

# Domain-specific Knowledge Management in a Semantic Desktop

**Ola Leifler**

(Linköping University  
olale@ida.liu.se)

**Henrik Eriksson**

(Linköping University  
her@ida.liu.se)

**Abstract:** Semantic Desktops hold promise to provide intelligent information-management environments that can respond to users' needs. A critical requirement for creating such environments is that the underlying ontology reflects the context of work properly. For specialized work domains where people deal with rich information sources in a context-specific manner, there may be a significant amount of domain-specific information available in text documents, emails and other domain-dependent data sources. We propose that this can be exploited to great effect in a Semantic Desktop to provide much more effective knowledge management. In this paper, we present extensions to an existing semantic desktop through content- and structure-based information extraction, domain-specific ontological extensions as well as visualization of semantic entities. Our extensions are justified by needs in strategic decision making, where domain-specific, well-structured knowledge is available in documents and communications but scattered across the desktop. The consistent and efficient use of these resources by a group of co-workers is critical to success. With a domain-aware semantic desktop, we argue that decision makers will have a much better chance of successful sense making in strategic decision making.

**Key Words:** knowledge management, semantic desktops, domain-specific ontology

**Category:** H.5.3, H.5.4

## 1 Introduction

Inconsistent or improper use of information is a major cause of concern in environments where the correct, timely, joint assessment of a situation is paramount to success. As part of knowledge management at the strategic level of a larger organisation, documents are communicated within as well as between different levels of management, using both general desktop applications as well as specialized systems. Being able to correctly assess the information contained in those documents as well as in supporting information systems is critical to success. For efficient information management, Semantic Desktops [Karger et al. (2005), Sauer mann et al. (2006)] have introduced Semantic Web technologies to a desktop environment in an effort to enhance their reasoning capabilities. Such capabilities are believed to empower users through a unified knowledge management platform that integrates several application-specific information sources.

To be effective in supporting higher level strategic decision-making, however, semantic desktop ontologies need to accommodate the specific context of work to provide

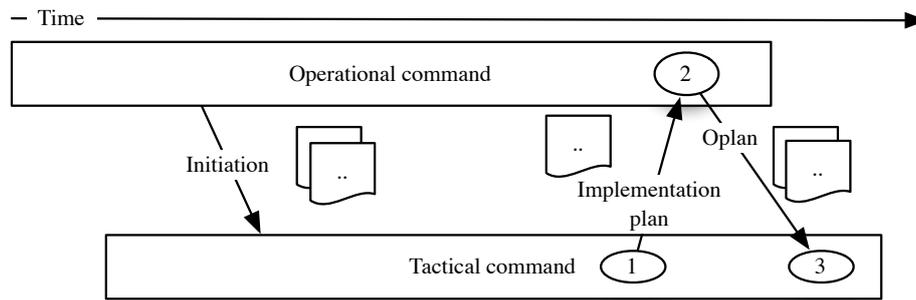


Figure 1: The information flow between the operational and tactical levels of staff. Both staffs work simultaneously with developing their understanding of the current situations and their options. As part of this process they exchange and modify a set of plan documents. In particular, the tactical command develop their suggestion for how to implement the operational objectives prior to the final operation plan.

useful reasoning services [Leifler and Eriksson (2008)]. Without such domain models, reasoning would be limited to information-based inferences, not knowledge-based ones. We present in this paper (1) how ontology-based *information management* in current semantic desktops can be extended with domain-specific features, and (2) an implementation of such extensions in a semantic desktop environment that can provide significant benefits to *knowledge management* for command and control.

## 2 Command and control

In strategic decision-making environments, which we exemplify with military command and control, correct sense-making is critical to success [Brehmer (2007)], which in turn places high demands on knowledge management. As an example of the knowledge management process in command and control, consider Figure 1, where two levels of command communicate in part by submitting documents to one another prior to launching an operation. The tactical command reports to the operational command how they would prefer to solve the task given to them [Fig. 1, step 1]. These preferences are taken as input by the higher command when they outline the final plan of the operation [Fig. 1, step 2]. Finally, the tactical command receives a plan that they are responsible to execute [Fig. 1, step 3]. In the process, it is rarely the case that specialized, intelligent subsystems are used to automate work being done by humans.

Although research on intelligent systems for automating strategic decision making has yielded important results [Smith et al. (2005), Hayes et al. (2005), Becker and Smith (2000)], and recent research has attempted to bridge the gap between automated systems and conventional tools [Leifler (2008)] to improve the acceptance

of using intelligent support systems, mission plans are still crafted as a set of structured text documents. These documents describe separate aspects of a plan such as the task organization, rules of engagement, available time frames, force composition, and relevant intelligence information. From Figure 1 we can deduce that, prior to the delivery of the final *operational plan* from the Operational command, the subordinate staff at the tactical level will have to provide an *implementation plan* as part of their preparations. Also, according to NATO Guidelines for Operational Planning (GOP), there shall be a *synchronization matrix* during planning, in which responsibilities of military units at the various stages of the operation are listed [Fig. 3]. Documents and communications in command and control thus generally follow a well-specified pattern in terms of content and structure, and can be used by a semantic desktop-based environment for reasoning about the current work context.

### 3 The IRIS semantic desktop

The IRIS semantic desktop [Cheyer et al. (2006)] is a modular, service-based semantic desktop project which features an OWL-based object store, integration via plug-ins with external applications such as web browsers, integrated applications such as a calendar as well as a set of support applications for navigating and manipulating the underlying ontology. The applications included within IRIS are general in the sense that they are not specific to a particular work context. This is a feature of IRIS since it is intended to be a general framework for information management. For the purpose of creating domain-specific support in information-intensive work domains, however, we have identified a need for more powerful mechanisms for handling semantic entities that can be found in documents and communications.

Using the semantic desktop, we can aid the decision makers in this case by dealing with issues such as *What locations are referred as **Secure Point of Debarkation** in the current mission?* and *What are the different stages of the operation and what are the responsibilities of each unit during the first stage?*. To provide commanders with technical assistance with these issues, we propose the following use of the semantic desktop:

- When the operational command receives information from the tactical command that they have prepared an initial implementation plan [Fig. 1, step 1], the operational command opens the implementation plan in the form of a text document and reviews it [Fig. 1, step 2]. In doing so, the semantic desktop triggers domain-specific information extraction, which updates the ontology with locations referred to in the document. One of these locations has no previous connection to this document, causing a property update and a notification to the user via messages in the domain-specific information view related to geographical information.
- As part of writing the final operational plan, the operational command formulates a synchronization matrix to answer the questions *who* is responsible for *what when*.

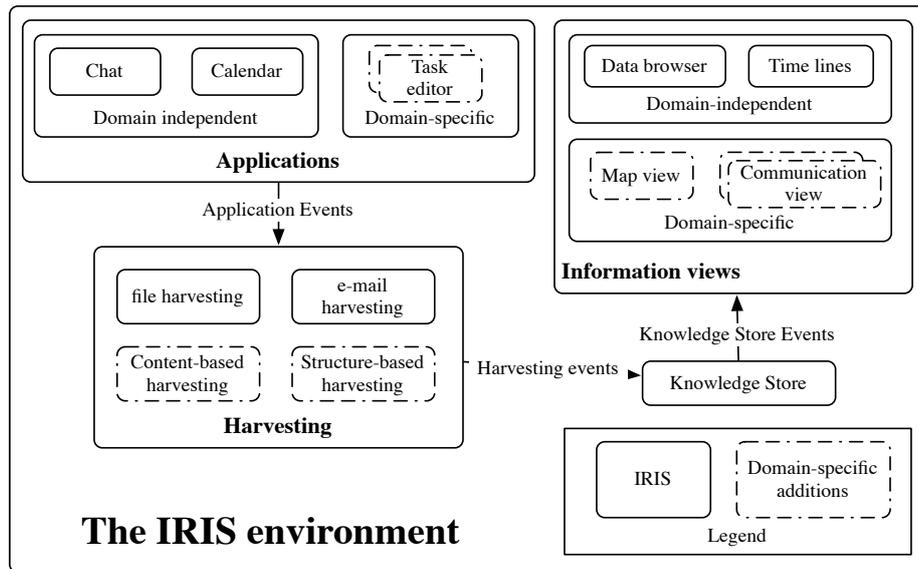


Figure 2: An overview of the IRIS semantic desktop enhanced with domain-specific additions.

The file containing the synchronization matrix is classified as a synchronization file by the semantic desktop, causing structure-based harvesting of information on tasks, responsibilities and timing. Both levels of staff shares the same semantic desktop environment with the same underlying ontology, so changes of the task descriptions at the operational level are also reflected at the tactical level. Whenever documents containing synchronization information are manipulated at either level, an explicit task representation is updated to reflect these changes [Fig. 3].

To effectively provide commanders with support in the situations described above requires features that are beyond what is provided by current semantic desktops: (1) the ability to *classify domain-specific words and phrases* by their semantic class, and (2) the ability to *recognize domain-specific structures* of work-related documents. As a response to these specific challenges, we have extended the IRIS semantic desktop with *named entity recognition* and *document structure analysis* [Fig. 2]. We have also extended the IRIS ontology to accommodate our specific domain.

### 3.1 Domain-specific extensions to IRIS

Current IRIS functionality for harvesting information from files and populating the ontology database relies on the use of the SEMEX [Cai et al. (2005)] information management framework. SEMEX is domain-agnostic just as IRIS in general in that no

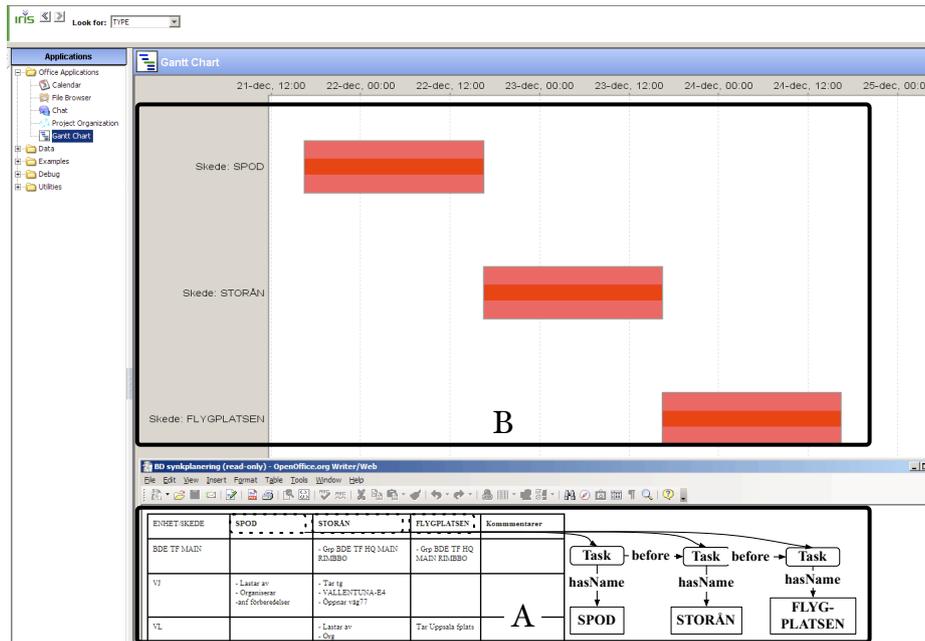


Figure 3: The IRIS environment contains a graphical component to show tasks in a time line. By automatically harvesting synchronization matrices (lower section, A), we can populate the time line with tasks that correspond to those being planned for (upper section, B).

information that is specific to a particular work context is encoded using the current harvesting techniques. To extract semantics pertaining to the work context, we have deployed natural language processing components from the GATE framework [Cunningham et al. (2002)] in IRIS to recognize semantic objects by their linguistic features. Some semantic objects are written in a particular format, such as force descriptions, directions and locations in military documents. These references can be found by searching for noun phrases that have a domain-specific signature.

Also, we extract information from the domain-specific document structure of files containing synchronization matrices. This addition to the IRIS harvesting environment was created using XPath-based document structure filtering of OpenDocument files. Figure 3 (lower section, A) describes how specific parts of a synchronization matrix are used to create semantic objects denoting stages of an operation. The information from named entity analysis and document structure analysis is harvested from documents automatically in IRIS, and is used for navigating domain-specific ontology information more efficiently using special information views. The current IRIS ontology is written in

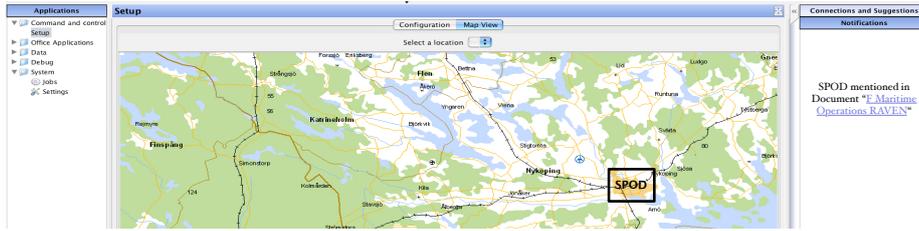


Figure 4: Map view showing a geographic reference plotted on a map and linked to a document.

OWL-DL and contains concepts related both to the activities performed by applications in a desktop environment, such as opening files, launching applications and reading e-mail, as well as to what users reason about in the environment, such as people, tasks, and places.

Once domain-specific harvesting has been set up by coupling named-entity recognition to ontology classes in IRIS, and by recognizing structure-dependent semantic entities and their relations in office documents, this information is processed by IRIS in an event-driven manner. In Figure 3, a time line describes the stages of an operation, as harvested from a synchronization matrix. Another example of how domain-specific, event-driven information can be used to assess the current situation is shown in Figure 4 where the Secure Point of Debarkation is marked on a map and provides direct access to the documents in which it is currently being referred to. When documents are manipulated, the set of geographical references is updated by using GATE for harvesting location references. Since such location references can include code names or functional locations such as SPOD in the military context, we cannot rely on GIS services to automatically plot them. However, once their correspondence to locations on a map has been established, we can provide visual cues and links to documents that refer to the locations using the IRIS notification framework.

#### 4 Related Work

Reeve and Han surveyed extraction of semantic information from documents using automated methods. These methods do not require any modification to the editing environment for producing documents, and attempt to minimize the user effort when extracting semantic information [Reeve and Han (2005)]. However, the presented methods do not treat any domain-specific information about documents, and are not concerned with finding document information that depends on guidelines for document creation such as the GOP. In other projects aimed at defining the semantics of documents, researchers have integrated ontology editing capabilities within document editing environments to provide support for creating semantic annotations when writing a document

[Groza et al. (2007), Eriksson (2007)]. Neither of these projects rely on, or require, any knowledge of the domain at hand. Because of this, domain-specific ontology elements are not possible to capture. Franz, Staab and Arndt [Franz et al. (2007)] have demonstrated how to integrate many application-specific information sources in a semantic desktop framework. Although they refer to *domain*-specific information sources, their framework primarily deals with *application*-specific ones.

## **5 Discussion**

In this paper, we have outlined how to extend a semantic desktop environment with a domain-specific ontology to support information visualization and navigation. Semantic Desktops are open platforms since the goal of a semantic desktop is to merge information from different sources and provide an intelligent framework for reasoning with pre-existing entities. Until now, however, few domain-specific applications have been demonstrated.

Our approach to extending the semantic desktop with a richer context model provides a specific advantage for strategic decision-making. In general, any domain with semantic entities in well-defined locations or as well-defined elements of vernacular language can benefit from similar extensions, thereby enabling an ontology which much more closely captures the context of work. By providing a knowledge base that reflects relevant semantic entities in the context of work, a reasoning agent can help people in knowledge management, and not only with information access.

## **6 Conclusions**

This article describes a model for enhancing semantic desktops with domain-specific ontologies for supporting strategic decision making in heterogeneous desktop environments. The model includes the use of natural language processing and document structure analysis, and has been implemented on an existing semantic desktop (IRIS). Our results show that using domain-specific extensions, a semantic desktop can provide directed support for richer environments and at a higher level than otherwise possible. By domain-specific extensions to the ontology and harvesting mechanism, rich and custom-tailored support mechanisms can be produced with only minor additions to the semantic desktop.

## **7 Acknowledgements**

This work was sponsored by the Swedish National Defence College.

## References

- [Becker and Smith (2000)] Becker, M. A., Smith, S. F.: “Mixed-initiative resource management: The AMC Barrel Allocator”; Proceedings of the 5th International Conference on Artificial Intelligence Planning and Scheduling (AIPS-2000); Breckenridge, Colorado, USA, 2000.
- [Brehmer (2007)] Brehmer, B.: “Understanding the functions of  $C^2$  is key to progress”; The International  $C^2$  Journal; 1, 1 (2007), 211–232.
- [Cai et al. (2005)] Cai, Y., Dong, X. L., Halevy, A., Liu, J. M., Madhavan, J.: “Personal information management with SEMEX”; SIGMOD ’05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data; ACM Press, New York, NY, USA, 2005; 921–923.
- [Cheyer et al. (2006)] Cheyer, A., Park, J., Giuli, R.: “IRIS: Integrate. relate. infer. share.”; Proceedings of the Semantic Desktop and Social Semantic Collaboration Workshop (SemDesk-2006); 2006.
- [Cunningham et al. (2002)] Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: “GATE: A framework and graphical development environment for robust NLP tools and applications”; Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics; 2002.
- [Eriksson (2007)] Eriksson, H.: An Annotation Tool for Semantic Documents; 4519; 759–768; Springer Berlin/Heidelberg, Innsbruck, Austria, 2007.
- [Franz et al. (2007)] Franz, T., Staab, S., Arndt, R.: “The X-COSIM integration framework for a seamless semantic desktop”; K-CAP ’07: Proceedings of the 4th International Conference on Knowledge Capture; ACM, New York, NY, USA, 2007; 143–150.
- [Groza et al. (2007)] Groza, T., Handschuh, S., Möller, K., Decker, S.: “SALT - semantically annotated latex for scientific publications”; Proceedings of the 4th European Semantic Web Conference (ESWC 2007); Innsbruck, Austria, 2007.
- [Hayes et al. (2005)] Hayes, C. C., Larson, A. D., Ravinder, U.: “Weasel: A mixed-initiative system to assist in military planning”; Proceedings of the IJCAI-2005 Workshop in Mixed-Initiative Planning and Scheduling; Edinburgh, Scotland, 2005.
- [Karger et al. (2005)] Karger, D. R., Bakshi, K., Huynh, D., Quan, D., Sinha, V.: “Haystack: A customizable general-purpose information management tool for end users of semistructured data”; Proceedings of the Second Biennial Conference on Innovative Data Systems Research; 2005.
- [Leifler (2008)] Leifler, O.: “Combining Technical and Human-Centered Strategies for Decision Support in Command and Control — The ComPlan Approach”; Proceedings of the 5th International Conference on Information Systems for Crisis Response and Management; 2008.
- [Leifler and Eriksson (2008)] Leifler, O., Eriksson, H.: “A model for document processing in semantic desktop systems”; Proceedings of I-KNOW ’08, The International Conference on Knowledge Management; 2008.
- [Reeve and Han (2005)] Reeve, L., Han, H.: “Survey of semantic annotation platforms”; Proceedings of the 2005 ACM symposium on Applied computing; (2005), 1634–1638.
- [Sauermann et al. (2006)] Sauermann, L., Grimnes, G. A., Kiesel, M., Fluit, C., Maus, H., Heim, D., Nadeem, D., Horak, B., Dengel, A.: “Semantic Desktop 2.0: The Gnowsis Experience”; I. Cruz, ed., Proceedings of the International Semantic Web Conference (ISWC 2006); 4273 of Lecture Notes in Computer Science; Springer Verlag, 2006; 887–900.
- [Smith et al. (2005)] Smith, S. F., Hildum, D. W., Crimm, D. R.: “Comirem: An intelligent form for resource management”; IEEE Intelligent Systems; 20, 2 (2005), 16–24.