How Semantic Analytics Delivers Faster, Easier Business Insights

Improved analytics of the big data already at their fingertips can help transform organizations for the digital age, giving them answers to pressing business questions and uncovering previously unknown relationships and trends.

Executive Summary

It's long been known that organizations that can act on fact-based insights gain competitive advantage. Reducing the time from raw data to informed insight to action reinforces and extends that advantage.

But for many organizations, their analytic efforts are hobbled - not by too little data, but by too much. We are now in an age where data is measured in zettabytes (ZBs, each of which is a billion terabytes), pouring into the enterprise from a growing number of internal and external sources. IDC estimates that the amount of data generated by everything from point-of-sale systems to sensors will rise from 4.4 ZB in 2013 to 44 ZB in 2020.1 Sadly, many organizations have too little information about what this data represents, and how it has been or could be used. Even when the organization does have this information, it’s often held by “data experts” who are too hard to find or too busy to help front-line business analysts with their queries.

With the ongoing explosion in the volume, variety and speed of data flows, organizations face a two-fold challenge:

• Understand what data is most relevant and accurate to meet a given business opportunity.
• Make it easier and quicker for business users to perform analyses and understand and share the results, without needing to understand the underlying databases or complex data science principles.

Semantic analytics delivers these capabilities, and more. It uses metaknowledge - knowledge about the nature of, and relationships among, your business data - to allow even nonspecialists to uncover the complex, nuanced insights needed to meet today’s business challenges.
Zipping from Metadata to Metaknowledge

<table>
<thead>
<tr>
<th>Metadata About Zip Codes Used in Traditional Analysis</th>
<th>Metaknowledge About Zip Codes that Drives Faster, Lower-Cost and More Insightful Semantic Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is a required part of a customer record.</td>
<td>Zip codes represent a bounded geographic area served by a post office in the U.S. They may cover an entire town or multiple neighborhoods in a town. A city may include multiple zip codes, with some residents of adjacent zip codes living in close proximity, with potentially similar demographic or socioeconomic characteristics.</td>
</tr>
<tr>
<td>Must be either five or nine digits.</td>
<td>The equivalents of zip codes outside of the U.S. include many different formats and may be referred by other names such as postal codes. The concept map should define which characteristics are applicable for which countries, and how they relate to certain categories of analyses.</td>
</tr>
<tr>
<td>Is updated from customer-provided information.</td>
<td>In new markets, postal codes may come from third-party providers and are often less accurate than those in established markets, and thus require validation. The metaknowledge describes the automated techniques that can be applied to perform that validation.</td>
</tr>
<tr>
<td>No insight into relationship of postal code to other data such as street address.</td>
<td>If there is a discrepancy between street address and postal code, the postal code is often more accurate and should be used for transmission of legal documents or physical delivery. In some cases, organizations might want to allow customers to provide their own addresses if they find that most convenient.</td>
</tr>
</tbody>
</table>

Figure 1

This white paper explains how semantic analytics empowers business users to identify new opportunities and market trends more quickly and at lower cost than ever before. It also describes specific use cases, from marketing to regulatory compliance and security, where semantic analytics can deliver compelling business benefits.

Semantic Analytics Defined

Metadata (data about data) has traditionally provided basic information about data, such as who created it, its format and a brief description of its contents. Metaknowledge allows users to apply semantic analytics to modern information systems, giving them richer, deeper and more nuanced information to solve business problems using data their organizations already own. Metaknowledge also gives users a broader view of problems and opportunities, and the ability to see relationships they might not otherwise have discerned.

This metaknowledge includes the relationships between a data element and other data, context about its meaning and when a new context can change the meaning. It also includes the origins and nature of the data, and expert knowledge about how to interpret it. Semantic analytics uses this metaknowledge as a road map to guide and perform the analysis of that data, providing insights, correlations and understanding that otherwise would require a data scientist's expertise. It also allows the easy, and even automatic, sharing of such metaknowledge among applications, rather than keeping such knowledge locked within the business rules and program code of individual systems. (See Figure 1 for an example of the differences between metadata and metaknowledge.)

The metaknowledge that makes semantic analytics valuable to users includes:

- **Commonly used synonyms, abbreviations or even misspellings of frequently used terms:** With semantic analytics, for example, a user need only query for customers who “liked” a product and automatically find those who also thought it was “neat,” was “cool” or “rocked.”

- **The relationship of a data element to other, related data elements:** A user who asks for all customers who like “team sports” can receive a list that includes fans of soccer, football (American and otherwise), rugby and basketball, but not tennis or gymnastics. This could be used to identify prospects with similar interests or characteristics on whom to focus customer-conversion marketing efforts.

- **The relationship of a generalized category of data with other, more specific subcategories, such as degrees of satisfaction or dissatisfaction, and the ability to combine those categories in future analyses:** An auto
maker, for example, might wish to identify an “excited but value-conscious” category of auto buyer and provide these individuals with customized marketing messages, financing offers or trim options.

- The context of the data, such as expert knowledge that provides possible explanations for “out of range” data: For example, an analytics platform may generate alerts when weekly purchases of a product in a geographic area fall above or below a certain range. The “expert knowledge” here might be a human noting that a surge in milk sales happened before a bad snowstorm or a surge in snack sales during a major sporting event. This context can not only offer an explanation for that specific event, but embed that knowledge (such as weather data or sports schedules) into future analyses. As more context is added to the semantic model, it empowers even nonexpert users to perform more and more “expert” analyses of the data over time.

Business Value of Semantic Analytics
Semantic analytics helps the business in two ways. The first is to more precisely and easily analyze data to get answers to known questions. These “known unknowns” might include “How do iPhone vs. Android users feel about my mobile app, and which factors influence this sentiment?” “What do female millennials think of my clothing brand?” “What data is most relevant to analyzing purchase trends?” “Has new data been onboarded recently that can improve this analysis?”

The second way semantic analytics delivers business value is through analysis of the metaknowledge itself to find previously unknown relationships and trends. These “unknown unknowns” might include “Are any of my stock traders showing signs of suspicious behavior?” “Are new patterns of anomalous behavior emerging among my users, and what exposures do they create?” “Are there undetected attacks happening in my network?” It can also improve future analyses by informing users about what data is being used more often over time and becoming increasingly relevant, and thus should automatically be included in future analyses (such as the effects of weather and sporting events on food sales).

Quick Take

Case in Point: Sentiment Analysis for a Motor Vehicle Maker

Sentiment analysis based on social media is often limited to binary choices, such as positive vs. negative, that fall far short of capturing the complexity of customer sentiment. It often cannot, for example, distinguish between a mere satisfied customer of a car brand and a far more valuable brand champion or influencer, or the stages a customer goes through on their way to a purchase decision.

For example, automobile manufacturers looking to expand their share of the fast-growing SUV market need to understand what specific features various customer segments value most. Semantic analytics could identify, for example, “budget luxury” customers who through their social media comments signal their interest in a vehicle that appears luxurious but costs only slightly more than a mid-level model. Through detailed analysis of their comments the manufacturer could pinpoint which “luxury” features (e.g., leather seats and a stitched dashboard) are most important to them, and which features (e.g., good mileage or rear seat comfort) are less important. These insights can be a “real world” complement to conventional market research such as focus groups and even identify potential customers for such products.
It can also suggest known correlations (such as between adults traveling with young children and sales of extra space seats on an airline) to suggest to business analysts. Also, it can point to dimensions that should be collapsed or expanded to make the data easier to understand, such as combining data elements for home sales in “coastal” and “non-coastal” postal codes if proximity to the coast is not relevant for a given query. All this makes it easier, faster and less expensive to uncover business insights.

Semantic analytics is especially useful for some of the most complex analytic challenges that involve hundreds or thousands of dimensions of data and complex relationships among them. These include social media sentiment analysis, detecting security threats in complex IT environments and identifying fraud in intricate financial transactions. It helps meet these challenges by abstracting both the complexity of the dimensions and the relationships among them, and by embedding expert rules and knowledge within the data.

To more accurately and completely capture the sentiments of tens of thousands or millions of customers, a business may need to analyze a hierarchy of terms, to rank some higher than others or to understand how terms relate to each other. This includes not only which words are synonyms or antonyms, but which terms seem to be antonyms but are actually synonyms. (Examples include a Tweet that an actor is “really bad” or a song “rips it up.”) Semantic analysis not only provides such context, but can describe common abbreviations or even misspellings to assure more accurate analysis.

Semantic data also allows businesses to create and combine multiple, interoperating “mental models” of data to identify potentially significant
actions, trends or individuals. These may use taxonomies, concept maps or knowledge graphs to describe data. Such a mental model might organize the details of customer behavior (such as interest in or avoidance of a product or brand), demographic attributes such as age, gender or past actions such as purchase history (for a customer) or access attempts (by a potential hacker). These mental models can be easily shared among users and applications, providing ongoing cost and time savings in future analysis while empowering more and more users to perform increasingly useful analyses.

Semantic analytics also allows the use of multiple concept maps to analyze the same data using different terms, or the use of the same terms in different contexts. One example is combining the genre of a movie (such as action/adventure) with demographic information about the viewer (is he a child or an adult?) to infer whether their use of the term “terrifying” means they were pleased or displeased.

These concept maps can be used to represent any type of relationships, from the parts, subassemblies and assemblies that make up a motor vehicle to the roles and units within an organization or a business. This extends the number, and type, of business-critical analyses they can improve. For example, a concept map can assign names to any particular level of detail of the terms within it, reducing the time, cost and effort required to run natural language queries with a “Google-like” ease of use.

**Looking Forward: Near-Term Applications**

Semantic analytics is not a replacement of, but an enhancement to other analytic methods, using graph algorithms to generate metrics which can then be tabulated in conventional relational databases or using expert knowledge and cognitive techniques to speed and enrich the analytic process (see Figure 2).
Integration of Concept Maps

Concept Map for Product
Semantic analytics thus empowers more decision-makers to make better informed decisions more quickly than ever before. Many organizations already have the data and expertise they need to achieve this. Adding semantic technology and semantics practices to their analytics activities will enrich their analyses for additional competitive advantage (see sidebars on pages 3 and 4 and this page for case illustrations).

Quick Take
Case in Point: Detecting Security Anomalies

Sifting through millions of events in thousands of network devices and servers for hints of malicious attacks is one of the most complex, but most business-critical, analytic challenges facing a modern business. As with fraudulent trades, the aim is to identify the “normal” behavior of networks, devices and users so security analysts can better identify abnormal conditions. Semantic analytics enables automated systems to go beyond a simplistic check of whether, for example, traffic to or from a given port falls outside a normal range. It also allows such a system to learn which combinations of dozens of network characteristics are most likely to indicate an attack, and which other metrics it should check if one measure falls outside the normal range.

The use of semantic technology also allows security professionals to create concept maps, including abstractions of various event hierarchies, that allow for the graphical analysis and correlation of events and alerts to help fight threats. The metaknowledge associated with these events might include, for example, known defects in certain hardware and software, and the knowledge that the exfiltration or export of sensitive data must have been preceded by surveillance and intrusion of the compromised systems.

As is the case with finding hints of fraudulent transactions, information security organizations need the skills and tools to first capture knowledge from human security experts about patterns that have, and have not, signaled security breaches in the past. Semantic databases are required to capture this knowledge in easily accessible form, along with natural language processing tools so users can easily execute queries against it. As the automated systems learn about new types of threats, or gain more insights into older threats, semantic analytics makes it easy to add new systems, behaviors or threat types to the analytic process.

Looking Long-Term

Generating business insights requires combining the right data, algorithms and expert knowledge. Existing technology allows organizations to automate the data management and algorithm-driven analytics. Semantic analytics automates the third required element. By embedding expert knowledge in the data, it empowers businesses to generate faster insights even as the variety, volume and velocity of data increases at an exponential pace.
About Cognizant

Cognizant (NASDAQ: CTSH) is a leading provider of information technology, consulting, and business process services, dedicated to helping the world's leading companies build stronger businesses. Headquartered in Teaneck, New Jersey (U.S.), Cognizant combines a passion for client satisfaction, technology innovation, deep industry and business process expertise, and a global, collaborative workforce that embodies the future of work. With over 100 development and delivery centers worldwide and approximately 233,000 employees as of March 31, 2016, Cognizant is a member of the NASDAQ-100, the S&P 500, the Forbes Global 2000, and the Fortune 500 and is ranked among the top performing and fastest growing companies in the world.

Visit us online at www.cognizant.com or follow us on Twitter: Cognizant.

World Headquarters
500 Frank W. Burr Blvd.
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233
Fax: +1 201 801 0243
Toll Free: +1 888 937 3277
Email: inquiry@cognizant.com

European Headquarters
1 Kingdom Street
Paddington Central
London W2 6BD
Phone: +44 (0) 20 7297 7600
Fax: +44 (0) 20 7212 0102
Email: infouk@cognizant.com

India Operations Headquarters
#5/535, Old Mahabalipuram Road
Okkiyam Pettai, Thoraipakkam
Chennai, 600 096 India
Phone: +91 (0) 44 4209 6000
Fax: +91 (0) 44 4209 6060
Email: inquiryindia@cognizant.com

© Copyright 2016, Cognizant. All rights reserved. No part of this document may be reproduced, stored in a retrieval system, transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the express written permission from Cognizant. The information contained herein is subject to change without notice. All other trademarks mentioned herein are the property of their respective owners.