nCompass Service Oriented Architecture for Tacit Collaboration Services

David Schroh, Neil Bozowsky, Mike Savigny and William Wright
Oculus Info Inc.
{dschroh,nbozowsky,msavigny,bwright}@oculusinfo.com

Abstract

nCompass is a flexible, Service Oriented Architecture (SOA) designed to support the research and deployment of advanced tacit collaboration technology services for analysts. nCompass allows a significantly larger number of individual analytic capabilities, applications and services to be integrated together quickly and effectively. Service integration results are described from several computational tacit collaboration experiments conducted with open source intelligence analysts working with open source data. Key to nCompass is the technical framework and unique analytic event logging schema that supports context sharing across diverse applications and services. It is by combining the analyst with shared context across multiple advanced computational capabilities in a system of systems that a breakthrough in collaborative open source analysis can be achieved. This paper introduces the nCompass framework and integration platform, describes key nCompass core services, and provides results on functional synergies achieved through technology service integration with nCompass.

1. Introduction

In response to – and with feedback from – researchers investigating tacit collaboration computational services for Open Source Intelligence (OSINT) analysts, we have designed and developed a flexible, component-based Service Oriented Architecture (SOA) for the goal of supporting the research of synergistic advanced technology services for analysts, and the subsequent deployment of these advanced capabilities. This framework – its design, specifications and implementations – is called nCompass.

Key to nCompass is the technical framework and analytic event logging schema to support context sharing across applications and services, including user modeling, information and expert recommendation, computational linguistics, reasoning and image processing services. It is by combining the analyst with shared context across multiple advanced capabilities from a wide variety of third parties in a system of systems that a true breakthrough can be achieved.

In the next section we present our objectives in designing and implementing the nCompass Service Oriented Architecture. In Section 3 we describe technologies that were leveraged to achieve these objectives. Section 4 provides a tacit collaboration scenario to illustrate the use and impact of nCompass integration and new analytic event services. Section 5 introduces nCompass as both an architectural framework and a reference implementation, with descriptions in Section 6 of core services necessary to support our approach to tacit collaboration. In Section 7, we review experimental OSINT analysis uses of nCompass, and resulting impacts on service integration, experiment design, and tacit collaboration through context-sharing. Section 8 briefly discusses related work, and in Section 9 we conclude and suggest future work.

2. Objectives

To enable quick and effective integration of multiple individual computational analytic capabilities, we set out, with iterative feedback from other research teams, to design and implement the nCompass SOA framework and integration platform. The objective of the nCompass framework was to create a single unified environment into which third party analytic components could be easily integrated to produce new computational collaborative services. The integrated system would provide applications and services for tacit collaboration, suitable for OSINT challenges in exploiting massive unstructured data that is common to the domain.

The integration framework needed to provide access to functions and data without overhead costs, allowing efficient transfer of information among components, as well as customization of capability by specific end-user organization and by analyst. The framework also needed to support advanced visualization and analyst activity capture. Analytic event logging was implemented to support context sharing across applications and services. Every third party component contributed via logging to the analyst context. Every third party component had access to the common pool of activity logging. This enabled improved analytic collaboration and information sharing among Web 2.0 applications and services that were particularly suitable for OSINT analysis.

Within the widely distributed and diverse organizations, information can be difficult to discover or access. Analysts "don't know what they don't know" [1]. It is also difficult to discover other analysts with expertise and insights relevant to the task at hand. A key goal of nCompass in providing a framework to enable improved information awareness, was to not impose on the analyst any additional procedural or cognitive strain. The approach to achieving this goal was through tacit collaboration. In contrast to explicit collaboration, tacit collaboration allows analysts to discover important information and relevant or complementary expertise which they are unaware of, based on their actions and the actions of other analysts and computational services that identify similar interests and similar information.

3. Technical foundations

3.1. Oculus nSpace

To define requirements for the nCompass framework, we leveraged functional integration experience with the Oculus nSpace system of systems for OSINT analysis [2]. nSpace is a visual analytics work environment for unstructured data with novel information triage, evidence marshalling and sense-making capabilities. It is implemented with an Ajax browser front end with multi-tier computational and data services. nSpace also serves as a functional integration platform that marshals a variety of third party tools and data sources for analysts. Using open web services interfaces and protocols, nSpace integrates computational resources such as reasoning services, agent-based modeling and advanced computational linguistic functions, including entity extraction, supervised and unsupervised clustering, and automatic ontology construction. As a test bed for new technologies, nSpace is a platform enabling analysis science. The impact of new technologies deployed into the nSpace platform can be seen through features such as side-by-side comparison of results from alternative tools, an integrated workspace to perform experiments in a whole analytic workflow context, and a workspace to move seamlessly between components without loss in data or task context.

nSpace uses open web services interfaces and protocols [3] for components to exchange data about information objects critical for the process of analysis (e.g. hypotheses, evidence, models). This experience with functional integration led the way for the nCompass approach.

3.2. Service Oriented Architecture (SOA)

In a Service Oriented Architecture (SOA), the capabilities within an application are exposed as services [4] [5]. Each service is autonomous, reusable, stateless and discoverable. Capabilities within an application that are suitable to be services correspond to a strong business activity or recognizable business function. Services should be coarse-grained, and reusable, and suitable for well-defined interfaces. This allows existing capabilities to be easily recomposed into new applications to solve unanticipated problems.

Components in a SOA should be loosely coupled, highly interoperable, and have platform- and development technology-independent access mechanisms. Web service standards define ways for applications to compose themselves of coarse-grained, reusable components with well-defined interfaces. Standards for message packaging are often coupled with machine-processable interface descriptions such as Web Service Description Language (WSDL) to make these services more easily consumable.

These characteristics led to the adaptation of a Service Oriented Architecture as the basis for the nCompass framework. For the enterprise, standardizing on a SOA would allow the leveraging of existing capabilities within the organization. Enterprise IT would also be able to draw on new research, knowing it can be integrated quickly. For R&D, a SOA would allow the quick exploration and assessment of combinations of new computational capabilities. A SOA approach would also provide the R&D community a roadmap for transition from research into production. The use of a standards compliant SOA platform would give developers a single way of packaging capabilities for multiple customers and solutions. It was our belief that one-time investment in the engineering required to work in a SOA would reduce the cost for researchers to collaborate and explore multiple technology synergies on an ongoing basis. As discussed in Section 7, experiments sponsored by the IARPA Incisive Analysis program [6] demonstrated that nCompass allows a significantly larger number of individual analytic capabilities, applications and services to be integrated together quickly and effectively.

3.3. Tacit collaboration approach

To improve information awareness through tacit collaboration, the nCompass SOA framework needs to support three core technical capabilities. Analysis Modeling Services use captured indicators of analytic activity to build models of the user's analytic context and discern the user's information needs. Information Modeling Services build models of information objects and their relationships to each other. And finally, context-aware services use these models of the information space, and user models, to enhance the analyst's information awareness [7].

Previously, the Pacific Northwest National Lab (PNNL) developed the GlassBox environment for instrumentation of analyst workstations to log user activity [8]. This software was designed to capture low level detail, such as mouse clicks and keystrokes. Creating robust heuristics to infer high level analytic activity from these low level events was difficult and error prone. A resulting design objective for nCompass, therefore, was to incorporate a new framework for capturing higher-level, more meaningful indicators of analytic activity to be made available to user modeling services.

The nCompass SOA framework has been integral in connecting computational services that improve information awareness through tacit collaboration services. Document recommendation services use models of analytic context to search for information that corresponds to the user's information needs. Adaptive information retrieval services provide re-ranking of search results, presenting high-value information to the analyst based on an evolving user model. User modeling services match models of user context to one another to make recommendations of relevant or complementary expertise.

Web 2.0 technologies that leverage social networking and crowdsourcing are directly applicable to OSINT analysis. Social bookmarking [9], information recommendation engines [7] and web mining applications [2] are some of the OSINT tacit collaboration technologies that have been loosely coupled through the collection of indicators of analytic activity and interest by the nCompass SOA platform.

4. Scenario illustrating use and impact

The following scenario demonstrates how four OSINT analysts, separated both in time and space, are able to share work in progress, discover important new information, and discover each other through tacit collaboration. This scenario was used in an nCompass integration experiment. All names are fictitious.

- Emily Baker is a senior OSINT analyst and an expert in nuclear proliferation. Today she is investigating possible transfers of nuclear technology from CountryA to CountryB. She has created a network of key players in her nSpace2 Sandbox and now digs deeper by issuing a query in TRIST to retrieve documents about "proliferation" and CountryA's "PersonX."
- User modeling capabilities have been subscribing to her analytic activities – searches conducted, documents exploited, markups and annotations applied – via the Analysis Log Service (ALS). When she issues her query, the integrated system recognizes she is pursuing a new line of inquiry and automatically searches the repository for documents she hasn't yet seen.
- A document is recommended that refers to CountryC's Prime Minister. Emily learns that his son was involved in transferring dual-use technology to another country of interest through a corporation connected to an associate of PersonX.
- Emily updates her nSpace Sandbox with this new information, planning to next investigate whether there is also a connection to CountryB, but she is interrupted by an urgent request for a different tasking. Rather than suspend this current investigation, she places a note in her Sandbox and shares it with Adam Andersen, a junior analyst in her office.
- Adam sees what Emily has been thinking about a possible connection between CountryA and CountryB via CountryC. He queries for more information involving proliferation, PersonX, CountryB and CountryC, and finds a document that appears to confirm Emily's suspicions. He tags the document with keywords that aid him in organizing information of value.
- While Adam has been working, his activities have been logged and his user modeling service is automatically updated. The system now recommends another analyst who has been working on related tasks. Adam decides to follow up on the recommendation by contacting Gabriel Martinez, an imagery exploitation specialist at another organization. The system has alerted Adam that

Gabriel has been investigating smuggling of dual-use technology into CountryB, and Adam wants to learn more.

- At yet a third organization, Jason Risdal, a Mideast Affairs specialist, is working on determining which countries are supplying CountryB with components for uranium enrichment. His query results in a collection of retrieved documents and recommended analysts.
- At the top of the list is Emily Baker. Jason clicks on her name to access her Analyst Profile page, where he finds references to information with which she has recently been working. He's interested in the document with the CountryC connection, as well as the set of tags associated with it.
- Jason clicks on the tag "smuggling," and sees that Gabriel Martinez has tagged more documents with this keyword than any other analyst. Clicking through to Gabriel's Analyst Profile page, Jason learns of Gabriel's extensive experience investigating smuggling of weapons and weapons components. Jason contacts Gabriel, who may be able to assist in finding visual evidence of weapons components smuggling into CountryB.
- Jason has now been connected to Gabriel through the tag "smuggling," which he found associated with a document in Emily's collection, even though it was Adam, not Emily, who had applied the tag. Adam's routine action, done for his own benefit, provided the means for connecting Jason to Gabriel. Tacit collaboration is the key to enabling efficient insights such as this.

5. nCompass

nCompass is designed as a flexible, component-based SOA framework which supports the research of synergistic advanced technology for analysts, integrating reasoning services [10], agent-based modeling [11] and advanced computational linguistic functions [12] [13] including entity extraction, supervised and unsupervised clustering, and automatic ontology construction [14], and the deployment of these advanced capabilities. nCompass is an open, standards-based [15] SOA platform (Table 1) that is compatible with other enterprise SOA frameworks. The service interfaces are implementation agnostic, enabling deployment into existing enterprise architecture.

Table 1. Open web standards

Messaging SOAP 1.1 HTTP 1.1 WS-Addressing 1.0 MTOM 1.0 **Service Description** XML 1.0 XML Schema WSDL	**Service Publication & Discovery** UDDI **Security** HTTP Over TLS TLS 3.0 SSL 3.0 X.509

nCompass fulfills several roles and its design reflects these motivations. It is a research platform for managing the complexity of groups of services, allowing researchers who are combining efforts to spend more of their resources doing research. It is an experimental platform for collaborating with other researchers, exploring functional synthesis, prototyping concepts and reviewing with analysts. nCompass is also a deployment platform which includes a robust, tested reference implementation of key infrastructure services.

During deployment of new analytical services, nCompass is able to leverage enterprise SOA environments, integrating with existing enterprise SOA architecture and resources. The nCompass reference implementation also provides default capabilities, using open source SOA products that can be used as place-holders until specific enterprise products are in place. Default capabilities can also be used to minimize impacts on production systems until specific enterprise capacities are established.

6. nCompass core services

6.1. Analysis Log Service (ALS)

The Analysis Log Service (ALS) is a web service that collects and provides a repository for records of Analysis Log Events (ALEs). These logged events are captured high-level indicators of analytic activity and are used by Analysis Modeling Services to build models of individual and collective analytic context. These high level indicators are the foundation to this new approach for tacit collaboration services. The models of analytic context facilitate modes of tacit collaboration between analysts, and shared analytic context across applications that may not otherwise be integrated.

The design and usage of the ALS is based around a SOA infrastructure:

- Users interact with Applications, typically delivered in a web browser.
- Business Services deliver capabilities to Applications as web services.
- Applications, Business Services and User Modeling Services interact through open SOA specifications.

As users interact with Applications, these services report significant user events in the form of ALEs sent to the ALS, as shown in Figure 1. Business Services neither report nor directly consume ALEs, but obtain all information about the context of analysis by querying Analysis Modeling Services.

User Modeling Services obtain information about analysis activities and behaviors primarily by issuing requests to the ALS. They issue requests on whatever schedule and frequency most appropriate to the particular dimension(s) of analytic context they endeavor to model. User Modeling Services then supply their models to context-aware Business Services. User Modeling Services consume ALEs from the ALS, but do not (typically) supply ALEs back to the ALS [16].

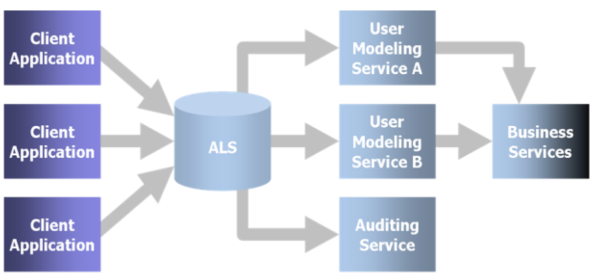

Figure 1. Analysis Log Service (ALS) architecture

6.1.1. Analysis Log Events (ALEs) The ALS supports an open specification in which each Analysis Log Event consists of a Message schema, and a single Event element which contains one or more Object elements.

The Event element is defined in a taxonomy of Event classes such as Search, Assess and Retain. In specifying these classes, the overarching goal is to define the smallest set of classes such that records of events of a given class are consumed by at least one Analysis Modeling Service, or records of events of a given class are required for metrics assessment and evaluation purposes. The objective is to provide sufficient precision in differentiating between events for analysis modeling services to produce effective models of analytic context, without rendering the taxonomy intractable and too difficult for model producers to use. Each event class is extendable, so that event producers can generate events that do not fit exactly into the taxonomy, or that can contain data that is not part of the standard event definition, without being bound by the specification governance process.

Each Analysis Log Event message contains Objects of various kinds. The Object class taxonomy defines Entity types along with their internal structures. There are a small number of base classes for different kinds of Objects, and some auxiliary types or classes. Entity is the base object that Resources, Relations and other Objects all inherit from. Resources include a wide variety of objects that might be transferred, excerpted, annotated, etc., including all sorts of text documents, images, audio files, and video data. All Object types have an extensions element that can contain service or application specific XML.

Analysis Log Service design and usage is predicated on the availability of a persistent store of information objects accessible across the enterprise. Information referenced in ALEs cannot be trapped within individual applications, but instead must be externally available for use by other services and applications.

6.2. Content Management Service (CMS)

The Content Management Service (CMS) is another core nCompass service that provides access to a persistent data store required to support the Analysis Log Service. It consists of a standard Simple Object Access Protocol (SOAP) web service interface to COTS and GOTS content and document management systems and can be used to store information objects ranging from individual arguments by a particular user, to large document collections. Applications can store their information objects in the CMS, either natively or for archival purposes. The CMS interface allows for retrieval of those artifacts, and discovery through keyword search.

The CMS SOAP interface supports standard operations such as Store and Retrieve. Content can be stored into the CMS by providing document metadata, as well as the document content. Each object stored in the CMS is assigned a unique identifier, which is returned when the object is stored. It is this CMS ID that is used to retrieve the content, as well as share it with others. The CMS supports the SOAP Message Transmission Optimization Mechanism (MTOM) specification [17] for retrieving content as binary attachments, rather than encoded text. This greatly increases the performance of using a web service to transmit and store binary data.

Metadata can be associated with an object upon storage, such as author, title, or Uniform Resource Locator (URL). Arbitrary fields are also supported. A metadata search service interface allows for searching on metadata. Optional additional modules are also supported, including a keyword search service and utilities for importing, exporting, indexing and processing the contents of the CMS.

Indexing content for keyword search is an optional feature in the CMS. Content added to the CMS's data store can be flagged as being suitable for keyword indexing. Logical collections can be configured to use the internal keyword indexer, based on the open-source Apache Lucene indexer, or can be configured to use an external, third-party indexer, such as a Google Search Appliance [18] or IBM OmniFind Yahoo Edition [19]. Logical collections that represent enterprise content services that support their own indexing will pass search requests directly to those services.

The CMS also supports a Representational State Transfer (REST) interface, both for retrieval and for discovery. For retrieval, all objects in the CMS can be retrieved via a Universal Resource Indicator (URI) for discovery. The CMS supports the OpenSearch 1.1 specification for information retrieval.

A plug-in architecture allows a single CMS instance to support multiple virtual content collections. Each named virtual collection can be stored in the same physical repository, or they can be located across multiple vendor implementations, all accessible through a single web service interface. This same plug-in architecture allows multiple implementations of content stores to be accessible through the same CMS interface. Implementations of the CMS interface support open source content stores including the XML content store eXist, and the relational database Apache Derby, as well as adapters for commercial products such as Oracle and MarkLogic.

For the researcher, or small-scale production environment, the CMS provides a standard way to store documents, data, metadata, annotations and other content in a repository that is searchable and from which every object can be retrieved as a URI. For the enterprise, the CMS can provide a data store that scales up. The adapter framework allows existing enterprise content repositories to be made available through the CMS interface so that existing capabilities that use the CMS can leverage enterprise resources with little or no additional modification.

Both the Content Management Service and the Analysis Log Service depend upon the availability of a unified authentication scheme. This is required to provide a common database of users, trusted communication between services, and to enable the controlled flow of information across application boundaries.

6.3. Authentication Management Service (AMS)

The Authentication Management Service (AMS) provides a common web service for managing credentials and user attributes across applications and services. The service presents a standard workflow for web applications that require authentication. By using a common authentication system, applications and services share a common database of users, and analysts do not have to log into multiple different systems. A verified user identity is also essential in providing accurate user modeling and tacit collaboration services.

The authentication service is not just for analysts; it is also for services. Just as users must authenticate with an application, so must that application authenticate with services that it consumes. This allows for trusted communication between services, and protection of information transmitted.

Using standard SOAP and REST access methods, users can be authenticated by either username and password against the web service, or through a client-side X.509 digital certificate. Once authenticated, applications and services identify the user by an authentication token. In the case of password-based sessions, this token is generated by the AMS. For TLS and SSL server configurations that support client-side digital certificates, this token is based on the Distinguished Name (DN) found in the certificate.

This token is passed along to other applications and services to represent the authenticated session. Downstream services can check the token, to ensure it is valid. The token can expire, or be revoked, ending the user's session. It is this standard authentication token that allows for single sign-on between applications.

The AMS also provides a user management interface for viewing common user attributes shared across applications. Behind the AMS web service can exist any proprietary authentication store; the reference implementation allows for a text file, relational database or LDAP directory. By implementing the service interfaces that are part of the specification, other authentication stores can be used.

6.4. Group Management Service (GMS)

The Group Management Service (GMS) provides a common web service to define groups for access control and social networking. The goal of the GMS is to bring the definitions of groups of users, and their roles, outside the boundaries of individual applications and services. Externalizing group definition from individual applications allows for management of group- and role-based access control. It also enables end users to create and manage their own Communities of Interest (COI) that cut across application boundaries. Externalization of groups, such as the standard ”friends” list, enables a broad variety of new social software to be brought to bear on analytic challenges.

The GMS consists of a SOAP interface that defines methods for creating groups, adding users to groups, defining user roles within groups, building hierarchies of groups, and setting access control lists (ACL) for management of these groups. These access control lists allow for the creation of public, private, and semiprivate groups. A group can be set public, so that anyone can join. Alternatively, a group can be semiprivate, requiring an invitation to get in, which anyone already in the group can provide. Finally a group can be private, and only the moderator of the group can add new members.

The GMS allows for manipulating the ACL of each group, for the purposes of configuring COIs. Mandatory access control on information objects, such as due to sensitivity or privacy concerns, can be performed by applications based on group definitions. If information objects are stored in the nCompass Content Management Service (CMS), access control can be configured there.

The GMS also supports the Group and People portions of the REST API from the OpenSocial specification. This specification defines a common API for social applications. There are many websites implementing OpenSocial, including Engage.com, Friendster, hi5, Hyves, imeem, LinkedIn, MySpace, Ning, Oracle, orkut, Plaxo, Salesforce.com, Six Apart, Tianji, Viadeo, and XING [20]. The GMS supports the retrieval of an analyst's “friends” list through this API, allowing existing applications that support OpenSocial to quickly support the GMS. The GMS extends the OpenSocial REST interface to support other group management operations.

Together, the Authentication Management Service and Group Management Service enable users and services to authenticate with security across multiple applications. This is a necessary capability to support authorized access for a distributed network of users, information objects and tacit collaboration services.

7. Experiments and results

The following describes integration experiences and results with three experiments that demonstrate the nCompass impact on system-of-system integration efficiencies for OSINT analysis work environments.

7.1. Ease of integration

Over the course of a two year period from 2006 to 2008, we participated as one of twelve research groups combining individual capabilities with the goal of producing an increase in analytic capabilities through tacit collaboration via a combined system of systems. nCompass served as the integration platform through the course of three integration experiments. The experiments were designed to investigate and demonstrate collaboration computational services to enhance analyst and system effectiveness, including information sharing through tacit collaboration, enhanced information value of items reviewed by analysts, and increased analyst effectiveness.

The first integration experiment demonstrated SOA integration of all 12 participating research capabilities focused on information sharing through tacit collaboration. The goal of the first experiment was to evaluate how well research capabilities, in a broad spectrum of maturity from concept through to ready-to-deploy, could be integrated in a Service Oriented Architecture, with a resultant benefit for the analyst. The first nCompass platform was deployed, consisting of a service bus and message framework. All of the research teams offered up their capabilities through either request-reply or event-driven web services.

The second experiment demonstrated increased functional complexity. Teams of researchers worked across their individual project boundary to leverage concepts from other researchers. Common service offerings were established in the nCompass platform, including the Analysis Log Service, Content Management Service, and Authentication Management Service. Functional integration focused on analysis logging for “analyst-aware” applications, data finding users instead of users finding data, social networking for tacit collaboration in analytic communities, and analysis auditing and review.

The third experiment consisted of an end-to-end solution using analyst modeling to enable expert recommendations, document and data recommendations, cross-tool and multi-modal shared analytic context, and enhanced information retrieval in an OSINT analysis scenario.

A key result confirmed by design specifications and project scheduling was that, while the second and third experiments each involved increasingly complex integrated systems of systems, the timeframes for planning, design, integration, and quality assurance grew progressively shorter from ten weeks to three weeks. The nCompass SOA proved critical in providing the capability to investigate, efficiently, breakthrough “combinatorics” of different research capabilities.

7.2. Impact on experiment design

In the fall of 2007, we supported the National Institute of Standards and Technology (NIST) in conducting an experiment focused on evaluating the software system known as Hydra 3 to demonstrate several key principles for enhancing analyst and system effectiveness, and set a baseline for further studies. Hydra 3 was an integrated system, that did not use nCompass, but combined the information triage and flexible analytic workspaces of Oculus nSpace, the entity extraction and document categorization capabilities [13] of the Fair Isaac Text Analysis Engine (TAE), and the natural language search and ontology generation [14] of Lymba's Power Answer and Concept Explorer (PACE).

The experiment consisted of two groups of users: a treatment group using Hydra 3, and a baseline group using Google search and Microsoft Word. Each group consisted of eight OSINT analysts who searched the corpus of approximately 30,000 documents harvested from the Internet, varying in type from news articles and digests, to blogs and forums. The task was an assessment of Country X’s political leadership influences, and the analysts were asked to use the Analysis of Competing Hypotheses (ACH) methodology [21] to produce a report using a specified, detailed Microsoft Word form. ACH is a tool for weighing alternative hypotheses, helping an analyst to minimize cognitive limitations and bias. Instruments included report ranking, logging, scaled and open ended questionnaires.

Experimental results on the impact of the Hydra 3 system on analytic effectiveness were inconclusive with differences in analyst reports difficult to assess. However, in terms of workload, analysts in the baseline group were observed to perform more queries rendering more results to scan through, while the treatment group significantly reduced the amount of work required to find and save information relevant to the task.

The PNNL GlassBox environment was used to instrument the analyst workstations, recording detailed information on usage of the technology. The logged GlassBox data served to verify proper logging of Analysis Log Events by the integrated Hydra 3 system, and to validate the use of ALE data for measuring characteristics of analytic activity.

The effectiveness of the experimental process in this earlier non-nCompass experiment was compared against that of other later experiments in which nCompass served as the integration platform. This earlier experiment did not take advantage of the nCompass SOA to facilitate open integration, and instead relied on multiple custom integration points. A key conclusion derived from the comparison of this experiment to others was that more time was spent on engineering and system testing, and less on experimental design and the research goals, when the experiment could not leverage the nCompass SOA [22].

7.3. Tacit collaboration through context-sharing

In the summer of 2008, as part of a focus on evaluating tools that provide user modeling capabilities, SET Corporation tested their User Modeling Service (UMS) [23] in the context of an nCompass integrated multi-component system. The experiment was developed to test three key areas of functionality: [24]

- Virtual Interest Group recommendations (data was captured online, and experimental recommendations produced off-line)
- Adaptive Information Retrieval (re-ranking of Google document results, with online user judgments)
- Document Recommendations (system generated queries based on the user's model, with online user judgments)

We generated a new version of the nCompass ALS to support sense-making analytic events in this experiment, and hosted the ALS, CMS and AMS for the experiment. We also provided two additional components. First the analysts used the nSpace2 web-based OSINT analysis environment as their workspace [25]. In this workspace they issued queries, organized information, and viewed documents. The queries issued in the nSpace2 environment were used during the Augmented Information Retrieval evaluation stage. Information organized in the nSpace2 Sandbox was used to establish a baseline for document recommendations. Events in querying, reading, and information organization were all used to build the user models.

A re-ranking web service collected search results on behalf of the user, submitted them for re-ranking, and populated rich web forms with the re-ranked documents to the analyst for judgment. For Augmented Information Retrieval, this service collected the last query issued by the user in nSpace2, submitted it to Google and collected 50 search results. These search results were passed to the re-ranking system to be ordered based on the user model. These reordered results were then mixed with the original Google results, and presented to the user in a rich form, again for user judgment of the effectiveness of the re-ranking

For Document Recommendations, the service collected a system generated query based on the user model, and submitted it back to Google. The search results returned were similarly submitted for reordering. The re-ranked results were mixed with a baseline generated at the start of the user's session, and presented to the user for judgments.

All searches, results, reordering, mixing, baselines, treatments, log files, and documents retrieved were recorded and archived during the experiment. The archives were made available to researchers for analysis.

With nCompass SOA as the underlying integration platform, researchers were quickly able to combine capabilities to effectively demonstrate breakthroughs from shared context across a combined system of systems. Integrating several existing and new analytic services into a powerful Web 2.0 application platform in the nCompass SOA environment saved significant integration time, compared to similar previous experiments, allowing researchers to focus on strong experimental design. Reported results of the experiment analysis indicated "significant impact of using user models to enhance finding better information in documents and in finding other people to work with" [26].

8. Related work

Context can be broadly defined as "any information that can be used to characterize the situation of an entity" [27]. Much work in the area of context-aware computing has focused on awareness of a situation in physical environments, with elaboration of associated data schemas [28] and technical frameworks [29].

With the Analysis Log Event schema, and the technical framework of the nCompass SOA and Analysis Log Service, we are endeavoring to provide *shared analytic context*.

Enhancing active collaboration sessions has been proposed through sharing tool state, and finding indicators of significant analytic events through examining textual messages between collaborating analysts [30]. In contrast, our goal is to enable tacit collaboration, without requiring explicit intent or intervention on the part of the user.

Some work has examined implicit human computer interaction, using physical sensors to provide context and alleviate the need for explicit input by the user [31]. Again, this work focuses on awareness of situation in a physical environment.

Our approach to enabling tacit collaboration relies upon unobtrusively capturing indicators of analytic activity as the user interacts with an instrumented workspace.

Prior to the work described in this paper, the Pacific Northwest National Lab (PNNL) developed the GlassBox environment for instrumentation of analyst workstations to log user activity [8]. This software was designed to capture low level detail, such as mouse clicks and keystrokes. Creating robust heuristics to infer high level analytic activity from these low level events is difficult and error prone. A resulting design objective for nCompass, therefore, was to incorporate a new framework for capturing higher-level, more meaningful indicators of analytic activity to be made available to user modeling services.

9. Conclusion and future work

Experiments have demonstrated the great potential in rich logging of analyst activity to support context sharing among users and across OSINT analysis tools to produce value-added information services. This is the key to enabling tacit collaboration, providing the analyst with improved information awareness without imposing any additional procedural or cognitive strain, and is particularly applicable to the Web 2.0 analytic tools emerging for Open Source Intelligence.

nCompass is designed to support the capture of meaningful indicators of analytic activity, and to allow a dramatically larger number of individual computational analytic capabilities, applications and services to be integrated together quickly and effectively. The nCompass SOA framework proved to be a key element in the success of researchers working to design solutions that increase analytic productivity and information value delivered to users through the use of larger frameworks of diverse, context-aware computational systems. It is by combining the analyst with shared context across multiple advanced capabilities in a system of systems that a significant improvement in analytic performance can be achieved.

Our planned next steps are to integrate with an existing analytic toolset to investigate the potential in modeling analytic workflow. At present, we have drafted an extension to the Analysis Log Event specification to support logging of workflow events in the analytic process. A prototype Analysis Log Service has been implemented to support this extended ALE specification.

An additional future direction of interest is implementing the Analysis Log Event schema in a machine-processable format, such as RDF (Resource Description Framework) or OWL (Web Ontology Language). This would enable us to explore the potential for machine reasoning over logged indicators of analytic activity.

10. Acknowledgements

This work was supported and monitored by the Intelligence Advanced Research Projects Activity (IARPA) and the Air Force Research Laboratory (AFRL) under contract number FA8750-06-C-0211.

The views, opinions and findings contained in this report are those of the authors and should not be construed as an official Department of Defense position, policy or decision, unless so designated by other official documentation.

The authors wish to thank the IARPA Collaborative Analyst and System Effectiveness (CASE) Program staff and AFRL staff for their support and encouragement.

11. References

[1] ODNI Information Sharing Strategy, February 21, 2008, p. 10.

[2] Proulx, P., L. Chien, R. Harper, D. Schroh, T. Kapler, D. Jonker, W. Wright, "nSpace and GeoTime – VAST 2006 Case Study", IEEE Computer Graphics and Applications, 2007.

[3] Whittenton, J., Editor, "Defense Intelligence Integrators Guide (DIIG)", version 1.2, February, 2007.

[4] Erl, Thomas, "Service-Oriented Architecture – Concepts, Technology, and Design", Prentice Hall, 2005.

[5] Krafzig, D., Banke, K., Slama, D., "Enterprise SOA Service-Oriented Architecture Best Practices", Prentice Hall, 2005.

[6] Intelligence Advanced Research Projects Activity (IARPA) Office of Incisive Analysis, http://www.iarpa.gov/office_incisive.html/.

[7] Brueckner, S., Downs, E., Hilscher, R., Yinger, A., "Self Organizing Integration of Competing Reasoners for Information Matching", In Proceedings of International Workshop on Environment-Mediated Coordination in Self-Organizing and Self-Adaptive Systems (ECOSOA'08), Venice, Italy, October 20, 2008.

[8] Hampson, E., Cowley, P., "Instrumenting the Intelligence Analysis Process", First International Conference on Intelligence Analysis Methods and Tools, MITRE, McLean, VA, May 2-6, 2005.

[9] General Dynamics Advanced Information Systems tag|Connect, http://www.gd-ais.com/index.cfm?acronym= tagconnect

[10] Bringsjord, S., Arkoudas, K., Clark, M., Shilliday, A., Taylor, J., Schimanski, B. & Yang, Y., "Reporting on Some Logic-Based Machine Reading Research". in Etzioni, O., ed., Machine Reading: Papers from the AAAI Spring Symposium (Technical Report SS-07-06), pp. 23-28, and also www.cogsci.rpi.edu/research/rair/slate.

[11] Hilscher, R., Sven Brueckner, Theodore C. Belding, H. Van Dyke Parunak, "Self-Organizing Information Matching in InformANTS", SASO '07: Proceedings of the First International Conference on Self-Adaptive and Self-Organizing Systems, July, 2007.

[12] Aarseth, P., M. Deligonul, J. Lehmann, L. Nezda and A. Hickl., "TASER: A Temporal and Spatial Expression Recognition and Normalization System", Proceedings of the 2005 Automatic Content Extraction Conference (ACE 2005), Gaithersburg, MD.

[13] Freitag, D., M. Blume, J. Byrnes, R. Calmbach and R. Rohwer, "A Workbench for Rapid Development of Extraction Capabilities", In Proc. International Conference on Intelligence Analysis, 2005.

[14] Moldovan, D., S. Harabagiu, R. Girju, P. Morarescu, F. Lacatusu, A. Novischi, A. Badulescu and O. Bolohan, "LCC Tools for Question Answering", Proceedings of the TREC-2002 Conference, NIST, Gaithersburg, 2002.

[15] WS-I Basic Profile, http://www.ws-i.org/Profiles/BasicProfile-1_2(WGAD).html/.

[16] CASE Analysis Log Service Specification, version 3.0, June 27, 2008.

[17] SOAP Message Transmission Optimization Mechanism, http://www.w3.org/TR/soap12-mtom/.

[18] Google Search Applicance, http://www.google.com/ enterprise/ gsa/.

[19] IBM OmniFind Yahoo Edition, http://omnifind.ibm. yahoo.net/.

[20] OpenSocial, http://code.google.com/apis/opensocial/.

[21] Heuer, Richards J., Jr, "Chapter 8: Analysis of Competing Hypotheses", Psychology of Intelligence Analysis, Center for the Study of Intelligence, Central Intelligence Agency, 1999.

[22] Morse, E., Sheppard, C., Grantham, J., Cheikes, B. and J. Boiney, "APEX System Evaluation", NIST and MITRE, October 30, 2008.

[23] Alonso, R., and H. Li, "Model-Guided Information Discovery for Intelligence Analysis", in Proceedings of CIKM '05, Bremen, Germany, 2005.

[24] Morse, E., "UMS Multi-Component Evaluation Plan", NIST, June 27, 2008.

[25] Chien, L., A. Tat, P. Proulx, A. Khamisa and W. Wright, "Grand Challenge Award 2008: Support for Diverse Analytic Techniques", IEEE Visual Analytics Science and Technology VAST 2008.

[26] Morse, E., "Recommender2 Evaluation: New Vectors/SET, Oculus, BAE, Fair Isaac", NIST, December 3, 2008.
[27] Dey, Anind K., "Understanding and Using Context", Personal and Ubiquitous Computing, Volume 5, Issue 1, February 2001, pp. 4 – 7.
[28] Chen, H. Finn, T. and A. Joshi, "An ontology for context-aware pervasive computing environments". The Knowledge Engineering Review (2003), 18, Cambridge University Press, pp. 197-207.
[29] Dey, Anind K. and Gregory D. Abowd, "The Context Toolkit: Aiding the Development of Context-Aware Applications", Proc. Conf. Human Factors in Computing Systems (CHI), ACM Press, New York, 1999, pp. 434–441.
[30] Hardisty, F., "GeoJabber: Finding Significant Analytic Events in Collaborative Visual Analysis Sessions", GIScience 2008, September 23-26, 2008, Park City, Utah.
[31] Schmidt, A., "Implicit Human Computer Interaction Through Context", Personal Technologies 4(2–3), June 2000, pp. 191–199.