# Integration of Unicheck plagiarism checker into DSPACE platform

## Introduction

Unicheck is a similarity detection engine used by educators and students to improve writing skills and prevent plagiarism (Unicheck scans submitted files /text for similarities with Internet and/or file collections). Unicheck is a cloud service that works as a standalone application or can be integrated into learning management system (LMS) or other platforms via standard integration technologies (LTI/API/plugin).

We would like to develop Unicheck extension for DSpace platform to allow repository administrators (or other relevant roles) to run on-demand plagiarism scans. The aim is to provide efficient and easy-to-use tools to help maintain high quality content in the repository. Automatic similarity check upon file deposit can serve as another layer of quality assurance. Prior to committing to this project, Unicheck has approached DSpace Community Advisory Team (DCAT). The committee suggested to gather feedback from community to understand what features and in what scenarios can be useful. Please provide feedback how you see the usefulness of suggested features and who might benefit from them and in what user cases.

## Suggested functionality

### Available to DSpace users at no cost (repository scan)

User can scan a file or multiple files for similarities across repository (most popular file formats are supported including .doc, .docx, .rtf, .txt, .odt, .html, .pdf). If there is significant text overlap (similarity score is high), it can be indicative of plagiarism, duplication of authors' content without proper citation, or other academic integrity misconduct. Automatic similarity scan is an additional quality control measure that helps to ensure that repository content is original (free from plagiarism) and of high quality (properly cited).

### Available to DSpace users for a fee (repository + Internet scan)

User can scan a file or multiple files for similarities across repository and the Internet. If there is significant text overlap (similarity score is high), it can be indicative of plagiarism, duplication of authors' content without proper citation, or other academic integrity misconduct. Automatic similarity scan is an additional quality control measure that helps to ensure that repository content is original (free from plagiarism) and of high quality (properly cited).

The difference between free and paid options is that paid option includes Internet scan - i.e. billions of pages and documents, published on the web. Since Unicheck is charged by web index provider (Microsoft Bing) for using fresh index, this scan option is offered for a fee.

## Unicheck report (examples)

Similarity matches are highlighted with yellow. Citations are highlighted with blue. Information section gives basic statistics about originality and has list of sources. User can click text matches in report and see relevant sources and vice versa.

Screenshot1. User can open the source and see the text match (color mask) in the source:

blocked URL

Screenshot2. User can exclude citations and references from the document, browse citations:

blocked URL

## Use Case

Automatic similarity check upon depositing research output - link.

## Feedback

Please, let us know what you think, what features you would like to see and what user cases come to your mind.