

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

Major educational publisher reaches 95% metadata accuracy with artificial intelligence

Malmberg's Challenge

Malmberg provides educational materials for many levels of learning, from primary education up to higher vocational education in the Netherlands. Malmberg, therefore, has a large amount of content covering a variety of subjects and age ranges. This content is only helpful when organized, which in the digital age means described by metadata. Metadata tagging involves giving different content digital tags according to criteria like subject, keywords, language level and intended age. This helps to classify information clearly, making it easy to navigate, store, search and reuse.

Traditionally metadata tagging has been done manually, by employees typing metadata into content. This way of operating is slow, expensive and prone to errors. Much of Malmberg's metadata had spelling errors, abbreviations or was placed in the wrong categories. Some content was missing metadata altogether, meaning searches for content were ineffective or limited.

To add the missing or replace the erroneous metadata would be time-consuming enough, but Malmberg also wanted to have the ability to implement a new metadata standard if necessary. This would mean manually tagging the content all over again, using more employee time and company money.

Malmberg wished to have their content classified according to Bloom's taxonomy model¹, a common framework for classifying educational content. This taxonomy classifies content according to learning objectives, making it helpful and relevant for educational publishers and teachers. Malmberg was also looking to classify its content according to CEFR² standards which are helpful for determining the language difficulty level of a piece of content.

Instead of fighting a losing battle against metadata, Malmberg decided to try something new. They asked EDIA to run an experiment with their automated metadata tagging model.

MALMBERG

Malmberg is a leading Dutch Education Publisher established in 1885.

The publisher provides educational materials for the Dutch education system including primary education, secondary education, and higher vocational education.



The EDIA process

EDIA uses machine learning to find and classify educational content in a fraction of the time it would take a human. The EDIA model is 'trained' on a sample set of content, allowing the algorithm to learn how to tag content efficiently and accurately. Once the model is operating correctly, it can be applied to larger amounts of data. Content can be quickly arranged into an organized and searchable system, including information on things like subject and language level.

In the Malmberg case, EDIA aimed to prove that our classifier could tag content according to Bloom's taxonomy as well as a human expert could.

The aim was to achieve 60% accuracy in metadata tagging of History, Chemistry and English.

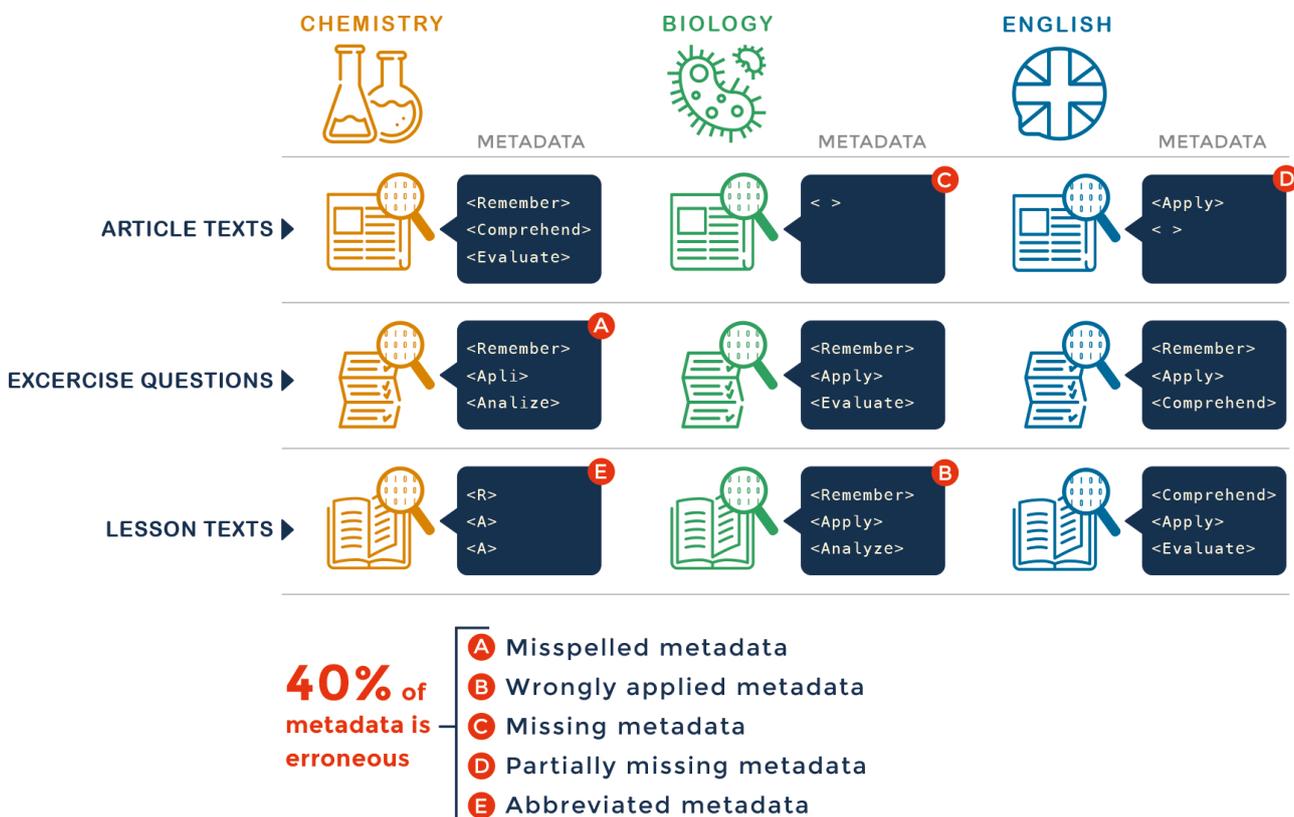
EDIA also aimed to prove that we could tag content according to CEFR levels with 60% accuracy. This was targeted at the subject of English as a second language. The accuracy of metadata tagging was measured against the metadata tagging efforts of human editors.

Results for Malmberg

After 4 weeks of operation and analysing 33,000 content items, EDIA had exceeded the goal of 60% accuracy, instead achieving:

- 79% accuracy** for Chemistry according to Bloom's taxonomy
- 68% accuracy** for History according to Bloom's taxonomy
- 95% accuracy** for English according to Bloom's taxonomy
- 94% accuracy** for English in CEFR classification

Analysis also revealed that up to 40% of Malmberg's existing metadata was erroneous.



These results had multiple benefits for Malmberg. Using the EDIA model, content was classified according to both Bloom's taxonomy and CEFR, improving storage, searchability and reuse of educational content. Malmberg alleviated manual processes of meta-tagging which are expensive and prone to error.

The Malmberg case demonstrates that automating metadata tagging and classifying content difficulty at scale is not only possible but provides high-quality results. EDIA allows educational publishers to create adaptive content relevant for the digital environment at a lower cost, while reducing the administrative burden on their employees.

The EDIA logo consists of the letters 'EDIA' in a white, outlined, sans-serif font, set against a solid blue square background.

EDIA develops artificial intelligence suites designed for educational publishers. The technology automates metadata tagging processes, which helps drive revenue, lower costs and manage educational content more efficiently.

1. **Bloom's Taxonomy**

Bloom's Taxonomy is a framework which divides learning into six models of activity. It is broadly used by teachers, publishers and other educational professionals to direct, plan and classify educational content against learning objectives. The taxonomy is widely used for metadata tagging.

2. **CEFR**

The Common European Framework Reference for Languages is a framework for language proficiency which can be applied to all types of content. Levels such as between A1 and C2 are applied to content as a measure of their language difficulty level.

The EDIA logo consists of the letters 'EDIA' in a white, outlined, sans-serif font, set against a solid blue square background.

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

www.edia.nl
info@edia.nl