

SyNTHEMA, LANGUAGE INTELLIGENCE & SEMANTIC TECHNOLOGY

BASED IN PISA (ITALY), SyNTHEMA IS A HIGH-TECHNOLOGY COMPANY THAT WAS ESTABLISHED IN 1993 BY COMPUTER SCIENTISTS FROM THE IBM RESEARCH CENTER.

Since then, the company has rapidly evolved, becoming nowadays leading provider of Language and Semantic solutions, with state-of-the-art technologies for applications like Enterprise Search, Audio&Text Mining, Technology Watch, Competitive Intelligence, Customer Relationship Intelligence e Management, Sentiment Analysis, Speech Recognition, Respeaking and Speech Analytics. Grounding its leadership into a strong and long lasting IT Research and Development, SyNTHEMA has pioneered a number of innovative applications and solutions, adopted on a daily basis by a vast amount of users to perform productivity tasks in different markets and industries, including Homeland Security, Intelligence and Law Enforcement, PA and Government, Healthcare and Media. SyNTHEMA is engaged in a systematic Research and Development activity both by participating to National and European projects and by reinvesting part of the profits back into internal research projects. Since 1999, SyNTHEMA has carried out several projects in the framework of the different IT program funded by the European Commission. Projects have also been funded by the Italian Ministry of Education, University and Research and by the Tuscany Regional Authority.



The main projects are:

SI-TAL

National infrastructure for linguistic resources in the sector of the automatic treatment of the natural spoken and written language. Synthema is involved in integrating Syntactic and Semantic TreeBanks, developed and provided by SI-TAL partners, into its MT System, and in assessing benefits of this integration for automatic translation process of technical documentation.

HERMES

An Information Delivery System based on Text Mining, Memory-Based Reasoning, and NetCasting technologies, and to exploit project results by creating an Experimental Service Center to provide SMEs with Full Text Information Selective Routing, Information Retrieval and Delivery.

PI-SPARK Enhanced Protection of IPR by Streamlined Provision of Access to Regulatory Knowledge

The aim of EPI-SPARK is to design, develop, manage, assess and commercially exploit a fully functioning prototype Internet e-commerce site providing IPR protection services. In this project, Synthema designs and develops a component for the information extraction and encoding of IPR (Intellectual Property Rights) Legal Knowledge, to provide experts with searching and updating facilities for legal information.

A multiplatform and multilingual interactive network for guiding people through leisure events

MOBIGUIDING aims at building a European, multi-platform, multi-lingual interactive leisure guide network, available on fixed Internet and mobile devices, featuring the most developed multimedia technology. This guide will be available on all Internet and mobile devices, the focus being on the 3rd generation mobile phones. This will be an interactive tool, helpful for both end-users, event organisers and content suppliers.

WISPER Worldwide Intelligent Semantic Patent Extraction & Retrieval

The next generation Internet requires a semantic expansion of digital content to fulfill the needs of intelligent retrieval systems. Content needs to be annotated with semantic information so that it can be processed by multiple agents and re-used in different contexts. This project will examine patent structured content, with the aim of demonstrating that flexible tools can be developed for automated mark-up. WISPER is a new multilingual portal for intelligent access to large patent databases. It encodes and visualizes semantic content of patents by using advanced Text Mining, Intelligent Image Analysis and User Profiling and thereby encourages SME participation in patenting. SMEs will be able to search both text and images for the first time by using natural language.

OPEN-IT Enhancing the SMEs internal innovation process by structuring available external knowledge

The OPEN-IT project aims to develop an advanced information system, based on Web Semantic technologies and innovation-oriented methodologies, able to support European SMEs in taking advantage from an "open innovation paradigm", by structuring internal innovation process focusing on available external knowledge. The proposed system is supposed to be used by internal resources of SMEs or by external innovation-management consultants. The OPEN-IT project final result will be an IT tool for searching, retrieving and managing external knowledge in weakly structured information (such as the Internet) as well as standardized databases (such as patent databases or scientific papers). The project proposal tool will support SMEs throughout the innovation process following three main modules: I) a methodological toolbox for product/process definition and analysis, which allows SMEs to make the best usage of their internal available knowledge; II) an information agent able to search and to retrieve relevant external information; III) an analytical IT tool to manage the external acquired information (Web, Patent and scientific papers) by advanced text mining, network analysis and NLP techniques.

KYOTO Making knowledge sharable between communities of people, culture, languages and computers, by assigning meaning to text and giving text to meaning

KYOTO addresses the need for global and uniform transition of knowledge across different target groups and organizations in society and across linguistic, cultural and geographic borders. The goal of KYOTO is to develop a system that provides semantic search and information access to large quantities of distributed multimedia data for both experts and the general public, and to apply this system to the environmental domain on a global scale. Information access is provided through a cross-lingual user-friendly interface that allows for high-precision search and information dialogues over a variety of data from wide-spread sources in a range of different languages: English, Dutch, Italian, Spanish, Basque, Chinese and Japanese. This is made possible through a customizable ontology that is linked to various wordnets and a

set of knowledge yielding text miners (so-called Kybots) for a variety of languages. Concept extraction and text mining are applied through a chain of linguistic and semantic processors that share a common knowledge base. This guarantees a uniform interpretation layer for the diverse information from different sources and languages. The system can be maintained and kept up to date by specialists in the field using an open Wiki platform.

GLOSS Integrated domain for sharing knowledge among different types of organizations involved in monitoring, analysis and management of environmental risks

The goal of GLOSS project is to create, in accordance with the current state of art in "Multi Lingual Information Access to Multi Media Contents" (MLIA2MMC) technology, an integrated domain for sharing knowledge among different types of organizations involved in monitoring, analysis and management of environmental risks. The "Environmental Protection Service and Civil Defense" of the Province of Pisa and the "Environmental and Territory Department" of the Province of Livorno participate in the project as domain-expert users. These are the main components of GLOSS: I) the "Semantic Crawling Engine" (SCE) retrieves information from the web in order to build augmentatively the documental base; II) the "Terminology Extraction Agents" (TEA) extract terms of interest and their possible semantic relations from the sources and connect them to the Wordnets of their respective languages; III) the "Georeference Extraction Agents" (GEA) extract terms and their geographic relations (district, area, region); IV) the "Fact Extraction Agents" (FEA) use terms and knowledge to identify facts (descriptions, events and references in the texts), which contain the terms self or concepts associated to them; V) the "Wiki Knowledge Editor" (WKE) enables specialists in the field to insert terms in their own language and to describe their meaning through concepts defined by the international community.

MOSAIC Multi-Modal Situation Assessment & Analytics Platform

The goal of MOSAIC is to develop a data intelligence capture and analytics platform for multi-modal data sources, including video and textcollaterals. The distributed intelligence within the platform enables decision support for automated detection, recognition, geo-location and mapping, including intelligent decision support at various levels to enhance situation awareness, surveillance targeting and camera handover; these involve level one fusion, and situation understanding to enable decision support and impact analysis at level two and three of situation assessment.

CAPER Collaborative information, Acquisition, Processing, exploitation and Reporting for the prevention of organised crime

The goal of CAPER is to create a common platform for the prevention of organized crime through sharing, exploitation and linking of Open and Closed information Sources. CAPER will support collaborative multilingual analysis of unstructured and audiovisual contents, based on Text Mining and Visual Analytics technologies. CAPER will allow Law Enforcement Agencies (LEAs) to share informational, investigative and experiential knowledge.

