

Master thesis : Automatic Quality Assessment of Digital Comics Using Machine Learning

Auteur : Haas, Bastien

Promoteur(s) : Geurts, Pierre

Faculté : Faculté des Sciences appliquées

Diplôme : Master : ingénieur civil en science des données, à finalité spécialisée

Année académique : 2025-2026

URI/URL : <http://hdl.handle.net/2268.2/25226>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Automatic Quality Assessment of Digital Comics Using Machine Learning

Author : HAAS Bastien

Supervisors : GEURTS Pierre, ROEKENS Joachim

2025-2026

Thesis presented to obtain the degree of :
Master of Science in Data Science Engineering

The objective of this master's thesis is to develop an algorithm capable of assessing the quality of digital comics using machine learning. This was conducted to enable *Stripik*, a mobile application that allows users to write comics with a generative tool and also read the content created, to recommend works that meet a certain quality threshold. The learning process must be based on comics that have been manually assessed beforehand.

The development of such an algorithm first involves selecting and managing the data to be used during the learning phase. Afterwards, the data were divided into three main categories : statistical data, textual data and graphical data. For each category, data were specifically processed, according to the corresponding category. A predictive model then had to be chosen to determine which one would be implemented in the final algorithm. This choice was made sequentially, beginning with the optimization of each model individually on a training set, followed by a comparison of the optimized models using a validation set. The resulting model, trained on the selected and processed data, ultimately constitutes the algorithm to be developed in this work.

This thesis concludes with encouraging results, although they are not yet fully satisfactory in terms of performance. Future perspectives for improving the assessment algorithm and its outcomes are proposed.