BEST
p r a c t i c e s

The Center for Information-Development Management

DECEMBER 2022
VOLUME 24, ISSUE 4

From Information Wars to Knowledge Wars
And why business leaders should care, right now

Michael Iantosca, Avalara, Inc.

THE WAR YOU DIDN'T KNOW YOU WERE FIGHTING
If you are a senior decision-maker, you're more than likely in an existential business battle in a war you didn't know you were fighting. Web3 is virtually upon us. The web has been the graveyard of countless businesses that didn't keep up, and Web3 could bring your own company's demise.

Bring the subject of Web3 up at any social gathering and you'll get widely differing opinions of what it will be. Web 2.0 ushered in information wars with the advent of the social web and eCommerce. The defining elements of Web3 remain vague to many. The marquee applications, and blockbuster monetization of them, are yet to emerge. But make no mistake, they're coming, and they may be coming for you.

Some will say that the defining technologies of Web3 are blockchain and cryptocurrency, while others will say that it's metaverses and virtual reality, while yet others will say it's the maturing of AI and machine learning. They're all correct. Just as the foundation of Web 2.0 was comprised of multiple converging technologies, so will Web3. But just as information was the single-most identifying artifact of Web 2.0, knowledge and the relational interconnecting of knowledge assets is the single most enabling technology of Web3. If you find that a bit perplexing and vague, there's the first clue that you have yet to become fully informed as a decision-maker.

History informs us that "the next big thing" doesn't just suddenly happen. Such a catharsis is preceded by years of technological development to which few had been paying much attention.

Many companies that paid attention crushed the competition, amassed incredible wealth, or both. The smart ones were paying attention as these nascent technologies were in the incubator while everyone else was scratching their heads wondering how they missed the last platinum opportunity that has already made others rich, disrupted, or devastated their business model. As usual, they're late to the party. There are many examples, Google, eBay, Facebook, Amazon - the list goes on and on.

Take the web itself. The technologies that became the foundation for the web existed long before the first web browser, Mosaic, went public in 1983. Interconnected computing was occurring long before Berners-Lee invented the protocols for the web. He couldn't have even conceived of the web without the invention of the internet itself. The internet was invented in the 1960s. You read that right, the internet is not the web, yet many erroneously think they are the same. They're not. Long before the web, many technologists were surfing the internet, including yours truly – without the web or a web browser.

The development of the first network switch at UCLA in 1969, followed by email in 1972, the IP and Ethernet communication protocols in 1973, and then the modem in 1977 were all foundational technologies that made what we know as the web, possible. Even HTML wasn't anything new; it was a loose application of an otherwise boring standard called general markup language (GML) based on hypertext invented in the 1960s to standardize formatting codes for (get this) – industrial printers!

CONTENTS

Michael Iantosca
page 1

The 12 Days of Christmas
Dawn Stevens
page 2

Reveal Your Value by Building Cross-Functional Partnerships
Miriana Nedeva
page 8

Outsourcing: Working with another company's content
Jennifer O Neill
page 11

Manager's Calendar
page 15

CIDM Sponsors Highlight
page 16

(continued on page 3)

710 Kipling Street • Suite 400 • Denver, CO 80215

1

(continued from page 1)

Fast forward to the early 1980s when many of us were happily tunneling between mainframe servers using archaic command-line instructions, downloading software and documents, and posting on character-based bulletin boards and discussion forums long before the first website was even conceived. Of course, the advent of the IBM Personal Computer and Apple Macintosh (that followed the Apple I and II) democratized computing for the public and opened the door to mass connectivity for the general public.

People made fortunes with Web 1.0. Who could have known? Many of us. We were building and evolving those foundational technology elements for quite some time. The magic, and the monetary opportunities, arise when the right subset of technologies converge, at the right time, with the right application. The applications are opportunistic; their invention is only made possible due to multiple foundational inventions on which they otherwise wouldn't exist.

So, is Web3 blockchain? Crypto? Artificial Intelligence? Machine Learning? Virtual Reality, and Metaverses? Partially, yes, these are only a subset of the foundational technologies of the applications that Web3 will spawn. The killer applications that will define Web3 will use a combination of these. However, it's the sleeper, boring technologies that become the tipping point. Just as Web 2.0 was powered by information, Web3 will be powered by knowledge and knowledge assets including taxonomies, ontologies, and knowledge graphs that will provide the rocket fuel for Web3. By the time most organizations learn about the necessity of semantic assets, it will be too late for most to catch up.

You might rightly ask: how does information differ from knowledge? That's a fair and insightful question. It can seem perplexing at first but being aware of one's perplexities and questioning them is part of gaining knowledge; too many dismiss their perplexities to their demise.

Just as data isn't information, information isn't knowledge. Data can be collected and organized into information that enables comprehension. Information answers the who, when, what, or where elements of knowledge. Information combined with experience, intuition, and inferencing, yields knowledge. Knowledge answers the eternal question of why. Knowledge increases understanding and consciousness. Information is readily transferred; knowledge requires learning, and learning is the act of association followed by assimilation. Knowledge with experience enables predictive outcomes where information alone cannot.

In the early 1990s, rules-based artificial intelligence systems called Expert Systems, based on obscure programming languages such as LISP and Smalltalk attempted to mimic knowledge but failed. It failed because it was based on logic and rules instead of objects and relationships between those objects. Machine learning emerged years later and has been incorrectly conflated with AI. Machine learning removes the linear logic and instead is based on objects, their relationships to one another, inferencing, and association, much like what we as humans do to acquire knowledge.

What is Cognitive Content

Cognitive content is a strategy, an architecture, and an operational model that enables dynamic, machine-based discovery, mining, analysis, retrieval, assembly, and delivery of non-linear content objects to provide humans or machines using advanced knowledge technologies that are based on predictive relationships between content objects and inbound signals.

Cognitive content takes intelligent content one step further that enables an entirely new genre of useful content-based applications. Cognitive content combines self-describing content objects with enriched structured metadata and object-orient relational (semantic) technologies. In a word, you can consider this next generation of content object-oriented content, much like object-oriented software.

Michael Iantosca
Avalara, Inc.

Michael Iantosca is the Senior Director of Content Platforms at Avalara Inc. Michael spent 38 of his 40 years at IBM as a content pioneer – leading the design and development of advanced content management systems and technology that began at the very dawn of the structured content revolution in the early 80s. Dual trained as a content professional and systems engineer, he led the charge building some of the earliest content platforms based on structured content. If Michael hadn't prevailed in a pitched internal battle to develop an XML platform over a planned SGML variant called WebDoc in the mid-90s at Big Blue, DITA, and the entire industry that supports it, might not presently exist. He was responsible for forming the XML team and a member of the workgroup at IBM that developed DITA.
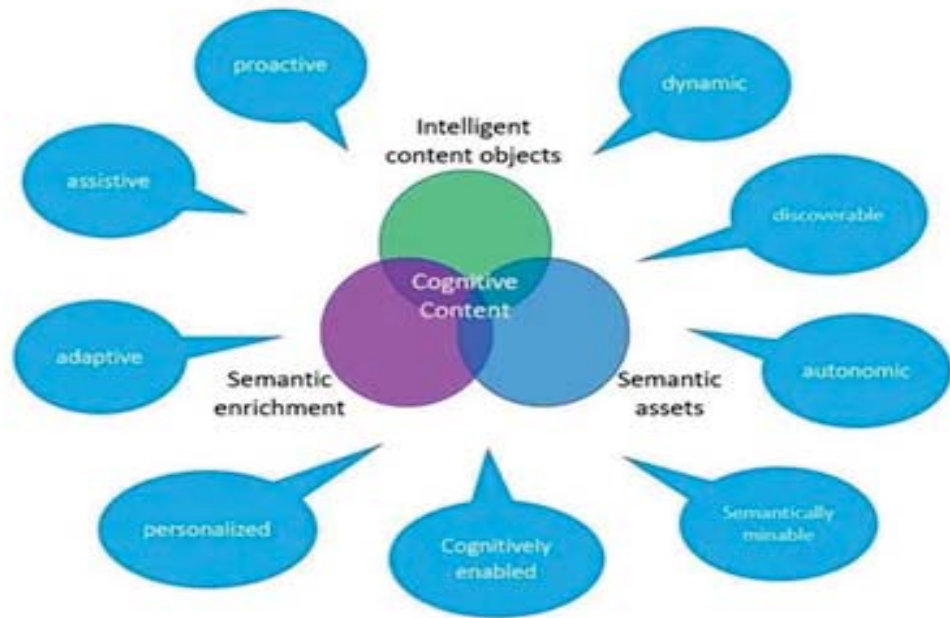
Sleeper technologies of Web3

While technologies such as blockchain, cryptocurrencies, NFTs, and metaverses grab the headlines, there's a group of technologies that, quite frankly, are desperately boring. Yet just as there would be no web today without the cryptic protocols of TCP and IP, so will there be no Web3 without these.

As a group, they are called semantic technologies and semantic assets. And as usual, they're not new, but have evolved over many years and are only now starting to gain attention in technological circles.

Berners-Lee coined the term "The semantic web" as far back as 1983. He envisioned the web being driven by AI and machine learning. The term semantic is about meaning. Just as humans associate and assimilate, it is the recognizing, understanding, perceiving, and then responding. That was quite insightful by Berners-Lee who had a vision of the power of linked data and information. But that wasn't enough. The notion of a semantic web couldn't be realized using only the foundational technologies that powered Web 2.0. What was needed were methods to represent semantic knowledge.

Would you hire your content?

If your content were a person, would you hire it?

Many business leaders would yell a resounding "No!" How do we know that? A recent study by Forrester, Component Content Management Systems: Supercharge Long-Form Content for Personalized and Consistent Experiences, bears this out in hard numbers from a diverse sampling of decision-makers.

Would you hire a person if all that they had to offer was information and not knowledge? Of course, not; information is finite. We want skilled people with knowledge and the ability to

grow and apply that knowledge. We live in the age of the knowledge worker, so then why do we settle for content that offers only information and not useful and relevant knowledge?

Forrester generated significant insights by surveying 450 decision-makers across fourteen countries with organizations that have annual revenue exceeding $500 million in revenue. A few startling findings:

Top challenges in content delivery:

♦ Production and delivery of personalized content (58%)

♦ Omnichannel publishing/content delivery (55%)

♦ Speed of publishing and updates (51%)

♦ Content consumption and customer insights (42%)

♦ Providing a rich documentation post-sales experience (34%)

Fifty-two percent also said that lack of integrated content analytics means that it is difficult for them to know if the content is successful or impactful with 49% saying that their post-sales help and support content experience is not as rich as their pre-sales content. Yet a whopping 67% believe that it's important to provide a great post-sales customer experience. Much of Forrester's findings corroborate prior studies. A study by IBM found that 88.7% of respondents believed that post-sales content was important or very important in customers' initial purchasing decisions. Other studies found that technical content generates more than 50% of viable sales leads (Forbes), is the second-most important pre-sales activity for technology buyers (IDC) and encompasses more than 55% of sales cycle time (vs. 21% spent talking to sales staff (Marketing Interactions).

All these studies and more, in aggregate, put to rest any doubt of their veracity. It's time for the C-suite to act.

So why the disconnect? And what are the solutions to these problems? The Forrester study and those like it tell only half of the story. They lead one to believe that the solution lies in a combination of tools, but just as information isn't knowledge, tools are not solutions, nor are they valuable business assets like knowledge. Tools come and go; knowledge persists and can also be greatly monetized.

The historical barriers to preventing the advancement of content as a burden to content-as-a-service (CaaS) have been latency and opacity. Latency because the ability to measure and identify deficiencies in the effectiveness and impact of content, especially post-sales content, occurs later in the customer journey, and opacity because it has been near impossible, short of occasional sampling and vague user feedback, to know how customers use content, whether the content is relevant, and if and how it results in user action and success. Some business leaders intrinsically know that post-sales content in particular plays a significant role in product sales, adoption, retention, and advocacy, yet there are many more that erroneously view it as a cost center only to be minimized since they haven't been able to accurately measure it.

That's changing. The content industry is on the precipice of making a generational leap – from providing information to providing knowledge. That knowledge will find its way to the customer and business leaders alike. The knowledge we're referring to will manifest itself as an entirely new genre of applications that were previously not feasible or technologically possible. This new crop of knowledge applications will smash the barriers of latency and opacity to bits and reveal the true value of and impact of content. Most importantly, these new knowledge applications will eliminate friction for the customer which leads to faster time-to-value, greater and more effective self-service, higher satisfaction,

better customer retention, improved revenue along with stronger upsell and cross-sell. Eventually, business leaders will come to view content as a competitive weapon, not just a cost center.

WHAT ARE THESE CONTENT KNOWLEDGE APPLICATIONS OF WHICH WE SPEAK?

The applications of which we speak are not buildable with conventional engineering tools and methods. They require the use of advanced semantic technologies including enterprise content taxonomies, ontologies, and knowledge graphs. The mere mention of taxonomy, let alone ontology and knowledge graphs conjure up visions of particle science and string theory for too many. Let's finally put these notions to rest – like or unlike Schrödinger's cat depending on your view.

Long the domain of Library Sciences graduate and doctorate programs, these knowledge technologies are now finding their way into the general business realm. They've been vague and poorly understood, mostly because too few content professionals come equipped with advanced semantic technology expertise. As technical communicators we ought to be ashamed, it's our job to investigate, learn, simplify, and communicate complex subject matter and simplify it for our customers. Worse, these are not head-scratching, mind-numbing concepts – they've simply not been explained well, nor has their immense business value been adequately framed. Many technologists toss around terms such as AI, and machine learning; the lack of a clear roadmap to practical applications has understandably led to a certain degree of doubt and skepticism.

TAXONOMIES, ONTOLOGIES, AND KNOWLEDGE GRAPHS – OH MY!

The core knowledge technologies that provide the infrastructure needed to build this new class of semantic applications are well-defined. They are best understood using the semantic content maturity model as a guide.

The semantic content maturity model (SCMM)
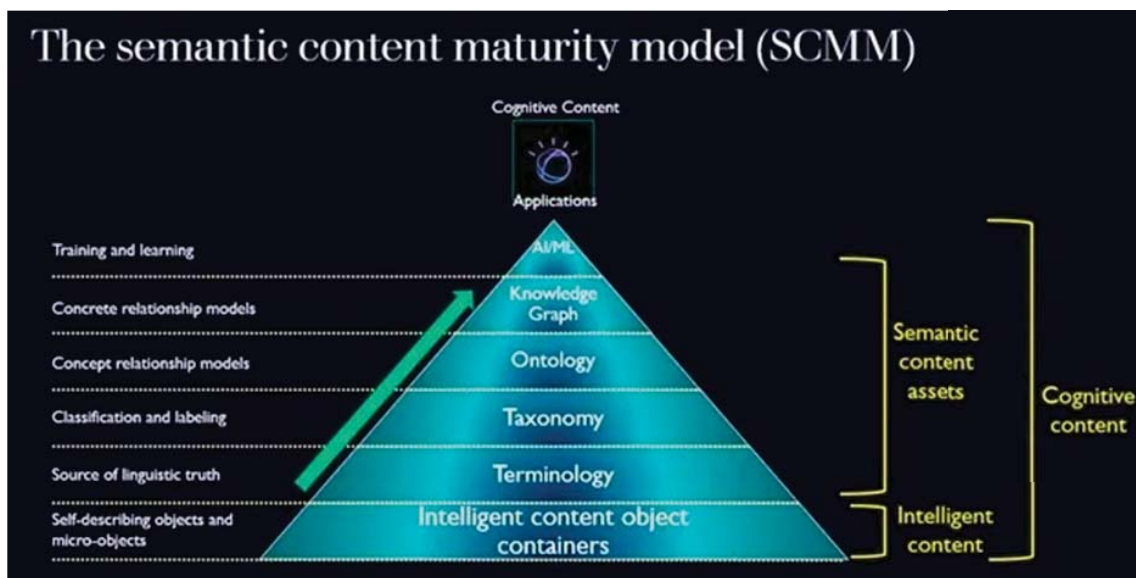
OBJECT-ORIENTED CONTENT

No doubt you understand the significance and value of object-oriented programming, but have you ever considered an object-oriented approach for content? Not all objects are created equal.

What organizations need to do is make a shift from monolithic, unstructured blobs of content to intelligent content. Many companies have already done so or are in the process of making that transition with technologies such as the Darwin Information Typing Architecture (DITA). So, what is intelligent content more precisely?

Intelligent content is modular, structured, reusable, separates format from presentation, and is semantically enriched such that the content is highly predictable for machine processing and automation. However, for our purposes of enabling cognitive content using artificial intelligence and machine learning, the most important aspect is that intelligent content is enabled for powerful and predictable algorithmic machine processing.

If your organization hasn't started making the transition to intelligent, object-oriented content, you are already late to the party.

TERMINOLOGY

We often forget that the very basis of human, and machine communication, are words. Words are the very foundation of language and understanding. Words matter. Yet we take language for granted; after all, we are all language experts, right? Wrong. You say tomato, I say tomato. You call it a hoagie; I call it a submarine sandwich. Heck let's call the whole thing off!

Before we even get started with the sexy semantic technologies (and I mean sexy only to a Library Science major), there's terminology. For most, terminology management is the least sexy of the semantic technologies, yet it is the most important. Forego formal terminology creation management and you might as well build a skyscraper on a bed of quicksand. Yet, to our astonishment, which is what many are doing.

Taxonomies are constructed from words, as are ontologies, knowledge graphs, and content itself. If words are used inconsistently across these we might as well speak different languages. A common terminology base serves as the lingua franca for your entire semantic stack. Without it, one is building a stack of cards that will inevitably collapse upon itself. Yet how many C-suite leaders understand why having a corporate terminology database managed by skilled terminologists is a core business necessity and investment every bit as important as product development? Sadly, few do, and the lack of having an enterprise termbase significantly hurts the business in hidden costs, let alone the opportunities we're discussing here. Establishing a managed enterprise termbase is step one and to be bypassed at one's own peril.

TAXONOMY

Some business leaders, but still far too few, recognize the existence of taxonomies in their organizations. Taxonomies exist because they are fundamentally required to build and operate the many business systems that companies rely upon.

A taxonomy is simply a grouping or hierarchical classification of things using labels. The list of your company's products and features is a taxonomy, whether one recognizes or calls it one. If you've used eBay and navigated a subject tree, from photography to cameras, and then down to lenses you've seen and used a taxonomy. We define taxonomies whose labels we can assign to our content objects as self-describing (metadata). There are manual and automated ways to formally create, manage, and apply these labels. Taxonomies are also used to build ontologies.

ONTOLOGY

If the mere mention of taxonomy makes your eyes water, discussions involving ontology are the stuff of migraines, but in reality, it isn't complex at all.

A great analogy top knowledge expert at Amazon used to describe ontology is how we learn as children. We learn based on concepts and objects. A young child learns that he or she lives in a house, and then we learn that the house has a front door, a sidewalk, and more houses (concept). There's a specific house next door where a playmate, Jackie, lives. Then we infer that if we want to play with Jackie we need to go over to our playmate's house – Jackie's house to be specific. While a house is a generic concept (there are many houses and many types of houses) we learn that the one next door is specifically Jackie's house. Jackie's house is a real-world object. What I just described, in technical terms, is called an ontology, which is a mapping of concepts and how each concept relates to one another, and a knowledge graph - actual real-world objects that use the ontology as an overlay, or a model. There's a bunch of folks out there that will use complex language to describe these that are suited for only Ph.D. dissertations – but as a decision-maker, it's no more complex than what I just described unless you want to get into the nuts-and-bolts of how this works and how they're implemented to create real-world business applications.

These knowledge assets are what powers Amazon suggestions and how social media seemingly magically knows you have an interest in horses for example when you think you didn't even say anything about horses. I'll bet dollars-to-doughnuts that some of you wonder if Siri is somehow secretly eavesdropping on you without permission more often than not.

KNOWLEDGE GRAPHS

As the pioneers of the US West once said, "There's gold in them thar hills!" But our gold isn't coming out the streams or mountainside by itself; you must do a bit of digging to get at it.

Just as an ontology is a mapping of concepts and their relationships, a knowledge graph is a mapping of real-world objects and their relationships. A knowledge graph uses an ontology as the model. In software programming terms, one might equate a knowledge graph where a schema is a model, and the actual implementation is a specific instance. A blueprint for a new building might specify a certain size front door, the actual door might be a specific make and model from a particular supplier at a specific price.

But why? What's the use and value of a knowledge graph? By mapping out these relationships we can add additional detail about the concepts and objects, how they relate to one another, and their characteristics. After we have that intelligence machines can discover new relationships and make inferences, we as humans cannot make, especially at scale. Those new capabilities break the dam holding back knowledge that can be used to create an entirely new genre of business applications that were previously not feasible nor possible, and the key to competitive success or failure. You know that when your competitors possess valuable knowledge your company doesn't know it can be deadly. Imagine that at scale; you might as well be flying blind. It will be as if you're fighting in the pitch darkness of night while your competitors are wearing night-vision goggles. Good luck.

Measuring for success

What steps can you do now and how do you progress to value? There are core questions you can and should ask of your content professionals.

♦ Are we designing and communicating with a single voice?

♦ Does our organization have a managed enterprise terminology base?

♦ Are we using consistent common vocabularies governed by taxonomy?

♦ Can we map and mine our engineering and content assets using knowledge graphs?

♦ Which applications enabled by knowledge graphs are yielding the greatest value?

There are many more questions you should use to interrogate your staff. What has been opaque and hidden is no longer a barrier. The type and number of applications that a knowledge-based approach enables are limitless. Let's take a (minuscule) look at the types of applications that semantic graph-based technologies can solve.

♦ Direct correlations between usage and customer satisfaction

♦ Effectiveness of user experience correlated with conversions

♦ Alignment/misalignment between engineering models preventing integration

♦ Gaps, overlaps, and conflicts in user content

♦ True personalization of dynamic user assistance

♦ Effective support case deflection with far greater relevance

♦ Scalable, effective chatbots that make dynamic use of "Big Content"

This is merely a sample of countless use cases. When you can correlate one body of data to another, and AI generates inferences that humans cannot, you're better armed with the knowledge needed to make fully informed, measurable, and empirical decisions.

Want to know more?

From microcontent to neurons: A practical guide for building a cognitive AI content supply chain for highly personalized user assistance is freely available at https://thinkingdocumentation.com/downloads Michael Iantosca, a content industry leader with more than 40-years of hands-on experience as a consummate content practitioner has generously developed this guide. It provides a detailed roadmap loaded with substance to help organizations understand and build a semantic content ecosystem.

Optimizing the DITA Authoring Experience
Instructor: Brianna Stevens
COMtech

Two-Hour Sessions Held Weekly Every Thursday January 12 — February 16, 2023
Design guidelines for making it as simple as possible for your authors to create and publish content that conforms to your information model and authoring guidelines.
https://comtechservices.com/training/dita-optimizing/