOA). The paper is organized as follows. Section 2 summarizes the basic problem this paper tackles. Section 3 describes, in general, a two-pronged approach to the classification of services. Section 4 describes the prototype service tagging and discovery tool we developed to illustrate the approach and to demonstrate its feasibility. Section 5 discusses our plans for extending the prototype in two basic directions: to apply it to the semantic classification of information, and to enhance its intelligent agent-based features.

2. Problem Statement

Two fundamental questions arise in connection with semantic tagging of services. One, why go to the trouble of using a structured semantic tagging scheme, rather than an *ad hoc* tagging approach—or no tagging at all? And two, if we accept the need for a semantic tagging scheme, what kind of—and how extensive—should the semantic meta-data structure be?

Why Use a Structured Classification Approach?

Regarding the first question, some writers have argued that structured semantic approaches to tagging content that utilize taxonomies and their (implicit or explicit underly-

---

tion of the mission of the department (force application, force support, and force management). The immediate question becomes, how to reconcile these separate perspectives to each other?

[3] Note that while this paper is focused on *services*, there is no reason why the approach cannot be applied to the meta-data tagging and subsequent discovery of *information* in the NCE as well.

ing) ontologies[4] are greatly overrated [4, 5]. They assert that Web sites or systems that use *ad hoc* tagging techniques (e.g., the Web site del.icio.us) or rely on full text indexing only (e.g., Google and other search engines) have been extraordinarily successful, to the point that taxonomies have become almost irrelevant. Specifically, the claim is that the less structured approaches provide more flexibility and power for content discovery with fewer maintenance issues when compared to a structured taxonomy approach to navigating through content (similar to what Yahoo emphasized in earlier years).

We acknowledge that such less structured and more decentralized approaches to discovering content (including service descriptions) can be valuable, but also argue that providing some further structure in the tagging approach also provides key advantages, particularly when the focus of discovery is narrowed to a set of domain-specific service descriptions and metadata rather than the entire Web. Services differ from information. Services have a clearly defined function and operate on specific kinds of content. In this sense they are fundamentally different from the unstructured web pages one finds using Google and similar tools. We therefore put forth three arguments for using structured approaches to classifying services.

First, if the service domain is at least somewhat focused and domain experts can usefully structure it through the careful choice of terms and relationships, this structure can be browsed and enable newcomers to the domain to better understand it, a capability not possible through *ad hoc* tagging or free text search. Second, less formal approaches do not capture synonyms or concept relationships, while taxonomies can easily be extended to capture this important information, thus potentially improving discovery results through "semantic search" tools [6]. Third, a service catalog is not a set of arbitrary Web pages; it has an inherent structure in terms of a number of description dimensions (e.g., see the Service Description Framework [SDF] outlined in the Enterprise Service Strategy [3]). These dimensions of structure, including any semantic dimensions and tag values, can be used collectively to attain more control over the search refinement process than

[4] We define an *ontology* as a detailed description of what exists in a domain (of interest), including the relationships obtaining between those entities. We view a *taxonomy* as a "tree" (i.e., hierarchy) of progressively more specialized concepts, easily obtained by selecting all of the "is a" relationships from a given ontology.

what can be achieved with a free text search. Such searches place a higher burden on the end user to formulate an “intelligent search” through the manner in which they assemble the search terms, as compared to a search that provides a variety of search dimensions with pre-selected terms.

What structured approach for Net-Centricity?

So, if we accept that a structured classification approach has potential advantages such as those outlined above, the challenge becomes one of creating a classification strategy that is feasible and appropriate for the NCOW/NCE domain. Among the key characteristics (and corresponding requirements) that define the domain are the following:

- *Very large*. Since NCE services will encompass virtually of all DoD’s capabilities, it seems clear that the set of services will grow very large over time. Moreover, individual services will evolve through multiple versions. These considerations in turn imply the need for powerful tools that enable unsophisticated users to navigate through a large body of services and associated classification data efficiently.

- *Many “federated” communities of interest (COIs).* Each community of interest acts as its own interest group and will have the authority to semantically “classify” services (and information) in any way that is most appropriate for that COI. In such an environment, a classification scheme that is intuitive to one user may make little sense to another who does not belong to that COI—and, as a result, be of little help in discovering the critical information or net-centric services needed to accomplish a mission. This implies that to be effective, any classification approach must accommodate and support multiple perspectives on the classification services.

- *Need for traceability and justification.* DoD develops and maintains a large body of policies and procedures that service developers and users must comply with. Accordingly, it is critical that the service classification approach aid in understanding and tracking why services were developed and what specific policies mandate or recommend their use.

The classification approach and prototype described in the rest of this paper aim to begin addressing these concerns and requirements.

3. A Two-Pronged Approach to the Classification of Services in the NCE

We propose a two-pronged approach to the classification of net-centric services[5] for service discovery and meta-data tagging in the NCE: taxonomic classification and faceted classification. By “taxonomic classification” we mean the entry of the name of a service at its appropriate place in a traditional hierarchical taxonomy that systematizes the domain to which the service belongs. A text messaging service, for example, could be classified as a messaging service within a larger communications service sub-tree of a more general networking tree or hierarchy of services. See Figure 1.

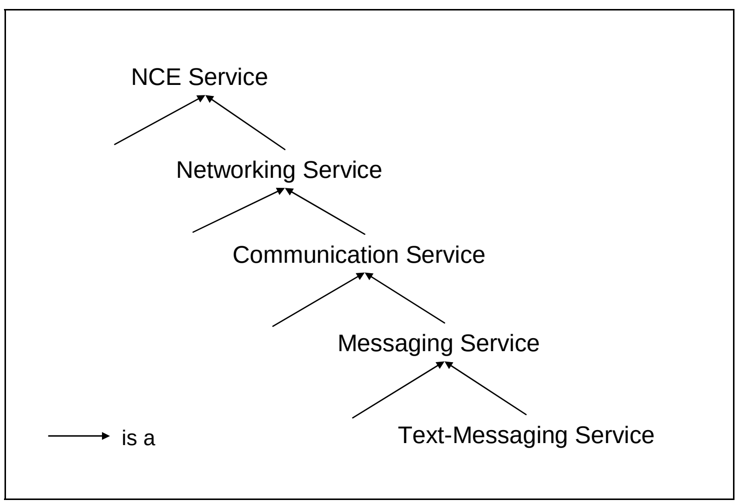

Figure 1. NCE Service Taxonomy Fragment

By “faceted classification” we mean the characterization of a service in terms of a set of attributes expressed as words or phrases. We call the faceted classification scheme used in the prototype (see below) the “7 Ws” scheme because, extending the familiar who-what-when-where-why paradigm, there are seven “W” questions used to classify a ser-

[5] Note that while this paper is focused on *services*, there is no reason why the approach cannot be applied to the meta-data tagging and subsequent discovery of *information* in the NCE as well.

vice. We chose this paradigm as a very pragmatic, yet powerful approach because a service is fundamentally an action, and in natural language one describes an action using subject, verb, object, and adverbial modifiers reflected in the facets cited above. Moreover, our approach allows for other reasonable faceted classification schemes. For example, one might consider every service as one that supports a phase in an overall management process (e.g., planning, data collection, analysis, decision-making, execution, monitoring) or a stage of Colonel John Boyd's OODA (observation, orientation, decision, action) loop [7].

The answer(s) to each of the following seven questions—not every question need have an answer—comprise our "faceted classification" of an NCE service:

1. *Who* uses the service?
2. *What* does the service do?
3. *On what* does the service act?
4. *To whom* is the service generally directed?
5. *Where* is the service used?
6. *When* is the service used?
7. *Why* is the service used?

Note that the second question, "*What* does the service do?" addresses the functionality of the service *per se*. Accordingly, it should mirror a service taxonomy that appears in the strictly taxonomic part of our classification approach. Note also that the third and fourth questions—the third and fourth facets—mirror the direct and indirect objects of ordinary declarative sentences and, depending on the nature of the verb (or action), may not be applicable to every service. The *where* facet is intended to capture the typical operational sphere of the service, while the *when* facet is meant to highlight the usual temporal scope or breadth of the service, for instance, near real-time, near-term, long-term, strategic, etc. The *why* facet would be particularly useful if it referenced particular DoD directives, instructions, guidance, or other prescriptive or advisory documents that could put the service into the larger warfighting context.

The novelty of our approach is two-fold. First, we admit the appropriateness of having multiple distinct taxonomies available for service classification. In other words we readily admit that no one taxonomic scheme will be adequate to classify the large variety of services likely to become available in the NCE. Even if one all-encompassing taxonomy were to be devised, it is unlikely that everyone within the Department of Defense would accept it. Second, the possible values for each of the facets used in the faceted classification scheme may themselves derive from a taxonomy appropriate to the domain of that facet. To take a simple example, to assign an appropriate value that answers the question, "Who is the (typical or most likely) user of the service?" we offer the user an organizational taxonomy, in effect, an "organization chart," from which the most likely user (organization) – or users or organizations – can be selected. And, as in the case of the purely taxonomic dimension of our proposed approach, multiple taxonomies will most likely be available from which facet values may be selected[6]. A variety of organizational taxonomies should be available, reflecting different organizational levels and types (e.g., bureaucratic, functional, etc.).

Overall, our approach is to take advantage of as much structure as possible that is already available from existing (and future) taxonomies, rather than attempting to invent a new language of our own. We argue that this approach will both increase the chances of its adoption in the community and offer benefits to users beyond those provided by a free text search through a less structured service catalog. If service developers and users have tools of sufficient sophistication to navigate and apply both types of classification structures, it will be easier to understand the mission domains to which the services apply, as well as to browse and search them.

More formally, we define a taxonomic classification of a service *s* with respect to a taxonomy Τ as follows. Let *SC* be the set of all service classes (i.e., nodes of a service taxonomy tree). Let *S* be the set of all services. An element *sc* of *SC* is a 2-tuple $\langle SC' \subseteq SC, S' \subseteq S \rangle$ such that $sc \notin \mathrm{pow}(SC')$, where $\mathrm{pow}(I)$ denotes the power set of *I*

[6] A service can also be classified in terms of facets that are not hierarchically structured like a taxonomy (e.g. a scalar number, a date, a developer name), but the implementation of this is straightforward and not considered here.

(i.e., *sc* is the root of a tree) and $(\forall s \in S')(\neg\exists\langle SC, S\rangle \in \text{pow}(SC') \mid s \in S)$ (i.e., a service is classified at most once in a taxonomy). We may now define a taxonomy as a set of service classes $SC' \subseteq SC$, and a classification of a service as a three-tuple $\langle \mathsf{T}, sc \in SC, s \in S\rangle$, stating that service *s* is classified as a member of service class *sc* in taxonomy T. If service *s* were—as is quite likely—also classifiable within another service taxonomy $\mathsf{T}' = SC'' \subseteq SC$, then $\langle \mathsf{T}', sc \in \text{pow}(SC''), s \in S\rangle$ characterizes *s* in terms of this second taxonomy: Service *s* is classified as belonging to some service class *sc* of T'. We envisage an indefinite number of service taxonomies available for the appropriate taxonomic classification of services. Most of these service taxonomies will be devised by COIs.

In the prototype we use (parts of) several currently available service taxonomies. The Navy's Common Services Function List (CSFL) provides a reasonably well-organized and documented set of 1,025 functions, taxonomized under three major headings (combat, infrastructure, and business support).[7] The *actions* and *capabilities* taxonomies of DoD's core taxonomy [8] were also selected to illustrate our proposed approach in the prototype.

With respect to faceted classification, the formal representation of the classification of a service *s* in terms of a faceted classification *FC* can be defined as follows. Suppose *FC* has a set of *n* facets $F = \{F_1, F_2, \ldots, F_n\}$. Remember that we can regard each facet as a taxonomy, and that we classify a service using zero or more terms from each facet. We can therefore represent the classification as a set $\{\langle f_1, n_{f_1}, s\rangle, \ldots, \langle f_m, n_{f_m}, s\rangle\}$ where $f_i \in F$ and $n_{f_i} \in \text{pow}(f_i)$. For instance, within the "7 Ws" scheme, a service *s* might be classified using the set:

$$\{\langle F_{\text{Who}}, \text{Defense\_Org}, s\rangle, \langle F_{\text{Who}}, \text{Government\_Org}, s\rangle, \langle F_{\text{Where}}, \text{CONUS}, s\rangle\}$$

[7] The CSFL was adopted to represent the services architecture within DoD's Federal Enterprise Architecture (www.feams.gov).

which means that service *s* has two likely categories of users, and is used in or has a geographical scope of the continental United States.

4. The Prototype

Our prototype contains functions for both tagging and discovery of NCE services. Services can be discovered interactively by a human user or by a software agent that invokes the service discovery tool as a service in its own right. Services can be discovered by judicious navigation through the hierarchy of a taxonomy, by facet-based searches, or both. The tool is also designed to enable a service producer to provide the other meta-data prescribed by the DoD ESS (name, textual description, developer, etc.). And, of course, this additional information is also provided to the user who is using the tool for service discovery.

The user initiates the prototype application and creates or opens a project. Each project consists of:

- A set of services the user wants to classify. This set may be empty if the user is only interested in service discovery.
- A set of taxonomic classifications the user deems relevant for classification or discovery.
- A set of faceted classifications the user deems relevant for classification or discovery.
- Classification information the user creates that relates services to the taxonomic and faceted classifications.

Each service, taxonomic classification, and faceted classification is uniquely identified by a Uniform Resource Identifier (URI) [9]. Currently the prototype simply asks the user to enter a textual URI. Eventually, we envisage a single and logically centralized (albeit replicated) directory of available services that the tool will link to automatically.

After the user (or agent) opens a project, the prototype displays the window similar to that shown in Figure 2. The pane on the left contains the set of services in the project; the user selects a particular service using the drop-down list in the upper left. The rest of the left pane shows information about the service. This information includes its name, a textual description, and its URI. We also foresee the need for information on how to invoke and use the service, plus additional information DoD might mandate. The prototype should incorporate tools to help users create this information, but currently it simply asks for two URIs: first, of a document containing a Web Service Description Language (WSDL) specification of a service's syntax [2], and second, of a document containing a DoD Discovery Metadata Specification (DDMS) [10]. These URIs are placeholders to suggest future functionality.

The right pane is split into two sub-panes. The top pane contains the taxonomic classifications in the project; the bottom pane contains the faceted classifications. The user can add classification schemes and services to, or remove them from, the project at any time.

Figure 2 shows how the user can classify a service with respect to a particular taxonomy.

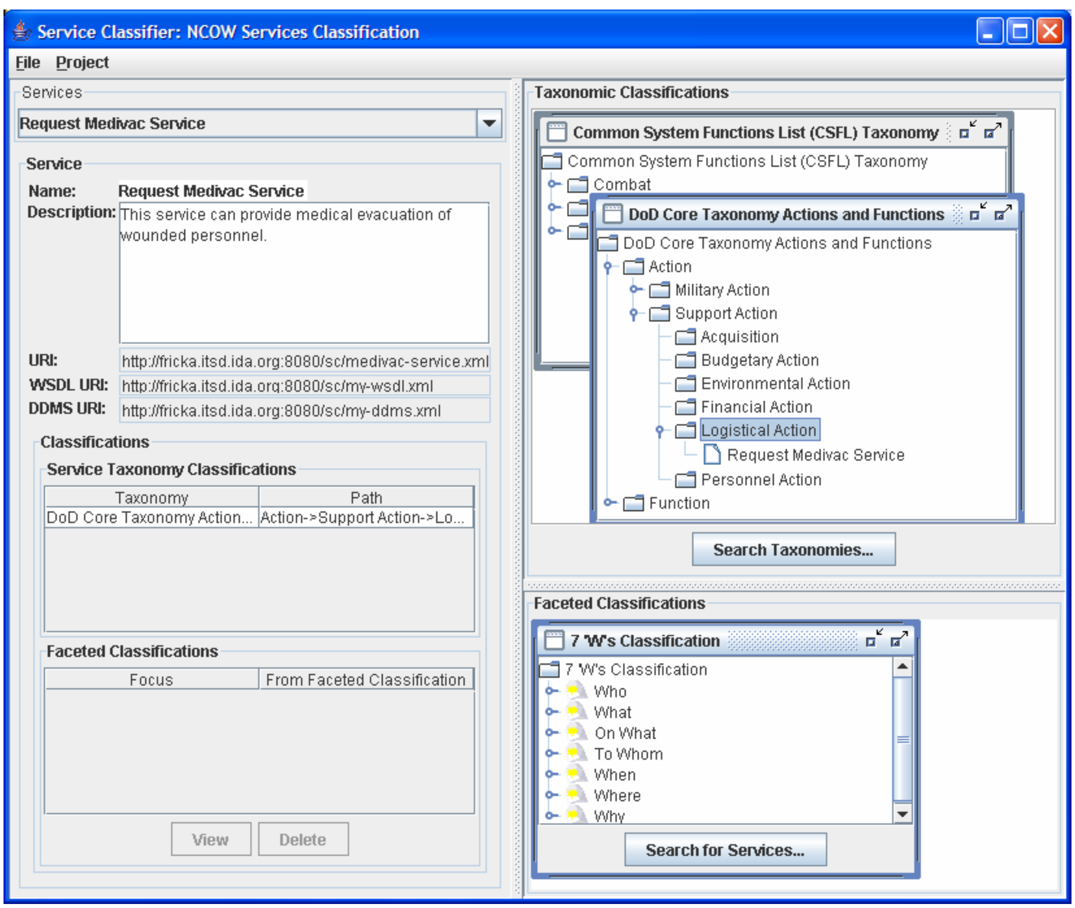

Figure 2. Taxonomic Classification of a Service

Here, the user has indicated that the Request Medivac Services is a Logistical Action in the DoD Core Actions and Functions Taxonomies. The prototype displays this fact both graphically (in the taxonomy tree in the upper right pane) and textually (in the Service Taxonomy Classifications table in the left pane).

Figure 3 shows how the user can tag a service using terms from the faceted classification. The user has chosen the “Military forces organization” term from the “Who” facet and the “Protection” term from the “Why” facet. As we discussed in Section 3, it is not necessary to assign terms from all facets, so the prototype is showing a valid faceted classification of the Request Medivac Services.

These two screenshots have illustrated some of the actions a user may take to classify a

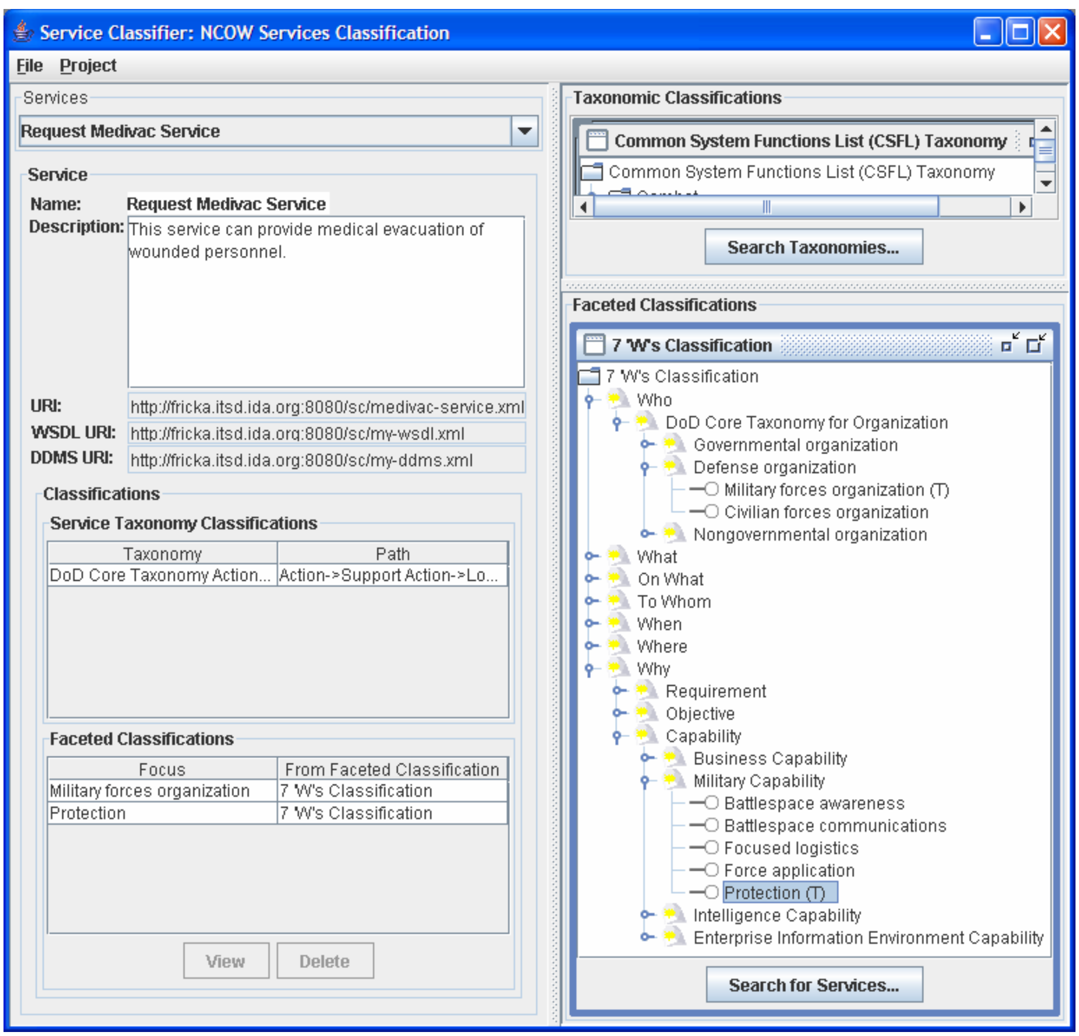

Figure 3. Faceted Classification of a Service

service. The user may also extend the classification hierarchies, adding new service classes to a taxonomy and new subfacets to a faceted classification. The prototype also allows the user to define synonyms in the faceted classification so as to increase the probability that a search by facets will yield a result.

The prototype currently saves classifications, extensions, and synonyms that a user defines in a local file. Ultimately, the tool would help the user publish the information in some globally accessible location. In this way, the user would propagate semantic information about a service.

The prototype supports discovery as well as tagging. Figure 4 provides an example. The user has searched known taxonomies (those displayed in the upper right pane) for ser-

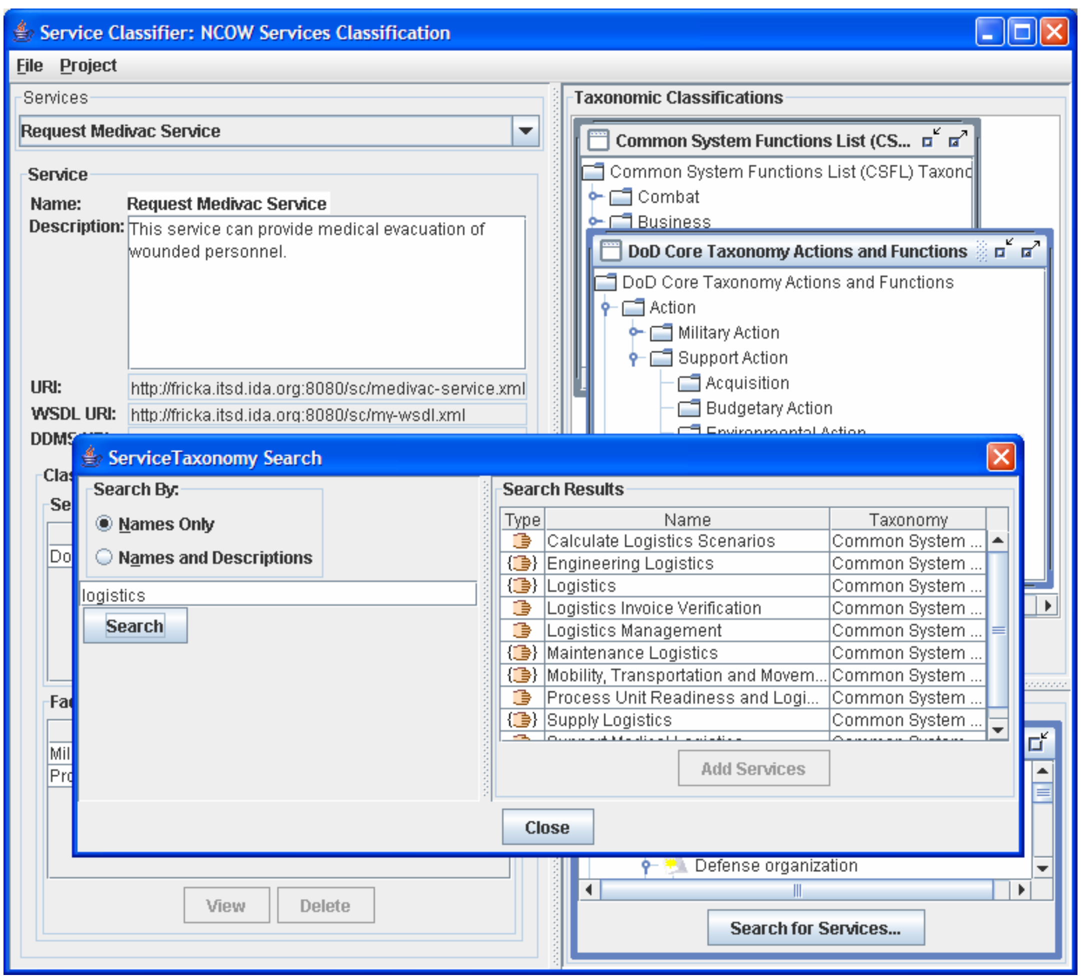

Figure 4. Service Discovery through Taxonomies

vices whose names contain the word "logistics," and the application has found and displayed ten such services. The prototype permits the user to add these services to the left pane for further examination and perhaps additional classification. In this way the user can continue to add semantic metadata to a service.

The prototype also supports searches using faceted classification terms. This is a controlled vocabulary search in which the user, presented with a fixed set of terms, chooses the subset of them that seem most relevant to his needs. Figure 5 shows an example. The user has performed a search using two terms: "Intelligence organization," from the "Who" facet, and "Theater," from the "Where" facet. The prototype has listed all services that are used by intelligence organizations and/or operate in theater. Once again the user can now examine each discovered service more closely and perhaps further classify it.

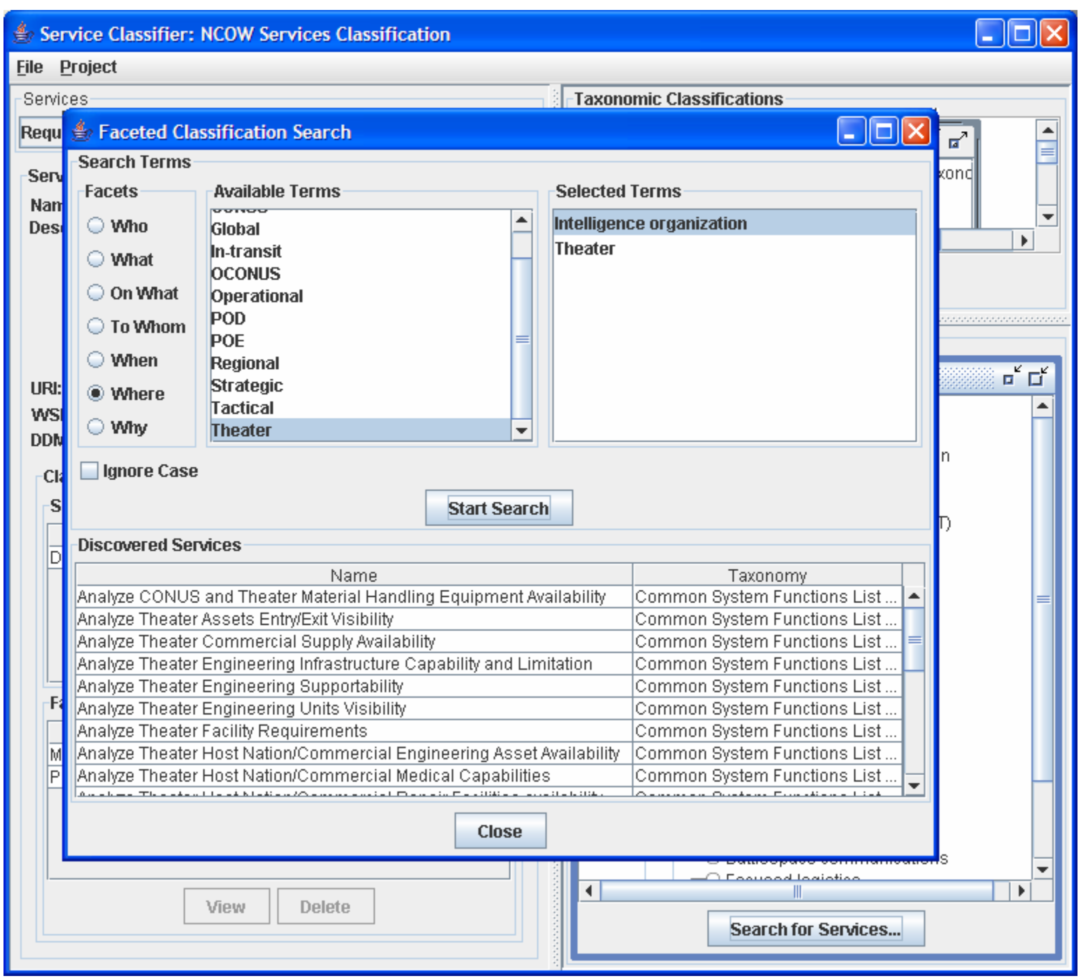

Figure 5. Service Discovery by Faceted Classification

The currently implemented searching capabilities are basic. A real-world application would let the user choose between conjunction and disjunction of terms, for example. However, the prototype's capabilities serve to demonstrate our vision of classification, discovery, and their interconnectedness.

5. Summary and Plans for Future Work

In summary, we presented an NCE service classification approach and described a prototype tool that promises to make more effective use of NCE services. The approach is powerful, yet flexible, and is consistent with the federated COI perspective.

In terms of future work, we hope to demonstrate how the prototype can be invoked by an intelligent software agent to automatically return the name of (and eventually invoke) a service that has been "discovered" as sharable within the NCE. In addition, we will explore implementing an unsupervised learning algorithm that enables the prototype to offer users candidate selections based on its tracking of the general user community's most successful historical search (navigation) experience. Not only will this approach facilitate efficient discovery and good service tagging, it will also allow NCE managers to easily determine which taxonomies and/or faceted classification schemes have little if any value to the NCE community due to minimal use by the larger community of users.

6. Acknowledgments

We would like to thank Mr. Terry Mayfield of the Institute for Defense Analyses for his valuable critique of the central ideas of this paper. This work was motivated by related work in support of the Office of the Assistant Secretary of Defense (Networks & Information Integration), DoD CIO, Architecture and Interoperability Directorate. The views expressed herein are those of the authors and should not be construed as those of, or endorsed by, the Department of Defense.

7. References

[1] Lichtblau, Dale E., "The Critical Role of Intelligent Software Agents in Enabling Net-Centric Command and Control," 2004 CCRTS, Command and Control Research and Technology Symposium, San Diego, CA, 15–17 June 2004.

[2] Web Services Description Language (WSDL), http://www.w3c.org/ws/desc.

[3] "Department of Defense Enterprise Services Strategy," draft of 17 November 2005, prepared by the DoD CIO.

[4] Shirky, C. "Ontology is Overrated: Categories, Links, and Tags," http://www.shirky.com/writings/ontology_overrated.html, 2005.

[5] Flank, S. "Why Taxonomies are Doomed," *DataStrategy Consulting*, 2004.

[6] Guha, R., McCool, R., and Miller, E., "Semantic Search," http://www2003.org/cdrom/papers/refereed/p779/ess.html.

[7] Hammond, G. *The Mind of War: John Boyd and American Security*, Smithsonian Books, 2001.

[8] http://dod.metadata.mil.

[9] Berners-Lee, T., "Uniform Resource Identifiers (URI): Generic Syntax," RFC 3986, January 2005, http://www.gbiv.com/protocols/uri/rfc/rfc3986.html.

[10] DoD Discovery Metadata Specification (DDMS), http://diides.ncr.disa.mil/mdreg/user/DDMS.cfm.