

Content Enrichment

An essential strategic capability
for every publisher



Bricks

Enriched content. **Delivered.**

Content Enrichment: An essential strategic capability for every publisher

Overview

Content is at the centre of everything a publisher does. Enriching that content delivers significant value across the whole content lifecycle. When implemented as a core capability, content enrichment delivers efficiencies in content production, supports the delivery of better information products, improves author services, adds value for the content consumer, improves marketing results and increases sales. Content enrichment should be a core component of every publisher's content strategy.

Content enrichment is the application of modern content processing techniques like natural language processing, machine learning and Artificial Intelligence (AI) to add structure, context and metadata to content to make it more useful to humans and computers.

Enriched content supports many internal and customer goals. When implemented as a strategic capability, rather than as individual features within single products, it delivers multiple benefits, for example having a semantic fingerprint for a piece of content can deliver a peer review recommender tool, but it can then also deliver other features like a relatedness feature, a smart notification feature etc.



Publishers are exploring how they remain relevant, how they maintain and grow their position in the publishing marketplace and how to differentiate themselves from their competitors. Publishers want to improve the authoring experience, create efficiencies in content production, reuse content across products, deliver more value to content consumers, deliver new content services, deliver value add product features, and make content more usable and discoverable. The problem is that publishers are often stuck with unstructured and siloed content, legacy technologies that are no longer fit for purpose, heavily manual content processes, a limited understanding of what modern content processing tools can achieve and technology teams that are already overburdened.

So far, most publishers have only dabbled in using modern content processing tools or they have implemented spot solutions for specific needs that do not deliver value across the organisation. If content enrichment is implemented as an organisational capability it delivers multiple benefits across the publishing lifecycle. Taking a strategic approach will maximise the return on any investments made.

67 Bricks' work with publishers has identified benefits from content enrichment in every part of the organisation (editorial, content production, product development, IT, sales, marketing, and finance). Implementing content enrichment as an organisational capability ensures that your investment delivers value across the whole organisation in addition to delivering more value to customers.

To remain relevant, increase efficiency and deliver new revenue streams every publisher should be developing a plan for how they will utilise modern content processing techniques and tools. 67 Bricks specialise in helping you take a strategic approach to building your core content enrichment capability and using it to achieve strategic business objectives.

Clearing up the jargon

Content enrichment comes with a number of terms that need some clarification. These are our definitions of the relevant terms and how they relate to content enrichment.

AI - Software or machines that are designed to simulate intelligent behaviour.

Content collection - A set of content items gathered around a specific subject area or topic. Typically used to aid discovery by helping content consumers locate content around the subject they are interested in.

Content enrichment - The process of adding structure, context or metadata to content to make it more useful to humans and computers.

Domain - A specific area of knowledge or research e.g. 17th century playwrights or Stem cell therapies.

Entity - A term or phrase that represents a concept with an identifier from a known knowledge model e.g. a drug name, a place, an economic approach.

How content enrichment delivers business benefits for publishers

Content enrichment is one of the key enablers in helping publishers achieve strategic business goals. It provides the necessary tools to help you succeed in the modern digital content environment.

The role publishers play and the value that they provide is being challenged. Remaining relevant to your authors and readers whilst increasing efficiency, developing existing and new products and adding value are challenges that need to be faced. Many publishers are using outdated processes and systems to tackle these challenges. Publishers need to complete the transition to become digital, data driven organisations. Utilising augmented and automated content processing approaches is fundamental to achieving this and in doing so supports the delivery of board-level strategic objectives.

Retain position in the marketplace

Content enrichment helps publishers remain relevant by enabling them to provide new services and experiences to content creators and content consumers. Publishers need to move beyond the current value proposition to find new ways to add value as existing services become commoditised or disrupted. Building content enrichment capabilities that leverage the breadth and depth of your content supports the development of new services and provides a barrier to new entrants to the market.

Increase efficiency

Content enrichment helps increase content production efficiency by giving publishers tools to augment or automate content production tasks. It also gives them a better understanding about what content they have and where it is, it helps break down content silos, makes it easier to effectively utilise outsourcing organisations and supports more efficient QA processes.

Drive discovery and usage

Making content more discoverable and usable is achieved by developing a better understanding of your content and building services that utilise that knowledge. For example, creating semantic fingerprints of all your content items enables you to compare them and propose highly related content to a reader. Once you have semantic fingerprints you can implement additional discovery tools like finely targeted content notifications, personalisation, faceted search, etc.

Develop powerful information delivery systems

Modern information products need to meet the expectations of users who regularly interact with applications from the likes of Amazon, Uber, Google and Apple. Content enrichment plays a key role in helping publishers meet and exceed those expectations. For example, providing highly relevant recommendations to a user only works when you have a detailed understanding of your content items.

Deliver new revenue streams

Managing, storing and enriching content in granular chunks facilitates the reuse of content allowing it to be repackaged and delivered as new products or in content collections. This can deliver new revenue streams for front list content, support the monetization of legacy content and lengthen the timescale over which legacy content can be monetized.



Clearing up the jargon

Graph - A collection of concepts (nodes) connected by defined relationships (edges). Typically a graph allows things like semantic queries. For example, a graph structure for a publisher might contain books, articles, authors, entities, institutions, subject areas, etc. and would support interesting queries of this data.

Keyword - A term or phrase that has been identified as important or relevant in defining a piece of content. Keywords can be identified using automated statistical or grammatical approaches or by author, editor or reader manual approaches.

Knowledge model - A controlled vocabulary, taxonomy, ontology or other data structure containing concepts that is used to organise a specific domain of knowledge.

Linked data - A set of best practices for publishing and connecting structured data on the web.

Machine learning - A form of artificial intelligence (AI) that enables computers to learn without being explicitly programmed.

Natural language processing (NLP) - The ability of a computer program to understand human language inputs. NLP is a component of artificial intelligence (AI).

Ontology - An ontology is a machine-readable knowledge model for a specific subject domain that aims to describe both the entities and the relationships between those entities in that subject domain.

Semantic enrichment - The process of adding meaning to content. For example, adding metadata to a drug name indicating the drug group that it comes from or identifying a person's name as an economist from the 18th century.

An introduction to content enrichment approaches

Content enrichment is the process of adding structure, context or metadata to content to make it more useful to humans and computers. It can be a manual or automated process or a combination of manual and automated (augmented) processes.

A content enrichment capability typically incorporates technologies, workflow and people.

67 Bricks helps publishers select the right combination of manual and automated processes, technologies, workflows and tasks to deliver an organisation's strategic objectives.

Statistical analysis identifies important and relevant words, phrases, sentences or paragraphs and other content features. Having identified relevant content features these can form part of a semantic fingerprint or be used to generate new metadata for a content item. Different approaches are used in different scenarios depending on the business objective. For example, a 'bag of words' approach is very fast and can therefore be used to analyse large quantities of legacy content, whilst a corpus based approach might deliver a better result in a specific well defined content domain.

Grammatical analysis looks at the structure of sentences and uses this information to identify important words or phrases and relationships. This approach is typically used in conjunction with a statistical approach to apply additional weighting to terms.

Rules based approaches can be used to identify concepts where they follow a regular pattern or structure, but a database of all known examples of that concept would be impractical e.g. car number plates or postcodes.

Machine learning is a form of artificial intelligence (AI) that enables computers to learn without being explicitly programmed. The process of learning typically involves introducing new data to the machine learning application from which it can improve its processing algorithms. Machine learning has many applications within publishing. One of the most tested applications is automated classification. In this case we can feed the application with training data about how content items should be categorised. The application learns from this and applies categories to new content items. Implementing

a feedback loop means that the application can continue to improve over time.

Entity identification is the process of analysing content to find known concepts from within a knowledge model. This approach typically utilises a knowledge model e.g. taxonomy or ontology that provides information to the system about the concepts we are looking for. Being able to confirm that a phrase is an entity from a knowledge model supports additional product features. For example, knowing that a phrase is a known drug we can use its identifier to pull in data from other sources and present more value to the content consumer.

Ontological classification is the process of using a structured ontology of terms to classify a piece of content. The ontology typically needs to be designed for this specific purpose. A software application is then used to build classification rules based on the ontology that can be applied to content items. This is a good approach in domains where specific terms identify the category a piece of content falls within. For example, if the content mentions the Supreme Court then we might be able to assume that this content item is referencing the US legal system.

Sentiment analysis is the process of analysing text to determine the opinion of the author. Typically, this is used to determine whether a piece of text is positive, negative or neutral towards a particular subject.

Relationship identification is the process of identifying connections between specific concepts within the content. This typically requires very specific rules for specific content domains and is hard to achieve good quality, but can deliver significant value when successful. For example, it might be extremely useful to know when a piece of content identifies a specific interaction between two drugs.

Clearing up the jargon

Semantic fingerprint - A machine readable collection of terms or phrases that uniquely identify a piece of content. Semantic fingerprints support features like relatedness and personalisation.

Text analytics - The process of analysing natural language text to derive information about a content item, such as identifying and extracting entities within it.

Why content enrichment should be implemented as a strategic organisational capability

So far, most publishers have only dabbled in using modern content processing tools or they have implemented spot solutions for specific needs that do not deliver value across the organisation.

If content enrichment is implemented as an organisational capability it delivers multiple benefits across the publishing lifecycle. Taking a strategic approach will maximise the return on any investments made.

We propose that organisations should focus on developing capabilities and adding value to their content assets ahead of publication rather than solely embedding features in delivery applications that can be lost when you move platform and cannot be reused or using black box approaches that do not add value to your content assets. We suggest that neither of these alternative approaches move the publisher towards being a modern digital content organisation that can meet the changing needs of content creators and consumers.

67 Bricks recommend taking a capability mapping approach to ensure that the implemented content enrichment capabilities can sustainably deliver the organisation's strategic business objectives. We use an approach that maps business objectives to user goals, user goals to content features and content features to capabilities. This ensures that the most appropriate content enrichment capabilities are implemented and the right technology choices are made.

Having built a capability map, a roadmap can then be drafted that implements the most important capabilities first and delivers early value to the business. R&D and prototyping activities should be incorporated into the roadmap to test ideas internally and externally, build awareness and buy-in across the organisation and support fine-tuning of the roadmap as the programme progresses.

Building capabilities rather than spot solutions or embedded delivery features helps organisations transition to being digital and data driven organisations and achieve board level strategic objectives.

How 67 Bricks works with publishers to help them develop strategic content enrichment capabilities

67 Bricks is a software development consultancy that focuses on designing and implementing content enrichment capabilities for publishers. We provide consultancy to help you define a content enrichment strategy and software development services to build prototypes and production ready content enrichment capabilities.

Our strategic consultancy can help you develop a content enrichment strategy incorporating a vision for how it will impact your business, a capability map that connects capabilities to business drivers and a roadmap that communicates how and when business benefits will be delivered.

Our software development teams build content enrichment capabilities using the most appropriate approaches and technologies to achieve your business objectives. We provide combined teams including technologists and publishing experts.



We work independently or alongside existing client technology teams and we transfer knowledge and skills to internal teams as required to ensure the successful ongoing use of the systems developed.

We are the recognised experts in designing and implementing content enrichment capabilities for publishers that deliver multiple strategic objectives. Get in touch to find out how we can help you.

Sam Herbert

CLIENT SERVICES DIRECTOR

sam.herbert@67bricks.com

www.67bricks.com