

# CONTEXT AWARE INFORMATION RETRIEVAL FOR ENHANCED SITUATION AWARENESS

Ali Bahrami  
Mathematics & Computing  
Technology  
Boeing Phantom Works  
P.O. Box 3707, M/C 7L-70  
Seattle, Washington, 8124,  
U.S.A

Jun Yuan  
Mathematics & Computing  
Technology  
Boeing Phantom Works  
P.O. Box 3707, M/C 7L-70  
Seattle, Washington, 98124,  
U.S.A

Paul R. Smart  
School of Electronics and  
Computer Science,  
University of Southampton  
Southampton, SO17 1BJ  
U.K.

Nigel R. Shadbolt  
School of Electronics and  
Computer Science,  
University of Southampton  
Southampton, SO17 1BJ  
U.K.

## ABSTRACT

*In the coalition forces, users are increasingly challenged with the issues of information overload and correlation of information from heterogeneous sources. Users might need different pieces of information, ranging from information about a single building, to the resolution strategy of a global conflict. Sometimes, the time, location and past history of information access can also shape the information needs of users.*

*Information systems need to help users pull together data from disparate sources according to their expressed needs (as represented by system queries), as well as less specific criteria. Information consumers have varying roles, tasks/missions, goals and agendas, knowledge and background, and personal preferences. These factors can be used to shape both the execution of user queries and the form in which retrieved information is packaged. However, full automation of this daunting information aggregation and customization task is not possible with existing approaches.*

*In this paper we present an infrastructure for context-aware information retrieval to enhance situation awareness. The infrastructure provides each user with a customized, mission-oriented system that gives access to the right information from heterogeneous sources in the context of a particular task, plan and/or mission. The approach lays on five intertwined fundamental concepts, namely Workflow, Context, Ontology, Profile and Information Aggregation. The exploitation of this knowledge, using appropriate domain ontologies, will make it feasible to provide contextual assistance in various ways to the work performed according to a user's task-relevant information requirements. This paper formalizes these concepts and their interrelationships.*

## INTRODUCTION

Situation Awareness (SA) may be defined as:

*“The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future.” (Endsley, 1988)*

Situation awareness is recognized as a critical enabler for operational effectiveness in a wide variety of military contexts and is a central element of contemporary system design approaches (Endsley et al, 2003). For military operators, the awareness of the temporal unfolding of events in a situation, together with their understanding of the implications of those events for both ongoing activities and future plans, is clearly a critical element in the effectiveness and adaptability of a military force, but the attempt to yield enhanced situation awareness, especially in coalition contexts, is complicated by a variety of challenges. Firstly, military operators must deal with the inherent heterogeneity and distributed nature of information content. Information may be contained in multiple formats (e.g. images, video, unstructured text, XML) and will often be distributed across multiple data repositories or nodes within an information network. Situation awareness ultimately depends on an ability to deal with this heterogeneity - to aggregate, integrate and process task-relevant information in ways that support decision-making in an operationally-effective manner. A second challenge in the effort to enhance situation awareness concerns the task-variant information needs and requirements of those engaged in problem-solving activities (e.g. individual human agents or hybrid teams).

The battlefield network currently consists of many information sources, including in-theatre sensors, platforms, intelligence reports and remote information such as archival

intelligence and satellite data. In future coalition contexts, strategies for operationally-effective modes of information exchange and exploitation will need to target a wider variety of disparate information repositories and communication systems, including digital datalinks, military information repositories, and the totality of the information space available via both internet-enabled and peer-to-peer computing environments. The exploitation of this 'Infosphere' will require an ability to integrate information from multiple sources and to process information in ways that respect the original semantics of the information content. Inter-operability solutions, in this sense, are not just about the physical exchange of information; they also relate to the meaning assigned to information items and the interpretive biases possessed by coalition partners. The point is that once we encounter distributed network environments that subtend a wide variety of information systems, sources and user communities (as is often the case in coalition contexts), we face a critical challenge in terms of our ability to integrate and share information in a semantically-sensible manner.

### APPROACH

The information sources relevant to coalition operations are, in many cases, created without the linkages that support rapid information retrieval and decision-making. However, in order to provide a more efficient and effective working environment, aggregated and personalized information from those systems is highly desired and is a prerequisite for Network Centric Operations. Military data systems are heterogeneous on many aspects, including: hardware and software platforms, process & information models, information access languages, query processing capabilities, etc. A generic approach needs to be established to manage all involved metadata and process models as well as mapping semantic knowledge among different pieces of information and its consumers. More importantly, process models, information consumers, and information resources along with their capabilities may change dynamically over time. Failure to make adaptive adjustments on information aggregation, in response to changes, may result in performance shortfalls in operational missions. Our approach is to leverage knowledge about

user preferences, information domains, workflow and context-aware information retrieval from heterogeneous data sources. Specifically, in order to manage the scope of the vast possibilities inherent in context representation and to focus this effort on a useful demonstration of context-influenced information integration, we propose to demonstrate a framework for context-aware information retrieval that avails itself of user profiles, process workflows, source availability monitoring, and real-time semantic integration. Semantic integration is generally recognized as an important element of enhanced situation awareness in military problem-solving contexts (Smart & Shadbolt, 2007; Smart et al, 2007) and, as part of our approach, we are developing techniques for semantic integration [2] and aggregation over multiple heterogeneous information sources. We have also developed a novel semantic query mechanism that can be used to increase the relevance of query results and reduce information overload.

The context-aware information retrieval framework is based on five intertwined fundamental concepts, namely:

- Workflow,
- Context,
- Profile,
- Ontology and
- Information Integration & Aggregation.

Our approach provides a contextual, user centric, and mission-oriented knowledge portal where each individual can access the right information in the context of their particular workflow. The ultimate aim of our approach is to provide a system that allows arbitrary navigation and semantic connections on any sources and products. Such a system must also take into account user individual interests and group constraints within a dynamic and evolving task context, and must allow collaborative work. In addition, it must include a variety of services, such as contextual search, task-oriented toolset, briefing production, configuration and security.

**The Workflow Service Layer:** A process model provides a comprehensive picture about the position of both a user and a piece of information within a larger task-oriented context. It may specify exactly how and where aggregated information is going to be used. For example, a process model may

indicate that the aggregated information is going to be shared among other users, in which case it links different user profiles together for better personalization purposes.

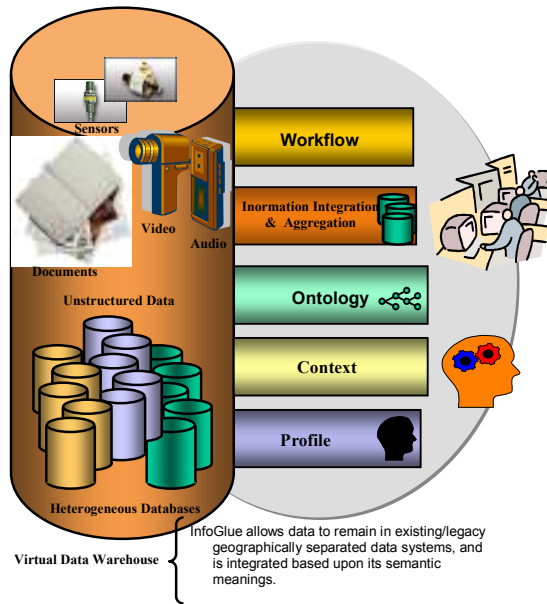


Figure 1. Major components of the context aware information retrieval framework.

**The Context Service Layer:** The notion of context awareness in information retrieval has received a great deal of attention recently. Research at Boeing's Mathematics and Computing Technology group has begun to show tangible progress in this area. We have developed approaches to representing aspects of the users' individual preferences, workflow, and information environment that can be leveraged in order to provide real assistance in the access, retrieval, integration and aggregation of information from heterogeneous sources (Yuan, Bahrami 2006). As an extension to this work, we propose to further augment the implementation of semantic queries over an ontology-mediated information integration service.

**The Profile Service Layer:** A user profile is an ontology fragment that describes the user's characteristics in regards to information consumption. Some of the characteristics are relatively static, such as a user's interests, skills, preferences, and security classification; but many others are highly dynamic (e.g. available devices, network bandwidth, current locations, etc.), and change significantly depending on the situation. Our idea here is that the user profile is used to adjust the information aggregation request such that aggregated information

content reflects a user's specific capabilities, authorities and responsibilities.

**The Domain Ontology Service Layer:** An ontology provides a semantically-enriched representation of some aspect of a domain of interest. Ontologies provide a mechanism for denoting the meaning of information content and they provide a substrate for semantically-mediated capabilities such as semantic queries and semantic integration. Domain ontologies are often used in conjunction with upper ontologies which define specific abstract concepts and can often be used to provide a useful common set of semantics for multiple lower-level or domain-specific ontologies.

Ontologies are an important element of context-aware information retrieval because they support information integration from semantically heterogeneous and physically-disparate information repositories (Smart & Shadbolt, 2007). Our approach to semantic integration is grounded in the user of semantic integration techniques, particularly ontology alignment solutions. Ontology alignment is a key element of semantic integration [4], and its aim is to establish mappings between the elements of multiple (ostensibly disparate) domain ontologies as a means of identifying semantically-equivalent (or similar) sub-components. Once established, these mappings can be used to drive information aggregation and integration activities (as might be required for goal-relevant information processing [5], for example), but most of all we see such mappings as providing a mechanism for the ontology-based mediation of inter-agent information transfer between conceptually or linguistically disparate communities.

In fact, ontology alignment constitutes only one aspect of semantic integration. Various forms of semantic integration have been proposed in addition to ontology mapping<sup>1</sup> (see Figure 2). These include ontology merging (the creation of a single coherent ontology from two or more sub-ontologies or ontology fragments), articulation (the creation of logical rules that express correspondences between some of the terms in the source ontologies) and transformation (the creation of transformation functions to transform one ontology into another).

<sup>1</sup> Ontology mapping is also referred to as ontology alignment.

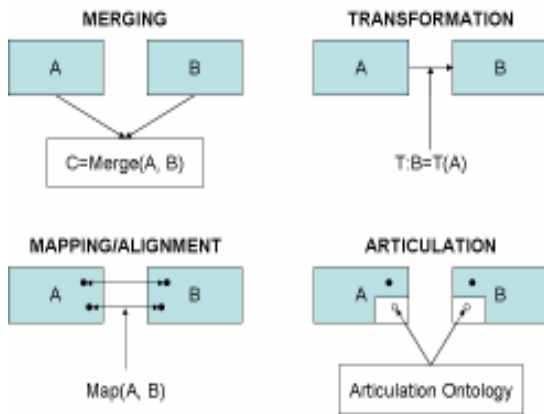


Figure 2: Semantic Integration Techniques.

### Information Integration & Aggregation (IIA) Service Layer

The proposed framework requires input from multiple information sources, in which there may be a combination of structured, semi-structured, and unstructured data. These information resources are inherently heterogeneous and extremely large. Existing technologies such as data warehousing have provided solutions to some extent. However, because of the issues such as: the sheer scale and volume of data, the extremely large number of participating systems, the variety of data sources and formats, and more importantly lack of necessary semantic understandability, existing approaches are costly and inefficient particularly when it comes to maintaining data consistency.

The nature of any complex system infrastructure requires those who operate and support those systems to have the ability to interoperate. Because these systems usually need to have ready access to multiple, disparate information resources, such interoperability can be achieved principally through information integration, as contrasted with system integration. The disparate nature of these information sources poses unique challenges and barriers having to do with access, accuracy, semantic understanding, completeness, and correlation of heterogeneous information and therefore successful information integration efforts depend critically on elimination of those barriers. The Boeing-developed approach to information Integration & Aggregation (IIA) can be configured to create a single, unified, coherent, and semantically-informed view of selected intelligence information. The semantics of information content are

explicitly represented by domain ontologies, and it is against these ontologies that queries can be issued to retrieve an integrated set of data in a coherent form. A system metadata repository is created containing logical design information for the targeted systems and mappings between the domain ontology and the logical models. This approach is known as a mediated approach to information integration. The domain ontology not only serves as the conceptual representation of the information domain, but it is also used to resolve the heterogeneity of the concepts in related data systems. Other metadata serves to encapsulate each member data system by providing a logical representation of the system's content – these logical elements are mapped to concepts in the ontological model and together “mediate” the system information requests.

The IIA run time system is implemented as a Web Service so as to fit into an open Service-Oriented Architecture (SOA). It uses the standard SOAP communication protocol to implement a distributed interface, as well as WSDL to expose its capabilities to outside callers for remote invocation. This allows the information service to be used and re-used without regard to implementation details by client applications. Client queries and query returns are implemented using XML standards, for instance the W3C XQuery specification. The IIA run time system can also incorporate a “publish and subscribe” proxy which further decouples the IIA from client applications.

IIA has been implemented using a variety of Commercial Off-The-Shelf (COTS) technologies as well as technology components developed within Boeing. The IIA architecture employs commercial technologies where possible to leverage industry development efforts while mitigating the risk of commercial supplier volatility. IIA specifically uses two COTS products in its runtime architecture. The ontology management system is based on the Interactive Ontology Development Environment from Ontology Works. It provides for ontology representation and management as well as deductive reasoning capabilities. The IIA distributed query execution engine is based on IBM WebSphere Information Integrator, and it provides a system-internal data-level operational interface to IIA member data sources. This is invoked to perform query decomposition, dispatching, and efficient

inter-database joins.

Heterogeneity can take many forms – when discussing mediated approaches. IIA successfully provides sort of “semantic independency” in classifying different types of heterogeneity into two major categories, i.e. semantic heterogeneity and non-semantic heterogeneity. We generally think of semantic heterogeneity as having to do with the meaning of the terms and concepts, and non-semantic heterogeneity as related to the platform, network, data model, and query mechanisms. With the separation of semantic and non-semantic heterogeneity, IIA is able to use a COTS-federated query engine to deal with non-semantic heterogeneity by providing a single unified query interface for different data systems. IIA components can then utilize the ontology and federated query engine components to provide the semantic mediation capability. Software adapters encapsulate the COTS components to shield internal subsystems from changes in the COTS implementations, providing loose coupling and allowing for greater flexibility.

#### **IMPLEMENTATION AND FUTURE WORK**

Currently, we are implementing a web-based prototype system that contains a few major components of this infrastructure, including workflow management, profile management, ontology management and query management. The prototype provides a unified web portal for both information producer and consumer to interact with the system. We’ve also extended our previous work reported in [2] by utilizing an innovative query rewriting technique to improve the dynamic adaptation of query processing. We also plan to do more experiments in the near future against available data sources to further test the performance and the scalability of the prototype system.

#### **SUMMARY**

The main objective of our approach is to develop a mechanism to enable context-influenced information retrieval from heterogeneous data sources. The approach lays on five intertwined fundamental concepts, namely Workflow, Context, Ontology, Profile and Information Aggregation. The exploitation of these five concepts will make it possible to provide contextual assistance in various ways to the work performed according to user’s preferences and workflow. We have developed

an initial algorithm to transform user workflow, profile, ontology and other contextual information into semantic queries to increase the relevance of query results and reduce analysts’ information overload and overall workload [2].

#### **ACKNOWLEDGEMENT**

Research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defense and was accomplished under Agreement Number W911NF-06-3-0001. 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 U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defense or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

#### **REFERENCES**

Ali Bahrami “Achieving Agile Enterprise Through Integrated Process management: From Planning to Work Execution,” International Journal of Cases on Electronic Commerce, October-December 2005, Vol.1 No. 4, pp 19-34.

Ali Bahrami, Changzhou Wang, Jun Yuan and Anne Hunt, “The Workflow Based Architecture for Mobile Information Access in Occasionally Connected Computing” 2006 IEEE International Conference on Services Computing (SCC 2006), September 18-22, 2006 Chicago, IL.

D. Martin et al (eds) OWL-S 1.1 Release, Upper ontology for Services, <http://www.daml.org/services/owl-s/1.1/>, Nov. 2004.

Ehrler, L., Fleurke, M., Purvis, M. A. and Savarimuthu, B.T.R., “Agent-Based Workflow Management Systems(Wfms) : JBees- A Distributed and Adaptive WFMS with Monitoring and Controlling Capabilities”, to be published in a special issue of the Journal of Information Systems and e-Business on Agent-Based Information (2005).

Fleurke, M and Ehrler, L, Purvis, M. A. (2003). “JBees - An Adaptive and Distributed

- Agent-based Workflow System”, in Proceedings of the International Workshop on Collaboration Agents: Autonomous Agents for Collaborative Environments (COLA 2003), Halifax, Canada, October 2003. IEEE/WIC Press. Ghorbani, A. And Marsh, S., Ed.
- Jun Yuan, Ali Bahrami, Changzhou Wang, Marie Murray and Anne Hunt, “The A Semantic Information Integration Tool-Suite”, Proceedings of the 32nd International Conference on Very Large Databases (VLDB), Seoul, Korea, September 12-15, 2006.
- José M. Vidal, Paul Buhler, and Christian Stahl. Multiagent systems with workflows. *IEEE Internet Computing*, 8(1): 76-82, January/February 2004.
- M. R. Endsley, (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society 32nd Annual Meeting* Santa Monica, CA: Human Factors Society, (1988) 97-101.
- M. Endsley, B. Bolte, and D. G. Jones, *Designing for Situation Awareness: An Approach to User-Centered Design*. London: Taylor & Francis, 2003.
- Marlene Gauvin, Anne-Clarie Boury-Brisset and Alain Auger “Context, Ontology and Portfolio: Key Concepts for a Situational Awareness Knowledge Portal” 2004 Proceedings of 37<sup>th</sup> Hawaii Conference on System Sciences.
- Maryam Purvis, Bastin Tony Roy Savarimuthu, Martin Purvis, A Multi-agent Based Workflow System Embedded with Web Services, Department of Information Science, University of Otago, PO Box 56, Dunedin, New Zealand.
- P. R. Smart, A. Bahrami, D. Braines, D. McRae-Spencer, J. Yuan, and N. R. Shadbolt, "Semantic Technologies and Enhanced Situation Awareness," presented at 1st Annual Conference of the International Technology Alliance (ACITA), Maryland, USA, 2007.
- P. R. Smart and N. R. Shadbolt, "The Semantic Battlespace Infosphere: A Knowledge Infrastructure for Improved Coalition Inter-operability," presented at 4th International Conference on Knowledge Systems for Coalition Operations (KSCO), Massachusetts, USA, 2007.
- Paul Buhler, José M. Vidal, and Harko Verhagen. Adaptive workflow = Web Services + agents. In Proceedings of the International Conference on Web Services, pages 131-137. CSREA Press, 2003.9
- Savarimuthu, B.T.R., Purvis, M. A. and Fleurke, M. (2004), "Monitoring and Controlling of a Multi-agent Based Workflow System", Proceedings of the Australasian Workshop on Data Mining and Web Intelligence (DMWI2004), Conferences in Research and Practice in Information Technology, Vol. 32, Australian Computer Society, Bedford Park, Australia (2004) 127-132.
- Sowa FJ: Top-level ontological categories. *International Journal of Human Computer Studies* 1995, 43:669-685.
- Stevens R, Robinson AJ, Goble C: myGrid: personalised bioinformatics on the information grid. *Bioinformatics* 2003, 19:302i-304i.
- Sycara, K.P., *Multiagent Systems*. AI magazine, 1998.
- Wang, M. and Wang, H. Intelligent Agent Supported Flexible Workflow Monitoring System. in *Advanced Information Systems Engineering: 14th International Conference, CAiSE 2002*. 2002. Toronto, Canada: Springer Verlag GmbH.
- Wooldridge, M.J., *Intelligent Agents*, in *Multiagent Systems*, G. Weiss, Editor. 1999, MIT Press: Cambridge. p. 27-77.
- Y. Kalfoglou and M. Schorlemmer, "Ontology mapping: the state of the art," *Knowledge Engineering Review*, vol. 18, pp. 1-31, 2003.