Analyzing and Optimizing Content for the Semantic Layer

Paula Land

Content Analysis Defined

Content analysis, or content auditing, is the process of assessing content against a set of defined, measurable criteria and the needs of the business and the content's audiences. The analysis also considers the content operations surrounding the planning, creation, production, and distribution of the content to all the systems that consume it, whether those be web content management systems, knowledge portals, digital asset management systems, enterprise search, or AI-enabled features.

When conducted in advance of the integration of content into a semantic layer—a structured representation of data, content, knowledge assets, and the relationships among them–content analysis becomes a powerful tool for preparing content for understanding, categorization, and enrichment.

The outcome of an audit is insight into the content's quantity, quality, and structure, and a set of recommendations for content improvement. When the semantic layer adds context and metadata to that analysis of the content, you have a content source that is well-optimized to support whatever your organizational needs may be. Examples of how content integrated into a semantic layer can be leveraged include

- knowledge discovery, as in a semantic search program or knowledge portal

- content generation–for example, the automated creation of reports or development of chatbots
- content recommenders that use the rich data about the content combined with user behavior to suggest relevant content.

Factors Affecting Content Readiness

When preparing content for a semantic layer, the quality of your overall content ecosystem is essential. Factors such as content duplication, recency, and availability will affect the content's readiness. Within that ecosystem, targeted content will need to be assessed along several dimensions to ensure effectiveness and usability:

- **Structure:** Content that has been modeled and structured in its source system allows for easier categorization and relationship identification.
- **Semantic richness:** Through techniques such as semantic tagging and entity extraction, the content can be enriched with metadata that describes its context, topics, and entities.
- **Presence and quality of metadata:** Metadata, including tags, keywords, and annotations, is vital in describing the content's context, meaning, and relationships.
- **Consistency and standardization:** Content should adhere to consistent formatting, naming conventions, and data standards to ensure interoperability and ease of integration within the semantic layer. Consistency facilitates accurate

data interpretation and enables effective knowledge extraction.
- **Content quality:** Content that is well-written to be current, accurate, and useful affects not only the end user experience but is critical for reliable semantic analysis and inference.
- **Volume and complexity:** The amount of ingested content can affect the performance and scalability of consuming systems.

By addressing these aspects of content preparation, you can enhance the effectiveness and value of the semantic layer, enabling more accurate, intelligent, and context-aware knowledge representation and discovery.

Auditing Content for Readiness

Designing a content audit to assess the readiness of content for ingestion into a semantic layer involves taking a structured approach to evaluating the content. General principles for auditing content begin with setting objectives and scope, including the metrics by which you will measure success; inventorying and categorizing the content sources for the semantic layer; assessing content structure; determining how well understood, available, and consistently used metadata is across content sources; analyzing domain-specific content; and evaluating the overall quality of the content.

When auditing content readiness for different use cases involving a semantic layer—such as knowledge discovery, chatbot development, content recommendation engine development, or content generation—you will need to tailor the steps to address each application's unique requirements and goals. Here's how you might frame the audit steps differently for each use case:

See Appendix for Figure One: Auditing Content for Readiness

General Principles for All Use Cases

- **Flexibility:** Each step must be adapted to the specific needs and characteristics of the use case. EK worked with the marketing operations team at a global telecommunications company to define the structure of products, the relationship between product components, and the taxonomy necessary to traverse those relationships. The product content model now enables the intelligent assembly of sales collateral, as well as multi-channel publishing to multiple user experiences including sales enablement portals, marketing websites, mobile applications, and social media.
- **User-centric approach**: Consider the end-user experience and how content will be consumed or interacted with. For example, EK recently worked with a global software vendor who needed to deliver more personalized, timely, and relevant release notes of upcoming product changes in a continuous implementation, continuous delivery (CI/CD) environment to both internal and external end-users. EK focused on developing a comprehensive content model supporting structured and componentized release note content, improving user experience (UX) interactions, and leveraging the organization's taxonomy to filter the content for more personalized delivery. We leveraged human-centered design practices and facilitated a series of focus groups across the teams of content authors, marketing, technical SMEs, and executive leadership to define the current state of content authoring processes and content management and ensure cross-team alignment on the target state for authoring, content management, and structured content model design. EK carefully considered the stakeholder

requirements in our delivery of the solution.

- **Feedback loops:** Implement robust feedback mechanisms to continuously improve content readiness and alignment with the semantic layer's goals. EK worked with a [bioscience technology provider](#) who needed EK's help to bridge the gap between product data and marketing and educational content to ultimately improve the search experience on their platform. EK incorporated ML, knowledge graphs, taxonomy, and ontology to redefine the user experience, allowing users to discover important content through an ML-powered content discovery system, yielding suggestions that resonated with their needs and browsing history. EK's flexible approach allowed for open dialogue and iterative development and value demonstration, ensuring that the project's progression aligned closely with the evolving needs of our client.

Taking Action on Your Audit

The outcome of your audit, as stated above, will be a set of issues to address that were uncovered by the analysis. To address those issues and ensure that your content is well-prepared for integration into the semantic layer and optimized for various AI-enabled applications, follow these steps.

1. **Optimize content structure:** For easier categorization and better integration within the semantic layer, reorganize and reformat content to adhere to standardized formats like JSON, XML, or DITA. Break down large content pieces into smaller, reusable components that can be individually tagged for targeted, dynamic assembly and delivery. For example, you might convert a set of technical manuals from PDF format to a structured content model that breaks out individual topics for reassembly and reuse in multiple contexts.

2. **Ensure consistency and standardization:** Improve consistency and interoperability of content by developing and enforcing content standards and naming conventions across the content domains. Implement governance as part of your content operations, by creating a style guide, training content creators, and providing templates that automatically enforce these standards.

3. **Enhance metadata quality:** To improve searchability and context-awareness in consuming systems, add missing metadata and improve existing entries. Use tools for semantic tagging and entity extraction to automatically generate and add metadata tags (for example, author names, publication dates, keywords, and abstract summaries for a set of articles) where they are missing.

4. **Increase semantic richness:** To enhance automated reasoning and inferencing capabilities, add detailed annotations and semantic tags to content. Use a natural language processing (NLP) tool to identify and tag entities and relationships in a database of content. Add these annotations directly into the content's metadata fields.

5. **Improve content quality:** Achieve higher reliability and accuracy by implementing a quality assurance program that includes regular audits to review content and plan for updates. To make this process more efficient, consider using a content management tool that scans content to find and correct errors and inconsistencies and identifies duplicated or outdated content.

6. **Optimize scalability and performance:** Enable efficient and reliable content operations that manage ingestion and processing by scaling up and optimizing tool capabilities and refine workflows for handling large volumes of content.

Implement iterative improvements: Enhance content readiness and alignment with semantic layer requirements on an ongoing basis. Establish a continuous improvement cycle based on feedback and audit findings. Set up regular review meetings with content teams to assess progress and adjust strategies. Analyze user feedback and audit reports. Track improvements and report status to content stakeholders.

Follow these audit steps and address the outlined factors to unlock your content's full potential to drive better decision-making, improve user experiences, and achieve greater operational efficiency.

Looking for expert assistance with your content audit project? [Contact us](#).

Conclusion

Incorporating a semantic layer into content management and distribution is a transformative approach to addressing the organizational challenges of handling vast amounts of information. As outlined above, the process begins with a comprehensive content analysis or audit to evaluate the readiness of content for integration into the semantic layer. This step is critical for enhancing the content's structure, semantic richness, metadata quality, consistency, and overall quality.

By enriching content with metadata and contextual information, organizations can significantly boost the capabilities of consuming systems such as AI-enabled chatbots, knowledge management systems, recommendation engines, and enterprise search. The audit process identifies gaps and provides insights into the content's quantity, quality, and structure, along with actionable recommendations for improvement.

Key factors such as structure, semantic richness, metadata quality, consistency, content quality, and scalability play vital roles in ensuring the effectiveness of content within the semantic layer. Properly structured and componentized content enriched with detailed metadata ensures accurate categorization and relationship identification, facilitating automated reasoning and inferencing.

Addressing these aspects during content preparation optimizes the content for semantic enrichment and enhances downstream distribution in relevant contexts. This leads to more accurate, intelligent, and context-aware knowledge representation and discovery, ultimately maximizing the content's potential in the semantic layer.

Appendix

Figure One: Auditing Content for Readiness