Carlisle Construction Materials
GDC Integrates Artificial Intelligence for Automated Metadata Tagging in Sitecore Content Hub

GDC
Technology You Can Trust.

🕐 4 Min Read

Carlisle Construction Materials LLC (CCM) is a diversified manufacturer and supplier of premium building products and related technologies for the commercial and residential construction markets. CCM has been a recognized leader in the roofing industry for half a century, offering high-performance single-ply roofing solutions that include EPDM, TPO, PVC and roof garden systems.

The Business Need

In the world of digital content management, organizations are constantly faced with the challenge of efficiently extracting and managing valuable metadata from a large volume of assets. For a prominent supplier of premium building products that continues to be at the forefront of the digital age, CCM needed a solution around cataloging and managing an extensive repository of assets stored within their Sitecore Content Hub environment. Among the crucial metadata, one challenge was extracting accurate information from various document formats, including PDFs. The initial approach of manual extraction and categorization of metadata proved to be time-consuming, resource-intensive, and prone to errors due to the diverse range of document formats. Furthermore, the limitations of Sitecore Content Hub's native AI Analysis tool, which could not process PDFs and lacked customization options, created a pressing need for an alternative, more efficient solution.

The Solution

To address this challenge, GDC IT Solutions (GDC) opted to use innovative Artificial Intelligence (AI) and Machine Learning (ML) technologies to create automated metadata tagging. They created an AI model for CCM that was built to employ state-of-the-art Machine Learning algorithms to intelligently detect and extract content. This content was then tagged as metadata from the assets stored within their Digital Asset Management (DAM) system called Sitecore Content Hub. The project unfolded in three key phases.

> "Over the past year, CCM has greatly increased its focus on leveraging Artificial Intelligence for both efficiency gains and breakthrough innovation— many projects being larger in scale. With their expertise and access to the latest technologies, our partners at GDC have proven to us that AI-related projects don't always need significant time and resources to deliver substantial value.
>
> **Robert Miley**
> Supervisor of Web Development, CCM

Data Collection and Training with Advanced ML Techniques: The first step involved collecting a diverse dataset of sample documents from Content Hub. These samples served as the basis for training the custom AI model. By annotating and confirming the locations and values of metadata in these documents, the AI model began to learn how to recognize and extract this vital information.

Custom Model Development: The process of creating an effective AI model proved to be iterative, given the diversity of document types. The team had to experiment with different model architectures, fine-tune parameters, and adapt the model to accurately identify metadata in various contexts. The model was trained using machine learning techniques, incorporating Natural Language Processing (NLP) for text-based documents and image analysis for other formats.

Integration with Sitecore Content Hub: Once the custom AI model reached an elevated level of confidence in detecting and extracting metadata, it was integrated directly into Sitecore Content Hub's workflow. Assets from Content Hub were obtained via its accessible REST API, and each document was fed into the processing pipeline for the AI model to obtain the appropriate information. The integration was designed to operate effortlessly and seamlessly within CCM's content management system. The result was a streamlined process, ensuring the automated metadata tagging was accurately extracted and associated with each asset. Content Hub's extensibility enabled the development of a customized solution, setting a real-world precedent for the application of AI driven technology in a business environment.

The Results

The implementation of a custom AI model to automate the extraction of content from the assets as metadata in Sitecore Content Hub yielded significant results, proving to be a creative and highly effective solution to an otherwise tedious problem. The outcomes of this project were as follows:

Efficiency Gains: The custom AI model significantly accelerated the process of metadata extraction. Manually reviewing and cataloging over 5,000 assets within Sitecore Content Hub wou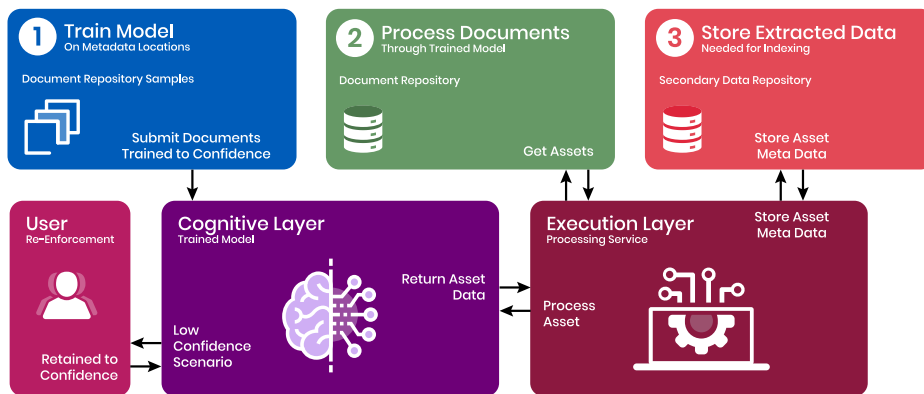ld have been time-consuming and resource intensive. With the AI model in place, this task was automated and performed with high accuracy.

Cost Savings: The automation of this process reduced the labor costs associated with manual data extraction. Additionally, the avoidance of potential errors in date extraction improved data accuracy and quality, saving resources that would have been expended in rectifying mistakes.

Resource Optimization: Human resources were freed from repetitive and time-consuming data entry tasks, allowing them to focus on more strategic and creative aspects of content management and production.

Scalability: The custom AI model is adaptable and scalable. As CCM's content library grows, the model can be updated and fine-tuned to accommodate new document formats and variations, ensuring continued efficiency.

The integration of a custom AI model for the extraction of metadata in Sitecore Content Hub demonstrated how creative and technology-driven solutions can solve real-world content management challenges. GDC successfully streamlined the metadata management process for CCM, resulting in savings of time, money, and resources. This enhancement also led to improved accuracy, consistency, and an enhanced digital experience within its content catalog, serving as a testament to the power of AI in transforming content management workflows.