IDC WHITE PAPER

How Semantic Technologies Enable Domain Experts to Steer Cognitive Applications

July 2017

by Nick Tahamtan, IDC #CEMA42878117

Sponsored by Semantic Web Company

This IDC White Paper discusses how companies can utilize semantic technologies to build cognitive applications. It examines the role of semantic technologies within the larger artificial intelligence (AI) technology ecosystem with the aim of raising awareness of different solution approaches.To succeed in a digital and increasingly self-service-oriented business environment, companies can no longer rely solely on IT professionals. Solutions like the PoolParty Semantic Suite utilize domain experts and business users to shape the cognitive intelligence of knowledge-driven applications.

Introduction

Scarcity is the fundamental driver of the marketplace. In the ever-expanding digital economy, however, "scarcity" is being replaced by "abundance" — of data. Yet studies show that less than 10% of data is used effectively by organizations.

Collecting more data does not necessarily result in more value. A digital asset gains value when it can be exploited to provide insight, influence decisions, and set directions. Semantic technologies close the gap between data collection and value creation from data. Every enterprise has a huge amount of data hidden in textual form (documents and files), and up to 90% of enterprise data is unstructured (IDC: Unlocking the Hidden Value of Information, 2014). This information, when properly utilized, can be used to create a better customer experience, bring better products to market, and reduce costs along the enterprise value chain. Companies that can process text-based resources have a critical competitive advantage.

Businesses are making uneven progress in exploiting the data they possess. Many companies are still struggling to make their internal documents searchable in a way that permits the content to be flexibly reused. At the other end of the spectrum, some technology companies are applying AI to rapidly automate and personalize their marketing and service communications. IDC predicts that, by 2019, 40% of all digital transformation (DX) initiatives and 100% of all effective Internet of Things (IoT) efforts will be supported by cognitive/AI capabilities (IDC FutureScape: Worldwide IT Industry 2017 Predictions, IDC #US41883016).

Most organizations lack the talent to utilize the technologies that will help them advance to truly agile data applications. Around 40% of companies say they are struggling to find the talent to implement Big Data solutions (Accenture: Big Success with Big Data, 2014). Another obstacle to technology-driven innovation is that some companies are simply unaware of the available technology solutions.

Embracing semantic technologies to deliver cognitive solutions can enable an organization to substantially reduce its dependency on developers and specialized IT professionals. Once the foundation of semantic data management is in place, the adoption of data-driven applications will be driven by domain experts and business users.

Standards-based solutions like the Semantic Web Company's PoolParty Semantic Suite can reduce a company's dependence on IT experts and provide many technological advantages. AI and cognitive solutions can assess the properties and characteristics of data and create value from it. Semantic technologies automatically enrich resources by providing metadata. They can supply extra contextual information, offering enhanced text analytics and data integration capabilities.

Semantic technologies enable the following:

- **Description Capabilities:** Description is a prerequisite for exploiting the reusability and querying capabilities of data. Semantic technologies describe heterogeneous data uniformly.
- **Context Enrichment:** Data can be enriched by contextual information such as time and location.
- **Interoperability:** The transformation of heterogeneous data in a standards-based data format allows for interoperability between data sources.
- **Advanced Search and Query:** Thanks to a rich metadata layer, structured and unstructured data can be queried and reused on a granular level.
- **Complex Event Processing:** This relates to the integration of data from multiple sources (e.g., sensor data), the identification of meaningful events, and the initiation of an appropriate response.

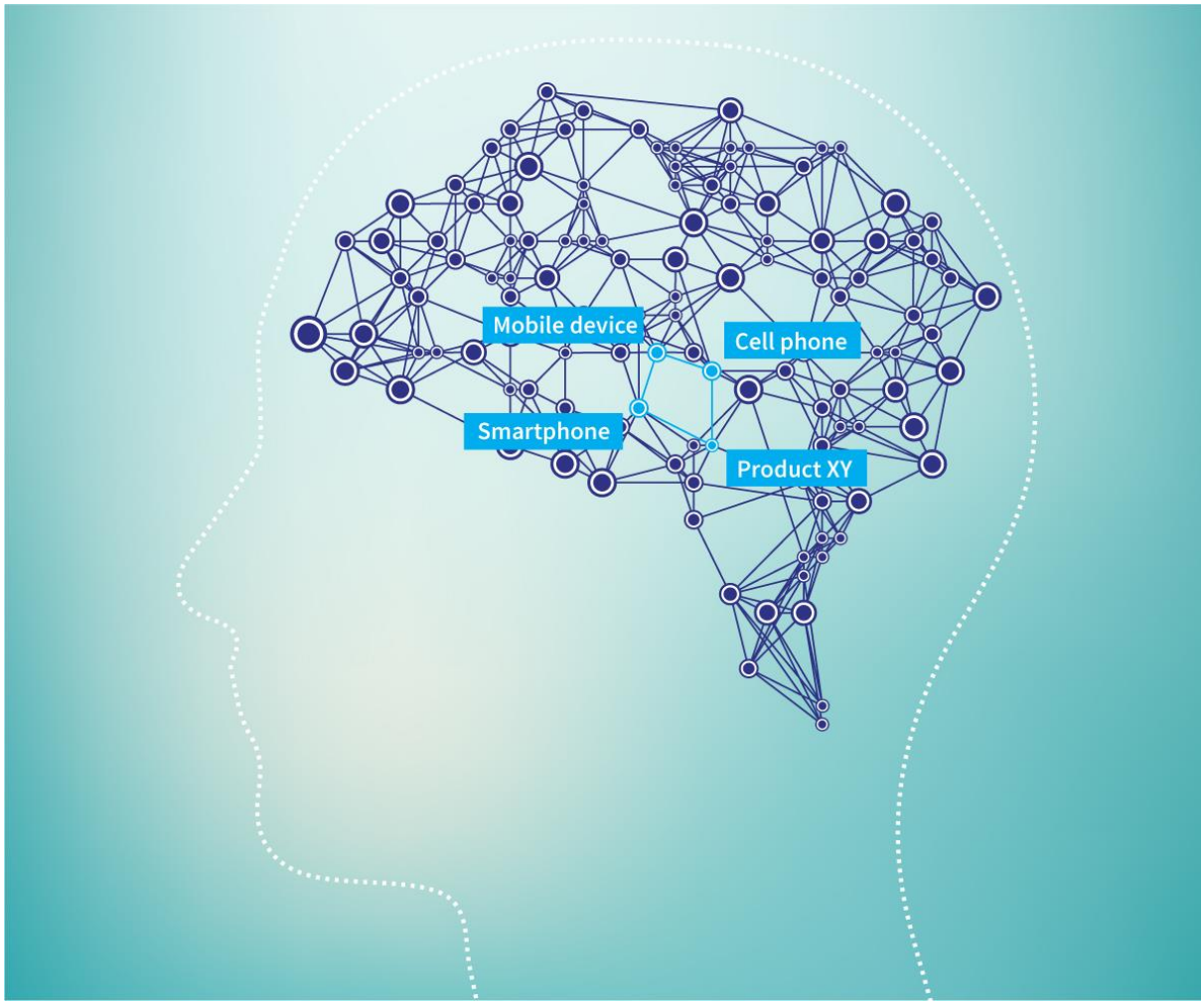
An overview of relevant semantic technology concepts and methods can be found at the end of this document.

The Era of Cognitive/AI Systems and Technologies

By 2020, spending on cognitive applications worldwide is expected to reach $47 billion (*IDC Worldwide Semiannual Cognitive/Artificial Intelligence Systems Spending Guide, 2016*). Exponential growth is occurring in smart applications that go beyond a concrete search query by responding to user questions and generating relevant insights. AI is being applied for a wide variety of uses, including medical diagnostics and treatment, predictive analytics for industrial equipment, customer advice and assistance in retail, fraud detection and prevention in financial services, and governmental citizen support systems.

FIGURE 1

Human Reasoning & Learning in a Nutshell



Source: Semantic Web Company 2017

The human brain's architecture features conceptual networks that enlarge and otherwise evolve through learning. Concepts are units of meaning that link to other concepts, creating contextual frameworks. When humans think, speak, and learn, their reasoning is determined by the relations of their conceptual networks.

Cognitive solutions essentially mimic how the human brain works. The search for cognitive solutions has challenged computer scientists for more than six decades. The research has matured to the extent that it has moved out of the laboratory and is now being applied in a range of knowledge-intensive industries.

There is no such thing as a single, all-encompassing “AI technology.” Rather, the large global professional technology community and software vendors are continuously developing a broad set of methods and tools for natural language processing and advanced data analytics. They are creating a growing library of machine learning algorithms to enhance the automated learning capabilities of computer systems. These emerging technologies need to be customized and/or combined with complementary solutions, depending on the use case.

Cognitive systems can be built on statistical models or knowledge bases — or on a hybrid of these. In the case of statistical models, text resources are analyzed and processed based on rule engines that use mathematical calculations (e.g., that consider term frequency). This approach may be effective with large data volumes, but it does not reflect human reasoning.

Knowledge bases, which are the core of semantic technologies, use knowledge graphs to carry out cognitive computing. Mimicking the human brain, knowledge models organize information using interrelated concepts. (For an introduction to knowledge graphs, see the "Background on Semantic Technologies" section, at the end of this paper.)

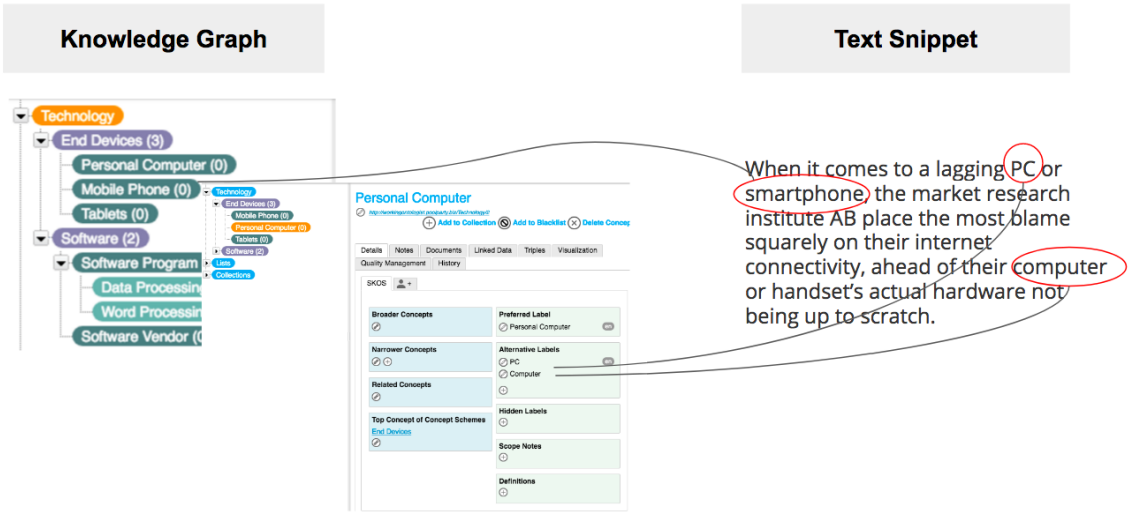
For many years, the creation of taxonomies and ontologies was regarded as excessively time-consuming and was not adopted on a wide commercial scale. However, advances in the semi-automatic creation of knowledge graphs — and their enhancement by machine-learning algorithms — has led to increasing demand in recent years for knowledge-driven approaches.

The statistical model and knowledge base approaches are both available as embedded software solutions. However, the statistical, rule-based, approach can be highly complex for end users and is usually not transparent. It may not be easy for the user to modify the solution if the results are unsatisfactory; indeed, modification of the data model may require the skills of a specialized IT professional.

In contrast, the knowledge base approach uses concepts and relations expressed in natural language, making the underlying data models accessible to domain experts and business users. It is not necessary to write code to include terms for text-mining functionalities; instead, the terms can be included in the knowledge graph. Because this approach mimics human reasoning, it comes natural to human users and is more user friendly than writing mathematical formulas and turning them into computer codes.

FIGURE 2

Text Analytics Based on Knowledge Graphs



Source: Semantic Web Company 2017

A hybrid approach to cognitive computing, employing both the statistical and knowledge-based models, will have a critical influence on the development of applications. Highly automated data processing based on sophisticated machine-learning algorithms must give the end user the option to independently modify the functioning of smart applications.

How Semantic Technologies Work in Practice

Since early 2000, the Semantic Web community has worked to transform the web into a Web of Data. The initiative marks a paradigm shift in large-scale knowledge management, taking it from a document-centric view to an entity-driven approach. The application of World Wide Web Consortium (W3C) data standards for web resources permits relations of the same concepts within different documents. The result is a tightly networked data landscape for search and question-answering functionalities, granting users access to related granular information that was originally available only in separate documents. The principles of Semantic Web technologies have been adopted in enterprise IT management, with the goal being to break down data silos and make unstructured data accessible in multiple contexts and application scenarios.

A new approach to data management must be embraced to allow domain experts and business users to drive smart applications by adapting and extending underlying knowledge graphs. The following are some technical and organizational prerequisites for companies that want to start working with data efficiently and rapidly in real time:

- **Standards-Based Technologies:** Unstructured data comes in many shapes. To be used effectively, it must be transformed into a common format — one that can be matched against a knowledge graph. Semantic technologies are based on Resource Description Framework (RDF) data. This standards-based format includes contextual metadata that can be traversed by knowledge graphs.
- **Graph Databases:** Actionable resources that are highly enriched with metadata must be managed with graph databases. Relational databases, in contrast, process data only via regular tables, with rows and columns.
- **Use Cases:** After a semantic data management infrastructure becomes operational, it is crucial to thoroughly define application use cases and, over time, expand their scope. Many companies fail to deploy cognitive solutions due to their perceived high complexity.
- **Management:** The shape of the knowledge graph should be strongly influenced by applications. Depending on the deployed software, the process of building a knowledge graph can be accelerated by using knowledge graph libraries and support tools. However, the human element remains the most decisive factor in establishing a new regime of information management. A range of stakeholders should be involved in creating the foundations of the smart application. This interdisciplinary approach requires well-structured organizational governance.

Semantic technologies offer many advantages — but integration also involves costs. These include recruiting experts who possess the necessary technology skill sets to develop shared information environments. Costs are also associated with data migration and the integration of the technology within the enterprise IT landscape, not to mention operational expenditures.

Applications Powered by Cognitive Computing

Semantic technologies can impact the entire business of an enterprise and its relationship with customers. Knowledge workers spend an average of 2.5 hours per day searching for information. IDC estimates that organizations' overall search time can be reduced by more than 50% if adequate enterprise applications are provided (IDC: The High Cost of Not Finding Information, 2004). Business-to-consumer and business–to-business purchasing and service behavior is highly information-intensive. Global companies understand this and are increasing investments in digitally enhanced customer experience solutions (Accenture: B2B Customer Experience Survey, 2015).

It can no longer be doubted: Cognitive applications are revolutionizing the workplace. By 2020, some 40% of the Financial Times' Europe 500 companies expect that most of their business will depend on their ability to create digitally enhanced products, services, and experiences (IDC: IDC FutureScape: Worldwide IT Industry 2017 Predictions, IDC #US41883016). The success of these initiatives will depend on how successfully data can be integrated, processed, and analyzed.

State-of-the-art smart applications are evolving into a fusion of search, recommendation, and question-answer machines. The most efficient will cover specific use cases and be defined by a focused knowledge domain. This is how high-precision, machine-generated results will be acquired.

Below are some AI-enabled use cases and application scenarios that are currently prioritized by corporate and digital business players.

Intelligent Content Authoring

Information is "produced" everywhere — and the reuse of content can have a substantial impact on business success in most industries. The duplication of content is widespread, as knowledge workers often do not know where to find interrelated resources. A powerful search portal that analyzes text resources on a granular level — and provides precise results — is a critical part of a modern content publishing lifecycle. Improved efficiency is the result of the ability to immediately reuse content in new work products. In the field of technical documentation, AI-supported content production is already having a significant impact. Text resources are being produced as structured metadata-enriched content "snacks" rather than as full documents. This kind of content is highly actionable: It can be automatically reused across different formats and digital communication channels.

In the field of intelligent content authoring, semantic and cognitive technologies can help in several ways. At the information collection phase, these tools can provide faster and more efficient ways to find, reuse, reconfigure, and adapt relevant information. At the creation phase, they can provide suggestions for new content.

Consider, for example, the case of a knowledge worker composing a smartphone market and technology report. It may be a tedious and time-consuming task to gather information — with no guarantee that all relevant results will be collected. Keyword-based search techniques may offer only limited help: In one document, the device may be referred to as smartphone; in another, as a handheld, mobile device, or cell.

In such use cases, the PoolParty Semantic Suite modules of Taxonomy & Thesaurus Management, Concept Tagging, and Text Mining & Entity Extraction provide suitable solutions. PoolParty integration with authoring tool FontoXML makes enriched content processable via dynamic publishing processes.

Smart Help Desk

Self-service portals were created to reduce costs. Today, many clients prefer to use these hubs because of their speed and efficiency, and they have become essential for companies and government organizations. A self-employed person working on a tax declaration, for example, will likely try to answer her questions on a self-service portal before consulting a human tax advisor. Portals empowered by semantic technologies can provide such information services even in highly complex domains.

When clients formulate questions in natural language, text analysis, natural-language processing, and semantic technologies are essential to match questions with relevant information. The development of an AI-enabled help desk requires a knowledge base steered by a knowledge graph. The knowledge graph should include concepts, synonyms, and relations. It should also include possible questions and their variations. Even in complex knowledge domains, it has been demonstrated that a highly precise help desk can be established after the 500 most frequently asked questions have been determined. Semantic technologies tackle the challenge of natural language heterogeneity and make help desks reliable — and extremely valuable — business enhancements.

The PoolParty modules of Taxonomy & Ontology Management, Concept Tagging, and Text Mining & Entity Extraction offer solutions to the scenario described above.

Cognitive Analytics

Data is the fuel of the knowledge economy. However, astonishingly, most data analytics applications still focus almost exclusively on quantitative data analytics (Accenture: B2B Customer Experience Survey, 2015). Structured and unstructured data analytics will uncover untapped business potential and point the way to efficiency gains. Research and development-intensive industries like healthcare and pharmaceuticals, vehicle manufacturing, and IT are already embracing cognitive analytics to enable their cross-functional expert teams to improve products.

Typical use cases could involve finding publications relevant to a specific topic, research topics linked to a specific problem, and researchers with expertise in a related subject. In these cases, a simple keyword-based search may not be enough. For example, searching on a topic like “bacteria” may not find articles in which the concept is referred to as “germs” or “bacterial chromosomes.”

Cognitive analytics deconstruct resources on a granular content level and match them with key performance indicators such as usage frequency. Typical queries could include:

- Show all documents related to "subject XY" that have never been displayed in search queries.
- Show the most popular documents and their search term histories.
- How many resources are available on "topic YZ"?

These kinds of insight can impact collaboration, research themes, and the fine-tuning of the underlying data models that steer cognitive analytics applications.

It must be emphasized that cognitive solutions are not limited to the cases mentioned above. Semantic technologies can assist companies in a wide range of use cases that rely on agile data integration and modelling.

The PoolParty module of Data Linking & Mapping, in combination with other PoolParty knowledge engineering solutions, enables companies to adopt cognitive analytics.

Software Vendor Profile: Semantic Web Company

Building cognitive solutions requires new ways of technology management on the front and back ends. Many companies consider themselves overmatched when faced with the challenge of assembling single software solutions. Full-scale, or market-leading, solutions might give the impression they are suitable only for large-scale IT-projects, and most companies prefer to develop innovative prototypes at their own pace. Working with a complete semantic technology suite can simplify the transition to dynamic data management.

The PoolParty Semantic Suite is provided by the Semantic Web Company, a pioneer and leading player in semantic technologies. The suite is a semantic middleware platform that includes complementary modules for semantic data integration and knowledge engineering, as well as a toolkit for setting up cognitive end-user applications. Users can license individual modules and extend their usage based on need.

FIGURE 3

PoolParty Semantic Suite Overview

Source: Semantic Web Company, 2017

The PoolParty Semantic Suite has been on the market since 2007. It has been deployed by organizations including Wolters Kluwer, World Bank Group, Credit Suisse Group, Healthdirect Australia, and Boehringer Ingelheim. It is standards based and provides out-of-the box integrations with leading content management systems, data management systems, search engines, and graph databases, including SharePoint, MarkLogic, Stardog, GraphDB, Sinequa, Drupal, and Confluence.

The Semantic Web Company and the PoolParty product have received multiple industry honors for excellence in the knowledge management field. PoolParty's strengths include:

Proven Innovation

PoolParty uses proven and innovative technologies to help companies improve the sophistication of their information architectures. The company has a strong track record in implementing its software in multiple industries.

Expanding System Intelligence

The PoolParty system learns continuously from information streams and documents. Applications and self-service portals gradually become more precise and reliable.

Pricing

A modular pricing model enables PoolParty to be used by small and medium-sized enterprises and large organizations.

Data Agility

Companies can increase the value of data by applying the W3C Simple Knowledge Organization System (SKOS) standard. PoolParty enables an organization to gradually develop customized structures without being overwhelmed by data complexity.

User Experience

PoolParty can be easily utilized by developers, information architects, and content strategists.

Background in Semantic Technologies

Semantic technologies are mature. Semantic technologies are referred to as technologies that enable or facilitate the use and/or interpretation of meaning by computers. Semantic technologies are based on different standards. Standardization offers several advantages: easier integration of complementary technologies, avoidance of vendor lock-in, and support from established experts and practitioners. Some of the most important concepts and standards are as follows:

- **Knowledge Graph:** A knowledge graph represents a domain's knowledge by making explicit the concepts and relationships between entities. Rules for interrelations between classes may be defined in an ontology. Knowledge graphs are the backbone of cognitive enterprise applications and semantic information management systems. Knowledge graphs substantially improve search, personalization, and recommendation capabilities.
- **Metadata:** Metadata is data about data. It describes the underlying characteristics of the data. Metadata may include keywords describing the content and structural information such as content type, date of creation, and author. Metadata can be managed by knowledge graphs.
- **Data:** The RDF developed by W3C is a standard, flexible, and extensible model for data interchange on the web. A variety of schemes can be mapped with RDF, enabling data processing irrespective of the underlying data model.
- **Identification:** A Uniform Resource Identifier is used to identify an entity. Each concept is included just once in a knowledge model but can be linked multiple times depending on the context.
- **Query Language:** SPARQL is a semantic query language to query RDF data.
- **Knowledge Models:** The Simple Knowledge Organization System (SKOS) is a standard for building knowledge models. Based on RDF, it includes a set of modeling rules. As its name suggests, SKOS can be used by domain experts who lack specialized skills in knowledge modelling.
- **Ontology:** Web Ontology Language (OWL) is an advanced knowledge modelling standard based on RDF. OWL enables knowledge engineers to model complex domains in the most specific way.

Complementary Semantic and Cognitive Technology Solutions

- PoolParty Semantic Suite: **https://www.poolparty.biz/**
- FontoXML: **https://fontoxml.com/**
- GraphDB: **http://graphdb.ontotext.com/**
- MarkLogic: **http://www.marklogic.com/**
- Sinequa: **https://www.sinequa.com/**
- Stardog: **http://www.stardog.com/**

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com