Enhanced Business Performance

Ontologies-Based Data Integration

Introduction

Companies are constantly seeking new ways to improve products, services, and support for their customers. This requires a thorough understanding of customer needs, which is most often accomplished by effectively utilizing information systems and applying Business Intelligence (BI) tools capable of converting data into meaningful, actionable information. Customer data, an important aspect of any successful business, is used by various entities within an organization to perform multiple job functions. For example:

- Product strategy, product management, and marketing may have a need to review customer data related BI reports in order to design better products and services.
- Customer care may have a need to collect and analyze customer information from multiple data sources to serve customers better.
- Internal IT may have a need to consolidate disparate customer data dispersed over multiple systems in order for the business to function smoothly. Business needs may include customer care, billing and processing, and BI data for intelligent decision making.
- Businesses may have a need for a Decision Support System (DSS) to process, review, and overlay complex data results depicted visually in order to arrive at a satisfactory customer-related decision.

It is clear that a comprehensive data view across various domains or areas of operation is best suited to support these activities. However, with customer data dispersed over multiple systems and often represented differently, semantic mismatches between terms and undefined or underrepresented relationships may cause pertinent data to be unavailable in a seamless manner when required by interested users and applications. Obviously, partial data and information-driven analysis can negatively impact decision quality and, therefore, business performance.

The use of data integration processes attempts to solve this problem. However, conventional data integration processes are poor at capturing concepts, relationships, and semantic differences between dispersed and disparate data. The context in which data is used, concepts within a particular domain (*e.g.*, products/services and customer service), and the frequently complex relationships between each of them are often lost or not easily captured. Therefore, full or meaningful data availability issues still persist, and businesses are forced to make decisions based upon partial data.

In today's globally competitive environment, enterprises that possess a comprehensive understanding of customer data have an edge over their competition due to better understanding of customer response to various stimuli. Businesses are coming to understand that this knowledge can be very effectively utilized in behavioral marketing.

Emerging trends in software ontologies and ontological engineering can help solve or reduce the issue of poor data integration. This paper examines how ontologies can support a fuller range and quality of data availability. We will examine a simple use-case scenario in the field of healthcare and demonstrate the difference ontologies can make in improving customer service. Though the example is in the healthcare field, similar ontological principles can be applied to various other fields, such as e-commerce, inventory management, bio-pharmacy, oil and gas, energy, GIS, speech processing, and many others.

Data Integration Support Using Ontologies

Dynamic data's ontological representation can fill the conventional data integration gap of mismatched data and lack of understanding of relationships between concepts. Ontological representation is useful where data needs to be inferred, relationships between data across the enterprise need to be assembled, and relationships between concepts need to be examined in order to formulate a coherent and intelligent picture that is usable by multiple applications and in intelligent decision making.

Such an inference mechanism and standardization of data with a common vocabulary can be accomplished using ontologies as an underlying structure, and when a comprehensive view of the data across various domains (*e.g.,* customer relations, sales, service, etc.) does become available, a positive impact upon business performance occurs.

By utilizing ontologies, businesses are enabled to understand how customer behavior is impacted by the actions within different company domains and then act upon that information. Because ontologies are particularly adept at making correct inferences based upon limited information, a comprehensive view across company domains translates into better decision making, faster data analysis, and consequently—better products and service to customers.

This naturally leads to greater profitability, reduced costs, and faster data processing—all of which help businesses remain ahead of their competitors.

Ontologies Overview

Ontologies and ontological engineering assist with storing hierarchical data and their relationships directly into the database structure. The result is fast queries, data accuracy, and inferences derived from existing data. Ontological structure also assists with the reuse of existing information systems, which allows for significant cost savings.

Ontology has multiple definitions and interpretations. As it relates to information systems, ontology defines relationships between data and its concepts—and also the relationships between concepts themselves. Ontologically represented data helps extract and correlate knowledge stored across multiple information systems or a single domain, such as customer service. Ontologically represented data also assists with knowledge sharing and data reuse between different information systems.

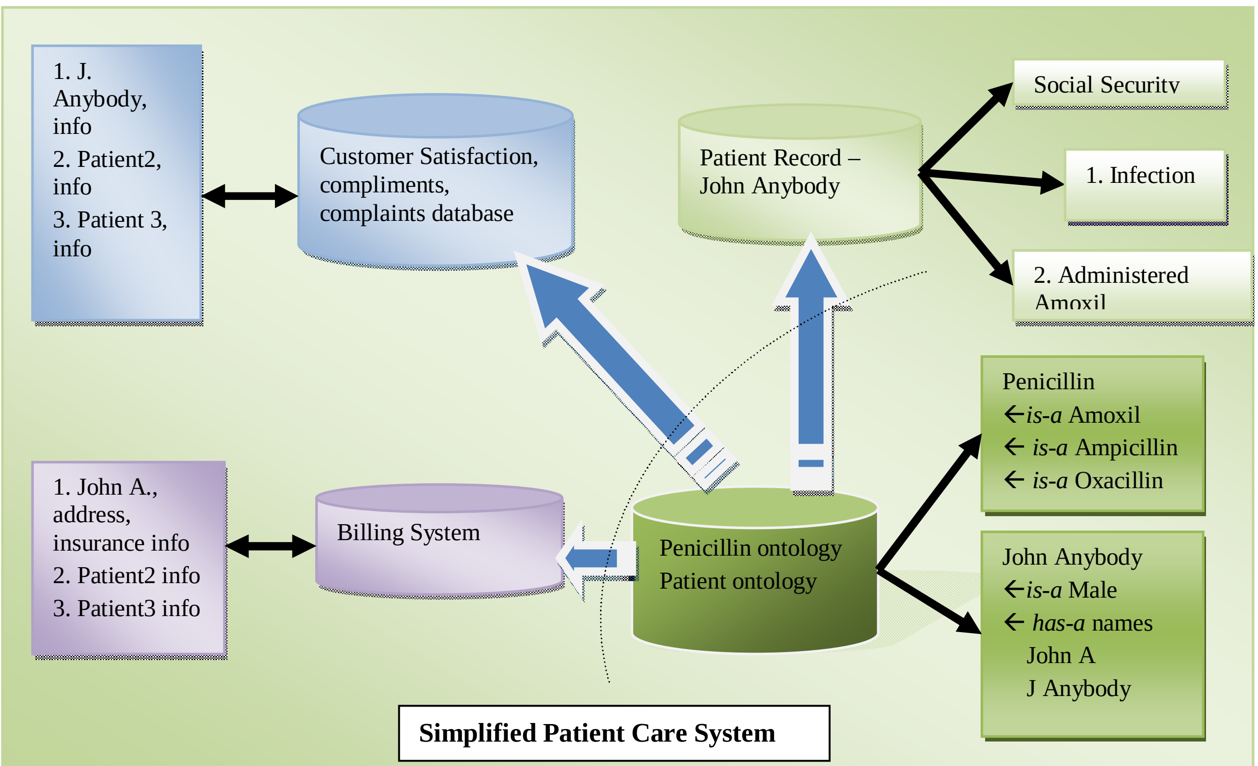
Simply stated, ontologies programmatically help accomplish what humans infer, using different data and conditions/context in which the data is being used. Such human-like inference capabilities bring tremendous advantages to business performance in terms of speed (and faster product launches), data accuracy (and thus better business intelligence), and provide a coherent picture for customer care personnel (which increases customer satisfaction). And the beauty of ontologies is that all of these results can be inferred from sparse information provided by customers.

Conventional systems utilize rule-based technologies to accomplish similar results in inferring data. A rule-based system utilizes "if-then-else" as one of the rules to draw a conclusion and present results. For example: If A is related to B, and B is related to C, then A is related to C.

However, such systems have a difficult time adapting to dynamic data traversing deeply embedded data or a tree-like structure, and then correlating disparate information. This can occur, for instance, when a product is moved from one category to another. Due to the nature of their implementations, ontologies help reduce the impact of these changes.

Patient Care Use Case

To understand the following Use Case effectively, consider the picture labeled "Simplified Patient Care System." It depicts a current state, which is to the right of the curved dotted line, and a post-ontology implementation state, which is to the left of the line. The post-ontology state shows two ontological structures—drug related ontology and patient (or "customer") related ontology.



Simplified Patient Care System

Let us consider the pre-ontology based system as the patient care system. Patient "John Anybody" visits a hospital for an emergency treatment for a serious infection. John's information is captured in the patient record database that maintains a history of prior and current treatment. John's name is stored in the system as "John Anybody" along with his social security number and address as "123 Main Street." John's condition is diagnosed, and he is administered Amoxil antibiotic for the treatment of his symptoms.

The billing system bills John for his co-pay and the deductible as per his insurance plan. The billing system, however, maintains John's name differently from the patient record database. The data in the billing system reflects his name as "John A." and his address as "123 Main St."

After a few days of follow-up patient care and John's recovery, a customer (patient) satisfaction survey is conducted. The customer care database captures John's experience and his commendations or complaints, if any. This system maintains John's name as "J. Anybody," and his home address as "123 Main St".

Due to a processing error, John's entire record is requested by the customer care administrator with his full name "John Anybody." Due to a name mismatch, a complete and consistent set of records for John cannot be found in all the systems, so only partial data can be retrieved. The administrator then must log into every system and query multiple times using different name and address combinations until a matching record is found in each system. The administrator then compiles John's data as a single comprehensive report to review the case history, payment, insurance claim, and other data in order to be able to effectively address the processing error.

This costs company time and money, and the customer can perceive the truly diligent effort on the company's part as "lack of interest in resolving my issue." This, of course, is due to the amount of time required to resolve the issue and the inconsistent customer information. Consequently, despite diligent efforts, the customer in this case responds unfavorably to the customer satisfaction survey.

Data Consistency Issue

In the scenario above, the issue is that multiple systems have stored John's information differently. In other words, the systems cannot "talk" intelligently to each other. Even if they have common interfaces to communicate, term confusion, mismatched data (*e.g.*, "John Anybody" name stored differently) and semantic differences between them may not produce the desired query results for John's record.

Consider John's information in the different systems:

- Patient Record System – "John Anybody," social security number, case history, drug administered, address as "123 Main Street"
- Billing System – "John A," social security number, address as "123 Main St."
- Customer Care System – "J. Anybody," no social security number, address as "123 Main St"

There are name mismatch, as well as common identifier, issues across the systems. John's name is stored three different ways in the three systems. Consequently, multiple queries based on social security number and name and address combinations must to be performed to collect John's information. A new report that consolidates all the information has to be compiled to meaningfully address John's issue.

Add to this, hundreds and thousands of patients who are handled by a hospital every day—or the number of claims an insurance carrier, for instance, must address for patients—and the issue becomes huge. Needless to say, hospitals and insurance carriers frequently spend an inordinate amount of time and resources addressing data disparity issues, despite corporate-wide CRM/ERP systems.

Prior to proceeding with the Use Case, but with the ontologically engineered solution, let us examine the two ontologies that are introduced to the solution.

Ontology Based Solutions

The patient/customer scenario above and variations thereof represent a common situation in many companies, especially those that have data dispersed over multiple systems. Corporate IT departments attempt to keep data standardized across different platforms. However, this is a Herculean task due to a wide variety of reasons, including new-system installs, multiple data collection points, storage methodologies, and—as in the previous healthcare example, the way in which a patient's data is entered into each domain or system.

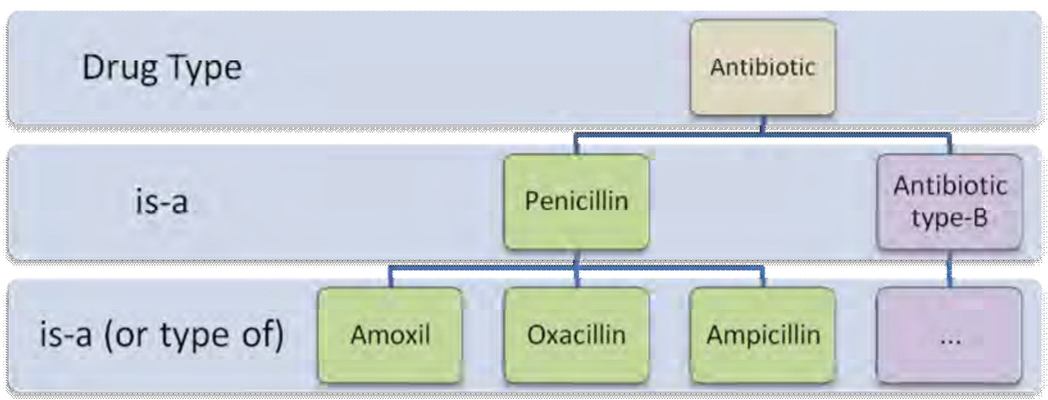
When ontologies and data mapping are combined, however, they aid tremendously in reducing processing times, increasing data accuracy and consistency, and establishing relationships so that they thus become a basis for meaningful interpretation by applications and users.

Let us now consider two brief examples of ontologies and their applicability to the healthcare scenario.

<u>Penicillin Ontology</u>

Amoxil, Ampicillin and Oxacillin all belong to the penicillin family. They work differently, but their active ingredient is penicillin; therefore, their relationship can be quantified as:

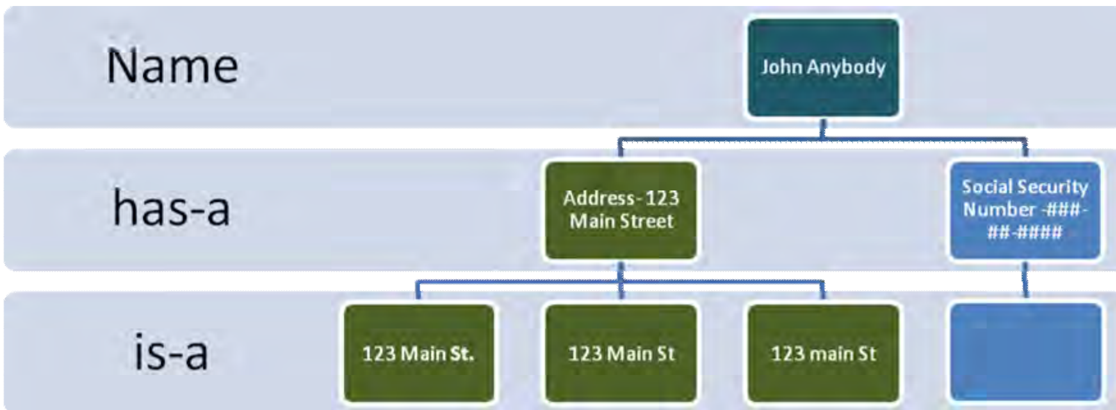
Amoxil *is-a* penicillin
Ampicillin *is-a* penicillin
Oxacillin *is-a* penicillin
and
Penicillin is an antibiotic



A query by using the word "antibiotic" will result in different antibiotics available, of which penicillin is one. As the example proceeds, one will see how context information leads to the inference of the three drugs listed below penicillin. Alternatively, a query on Amoxil may yield the related terms of Ampicillin and Oxacillin, and one of these terms may help build the proper context in joining this information with queries from other ontologies.

Patient Ontology/Data Mapping

The following is a patient ontology, utilizing "John Anybody" again. For simplicity, the example will work with address mapping-related ontological structure only. Similar solutions can also be envisioned for other pertinent patient data, such as drugs administered, allergies, etc.



All three variations to the addresses are mapped back to the address of "123 Main Street" for John Anybody. Therefore, query on any of the four types of addresses will always result in John's residence. It's easy to see, therefore, that by employing ontologies, a positive customer care survey outcome will be much more likely.

Use Case Analysis

Using ontologies, information residing in different information systems and domains can easily be extracted. To carry the illustration one step farther, if John were to call and report an issue, the interaction may now go as follows:

John calls the customer care center and informs them that he has a billing discrepancy. When asked, John indicates that he does not remember the exact name of the drug, but he remembers that it was some type of penicillin. Upon the request of the customer care representative, the ontological inference engine brings up Amoxil, Oxacillin, and Ampicillin as possible choices. John is asked the type of infection he had, and that results in the system's reviewing another ontology which helps determine the kind of penicillin generally provided for this type of infection. Note that *relationships* between domains are being examined here.

The customer care rep then describes the symptoms for the nearest drug match, and the patient is subsequently able to confirm the symptoms. Likewise, utilizing various ontologies that map patient provided data, and through a series of questions asked of the patient, the customer care rep is able to determine that the drug administered was Amoxil. Similarly, John's data is further narrowed by searching for his address as "123 Main St" and customer care's backend query is able to locate the correct address of "123 Main Street." Consequently, John's issue is satisfactorily resolved, resulting in a happy customer.

In this case, the use of ontologies helped join data from different information systems and knowledge domains vis-à-vis diagnostics performed, drug rendered, its effect on the patient, a billing error, and a rapid fix of that error. Significantly, all of this was accomplished by simply adding ontologies (and servers to host them) and by reusing existing information systems with *no change* to either their database representations (schemas) or to their code. Therein lies the power of ontological representations.

Conclusion

Since ontologies define relationships between groups of terms and are helpful in joining data from other systems and domains, they are extremely useful for integrating data across multiple systems without the expense of re-architecting the entire solution. Where data is complex and very hierarchical, programmatic representation of this data via ontologies can play an important role in assisting the various business, process, and customer care entities in performing their functions more effectively. Examples of complex and hierarchical data are:

- A drug and its derivatives
- A machine and its parts
- An insect species and sub-species

While the example provided was one of data integration in the healthcare field, the applicability of ontological engineering is limitless and applicable to almost any field. Of course, it goes without saying that data modeling and data mapping must be performed correctly and in an exacting manner that pertains to a particular industry, company, and domains of interest.

An added benefit to the use of ontologies is that, once architected properly, businesses can continue to reap increased benefits over time as the ontological structure becomes more intelligent with every new discovery of relationship or data. This represents a true leap forward as companies strive to manage their data in meaningful, comprehensive ways that serve to enhance business performance and increase the bottom line.