



WE HELP YOU CREATE SMART  
CONTENT THAT SAVES YOU COSTS  
AND DRIVES MORE REVENUE

# Dutch publisher leverages smart content with artificial intelligence

## Challenge

Our customer wanted to turn their content into Smart Content in a time- and cost-effective way. Smart Content is content that is context-free and well-described by metadata. This means that content is easily searchable and accessible in any digital system. Our customer had a specific classification system for content already established. The classification system was organized into a two-dimensional matrix based on descriptions of the educational purpose of the content.

Our customer was struggling with using and updating metadata<sup>1</sup> manually. Metadata types were difficult to assign objectively and accurately. The process of decision-making, combined with the manual metadata tagging of content, takes time and resources and requires experts to make appropriate choices. The customer asked EDIA to run an experiment with EDIA's metadata-tagger in order to discover a faster and more accurate way of metadata tagging.

## The process

EDIA uses artificial intelligence to find, classify and label content with metadata for educational publishers. EDIA's automated metadata-tagger is trained using a sample set of the customer's data. Based on what the algorithm learns, it can apply these rules to a wider set of content. This means EDIA can automatically label content with metadata both accurately and reliably.

In order to train the Metadata-Tagger properly, it must be provided with a sample set of data. We use a combination of different types of machine learning to ensure that the algorithm learns quickly and accurately from the data it has been given. Supervised and unsupervised approaches were combined to allow an algorithm to seek out new data, then learn from it. In the case of educational content, unsupervised learning grouped similar texts while supervised learning assigned these categories more specific metatags.

### **Machine learning**

*is the use of algorithms which allow a computer to perform a task without specific instructions. Machine learning algorithms learn based on vast amounts of data, from which it observes patterns and generates inferences, removing the need for step-by-step human instruction.*

### **Smart Content**

*is digital content which is clearly structured and well-labelled using metadata. This makes content more easily searchable, clearly organized and ready for digital learning platforms.*

### **Supervised learning**

*is a type of machine learning which involves training an algorithm to perform a specific task, using a set of training data. Supervised machine learning is used for categorizing content, or in the case of educational content, assigning it digital meta-tags.*

### **Unsupervised learning**

*harnesses an algorithm to search through vast quantities of data to look for patterns and structures. Unsupervised learning algorithms are often used for cluster analysis or to explore large quantities of data. In the case of educational content, unsupervised learning can be an initial first stage in working out how much content is available and how it should be organized.*

1. Metadata is data describing other data. It can be but is not limited to subject matter, topic, main keywords, reading level, etc.

## The goal

EDIA aimed to prove that automated metadata tagging could be done in an efficient manner, saving time and money spent on labelling large amounts of content. This would also demonstrate that automated metadata tagging was both accurate and reliable when applied to complex structures.

## Results

The results showed a high level of accuracy for most of the content – between **81%** and **95%**. This high level of accuracy was achieved with metadata types as well as using a combination of both supervised and unsupervised machine learning techniques. This combination meant that any data which was not fully represented in the training data did not produce false-positives. Metadata types which were moderately or minimally represented in the data training set required special attention of our AI experts.

| LABELS AI ▾  | Introduction | Theory | Processing | Reflection | Evaluation | TOTAL (Labels) | CORRECT |
|--------------|--------------|--------|------------|------------|------------|----------------|---------|
| Introduction | 1            | 0      | 31         | 0          | 0          | 32             | 3,13%   |
| Theory       | 0            | 109    | 0          | 0          | 0          | 109            | 100,00% |
| Processing   | 0            | 0      | 748        | 0          | 16         | 764            | 97,91%  |
| Reflection   | 0            | 0      | 5          | 0          | 0          | 5              | 0,00%   |
| Evaluation   | 0            | 0      | 13         | 0          | 77         | 90             | 85,56%  |

Visualisation of metadata tag allocations: 93,5% accuracy on average by AI

Any confusion for the algorithm in classification indicates little distinction between examples. It was confirmed that this confusion would have affected human labellers in the same way the algorithm was affected. If content was to be tagged manually rather than automatically, similar confusions would have emerged. This confusion indicates where problems in classification are - a confusion that is common to both human and AI metadata tagging processes.

## Conclusion

While helping the publisher explore the possibilities of Smart Content development, EDIA demonstrated that automated metadata-tagging could be performed quickly and with high accuracy while using techniques of machine learning.

**As an additional benefit accurate metadata tagging and classification allows publishers strategic insight into their content libraries. Making data-driven decisions becomes easier, therefore improving overall efficiency. Having Smart Content also enables publishers to explore new business models and personalized digital products, allowing them to move confidently into the future.**



Tesselschadestraat 11  
1015RA AMSTERDAM, The Netherlands  
+31 – (0)20 – 716 36 12

[www.edia.nl](http://www.edia.nl)  
[info@edia.nl](mailto:info@edia.nl)