



[JOINT WHITE PAPER]

# Ontos Semantic Factory

JANUARY 2009

## Executive Summary

In this paper we describe Ontos Semantic Factory—a platform producing semantic metadata on the basis of text (Web) content. The technology relies on the ontology-driven NLP engine OntosMiner™, which extracts information from natural-language texts and creates coherent semantic metadata for the processed content. Latest versions of OntosMiner™ process texts in English, German and Russian.

We are ready to provide semantic metadata and other Ontos Semantic Factory's services for Web Search, home and enterprise applications, and we invite software developers to co-operate in this area. In our opinion, the proposed integration allows to make a market breakthrough for diverse semantic applications, as well as to enhance capability of classic business applications, and to bring users new 'semantic' experience.

In what follows you will find our detailed vision of how software developers could co-operate with Ontos and AviComp Services to uncover the gist of the Semantic Web and to create business value from Ontos Semantic Factory.

## Desiderata for Technology and Market

Today we see the market of Web technologies as a quickly revolving system of approaches and applications. In Ontos and AviComp Services we try to make the Web as useful as possible both for Web consumers and for enterprise users.

The latest promising trend in Web development—the Semantic Web—is still rarely incarnated in large-scale solutions on the market. In our vision there are several obstacles to the advent of the Semantic Web for both end users and software developers. In this paper we would like to address the following desiderata for the market and for the technology, which in our opinion allow to overcome the aforementioned obstacles.

### **Desideratum One: Leverage Semantic Metadata within Existing Applications**

It is commonly agreed that semantic metadata greatly enhance business-oriented applications. These metadata come mostly from various structured sources (databases, thesauri, etc.). Still there is a lot room for further enhancement. According to a common expert estimate, about 80% of all reasonably accessible information underlying business processes is found only in natural-language texts. The major problem is that this text content cannot be readily understood by machines. It would be nice to have an ontology-based layer of metadata extracted from the text content and creating tight connections between pieces of knowledge coming from diverse applications.

### **Desideratum Two: A Semantic Factory**

To satisfy the above desideratum, one would need a factory producing full-fledged semantic metadata on the basis of regular content coming from corporate documents and from the Web. By ‘full-fledged’ we mean that the metadata have to be rich and customizable enough to allow users to facilitate their business processes, their research and entertainment activities. Not to invent a wheel, these semantic metadata should meet the standards recommended by the W3C (RDF, SPARQL and OWL).

The Semantic Factory should be built as a complex scalable architecture of semantic processors and adjacent modules, which ultimately form a new semantic platform for software and web developers. A great number of end-user-oriented products and services can be created on this platform.

We believe that there are two crucial prerequisites for the Semantic Factory supporting Internet and Intranet applications: an extremely elaborated system of data stores and a fine-grained NLP-based semantic analysis of the Web content.

\* \* \*

The remainder of this paper is organized as follows.

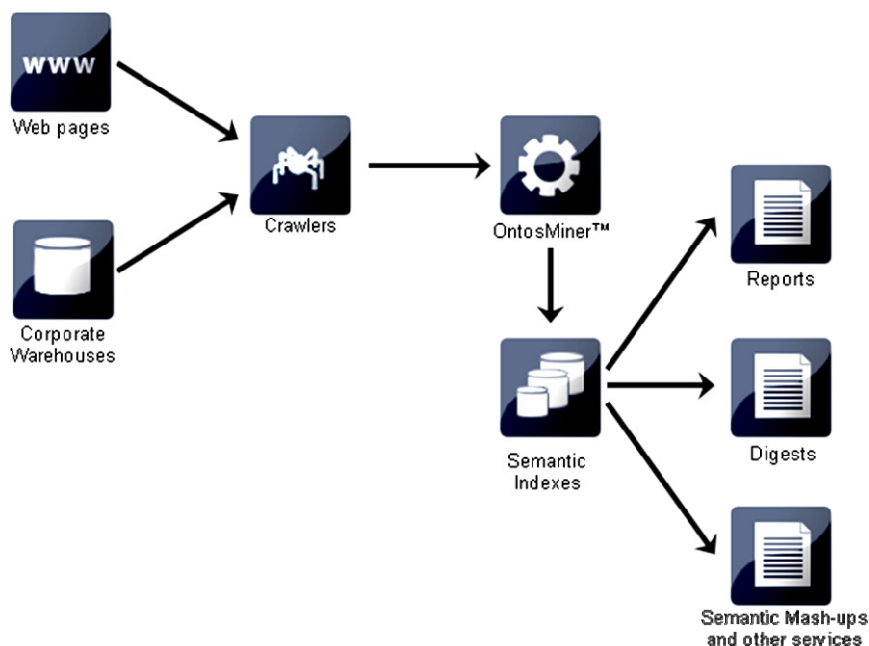
In Section 1, we describe a Semantic Factory created by Ontos, going in depth with our key achievements in Natural Language Processing and semantic metadata production.

In Section 2, we give a general overview of the areas in which we are ready to co-operate with business software developers. These fields include Enterprise Search, BI solutions and applications for Competitor Analysis, CRM systems, Corporate E-Mail, and Document/Content Management systems.

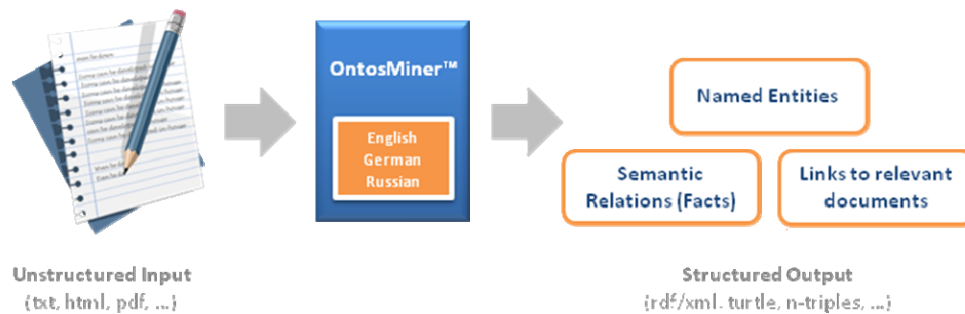
## Section 1. Ontos Semantic Factory

Since 2001, Ontos in co-operation with our partner AviComp Services has been developing a system of applications allowing to effectively produce semantic metadata on the basis of Web text content. This system crawls over the Web, extracts entities and relations on the basis of domain ontologies, creates NLP-based semantic metadata connected to source documents, and builds a range of analytical reports on the basis of these metadata.

The following picture illustrates the general architecture of Ontos Semantic Factory.



The system is based on the patented NLP engine OntosMiner™. The following diagram illustrates the process of text analysis performed by OntosMiner.



### Key characteristics of Ontos Semantic Factory

- ❖ extraction of entities and relations (facts about entities) from unstructured documents on the basis of domain ontologies
- ❖ complex identification of co-referent objects based on domain-specific roles within a document and across documents (incl. anaphora resolution)
- ❖ ability to operate on English, German and Russian texts (more than 10,000 linguistic rules per language/domain)
- ❖ high quality & performance metrics

On the basis of this Factory we developed a number of unique services, including the following:

- ❖ Search on the basis of semantic relevance
- ❖ Highlighting extracted objects on processed Web pages with a possibility to navigate through their relations to other objects and documents (semantic hypertext)
- ❖ Semantic Summaries & Digests (reports containing links to documents and document fragments semantically salient with respect to a given object)
- ❖ Clustering based on the semantic similarity of documents

These services are incorporated in our web portals:

- ❖ On-line news aggregator Ontos™ News
- ❖ Business on-line semantic library Ontos™ Biz
- ❖ Semantic Social bookmarking service Ontos™ Bookmarks (in beta)

This list just gives an idea about how much stuff can be invented on the basis of the Factory's semantic output. Our experience shows that Ontos Semantic Factory is mature enough to be incorporated in third parties' applications. This is the topic of the following part.

## Section 2. Consuming the Semantic Factory's Output

The semantic metadata (SM) produced by Ontos Semantic Factory can be used across a number of enterprise applications. We are happy to propose our partners to co-operate in the following areas:

- ❖ Web Search (including the Enterprise-oriented one)
- ❖ Business Intelligence (BI)
- ❖ Customer Relationship Management (CRM)
- ❖ Competitor Analysis (CA)
- ❖ Enterprise E-Mail systems
- ❖ Document/Content Management

We are ready to provide the functionality of Ontos Semantic Factory and its semantic output for software developers operating in the above areas. Crucially, there are two dimensions in which software developers can customize their use of Ontos Semantic Factory: the use of the Factory's existing technological core (the engine itself) and the use of the Factory's output (semantic metadata).

In the following table we draw four approaches emerging from this dichotomy.

|                     | Adapt SM  | Take SM as is  |
|---------------------|---|--|
| Adapt the Core      | Use Ontos Semantic Factory API  | Put the Metadata into your own database                          |
| Take the Core as is | Embed the functionality of creating user-defined SM in Intranet and corporate resources | Embed Ontos Semantic Metadata Subscription and Semantic Services |

## Conclusion

In this paper we addressed a promising prospect of integration of semantic metadata into Web-oriented software applications. The gist of our message is that the real advent of the Semantic Web in Tim Berners-Lee's sense is obviously conditioned by the measure of integration of semantic metadata into Web Search, and into home and enterprise applications.

Ontos semantic technology considerably enhances the coherence of business processes and entertainment activities. This ultimately brings a mutual benefit both to our partners and to us, and opens a whole new world of semantics-based business and end-user applications.

### AviComp Services CJSC

Vernadsky av. 84 build. 2  
119606 Moscow Russia  
Phone: +7 495 436-09-45  
Fax: +7 495 436-07-49  
[services@avicomp.com](mailto:services@avicomp.com)  
[www.avicomp.ru](http://www.avicomp.ru)

