Map Text Extraction and Parsing using Optical Character Recognition (OCR) for Facilitating Map Reproducibility Assessment

Thesis submitted in partial fulfilment of the requirements for the Degree of Master of Science in Geospatial Technologies

Yohannes Abrha Mulaw

Supervised by:

Dr. Christian Kray
Institute of Geoinformatics
University of Muenster
Muenster, Germany

Co-supervised by:

Dr. Jose Fransisco Ramos
Institute of New Imaging Technologies
University of Jaume I
Castellon, Spain

MSc. Eftychia Koukouraki
Institute of Geoinformatics
University of Muenster
Muenster, Germany

28 February 2024

Declaration of Academic Integrity

I, Yohannes Abrha Mulaw, hereby declare that the thesis titled "**Map Text Extraction and Parsing using Optical Character Recognition (OCR) for Facilitating Map Reproducibility Assessment** " is my original work, conducted and written solely by me with the guidance of my supervisors. I have faithfully and accurately cited all my sources, including books, journals, handouts, and unpublished manuscripts, as well as other internet resources. This work has not been approved for any other degree and is not currently being submitted for any other qualification.

____

Yohannes Abrha Mulaw

Münster, Germany

28 February 2024

Acknowledgements

I would like to express my deepest gratitude to almighty God, the source of all wisdom and strength, for guiding me throughout this journey.

I am immensely thankful to my supervisors, Prof. Christian Kray, Prof. Jose Fransisco Ramos, and MSc. Eftychia Koukouraki for their unwavering support, invaluable guidance, and mentorship. Their expertise, encouragement, and constructive feedback have been instrumental in shaping the path of my research.

I extend my heartfelt gratitude to the Erasmus Mundus Program for awarding me the scholarship that enabled me to pursue my MSc degree at leading universities. This opportunity not only enriched my academic experience but also allowed me to interact with students from diverse cultural backgrounds.

I would like to express my gratitude to my family back home, friends, and classmates for their consistent support and understanding during my academic journey.

Finally, I appreciate the contributions of all who participated in the survey, collaborated, and shared their expertise, enhancing the overall quality of this research.

Abstract

Reproducibility stands as a fundamental element in promoting transparency and openness in scientific publications and in geoscientific research as well. Figures, particularly maps, integrated into geoscientific research play a significant role in visualizing and representing crucial scientific results; thus, they should be reproducible. However, the assessment of map reproducibility for determining the success of map reproduction is limited due to the absence of standard metrics, criteria, and tools. In this study, a novel web-based application is developed to facilitate the map reproducibility assessment process based on textual elements of the map. The tool integrates an open source optical character recognition (OCR) technology for text extraction from maps and proposes a comprehensive comparative analysis workflow consisting of assessment criteria such as the text similarity between the extracted texts using fuzzy string matching techniques, the overlap ratio between the bounding boxes associated with the texts using the Jaccard index (intersection over union), and the Euclidean distance between the bounding boxes for effective map reproducibility assessment. The tool is validated and evaluated using real-world datasets and reveals its effectiveness compared to the existing map comparison methods in terms of accessibility, interoperability, and flexibility to accommodate diverse file sizes, image resolutions, and file types. As a result, the tool was found to be usable with a SUS score of 69.33 and useful for researchers and GIS professionals to extract and assess textual elements from maps. In addition, the study demonstrates promising results in the effective utilization of OCR technology for accurate text extraction from maps, even with the lowest map image resolution (60 dpi) and smallest font sizes (7 pt).

Keywords: reproducibility, map reproducibility, map reproducibility assessment, optical character recognition (OCR), text analysis, fuzzy matching

Table of Contents

List of Figures

List of Tables

List of Abbreviations

OCR	Optical Character Recognition
API	Application Programming Interface
GIS	Geographic Information System
ERC	Executable Research Compendium
OTD	Overlap Text Detection
PNG	Portable Network Graphic
BMP	Bitmap Image File
JPEG	Joint Photographic Experts Group
WebP	Web Picture Format
REST	Representational State Transfer
DPI	Dots per Inch
HTTP	Hypertext Transfer Protocol
HTML5	Hypertext Markup Language 5
CSS	Cascading Style Sheet
OSM	Open Street Map
CPU	Central Processing Unit
GPU	Graphics Processing Unit
SUS	System Usability Scale
IoU	Intersection over Union

1 Introduction

1.1 Research Background and Context

Scientific assertions or findings must be supported with evidence (Giraud & Lambert, 2017). In this context, the concept of reproducibility has become the key component for purposes of scientific validation and transparency, which entails the ability to re-run and re-create scientific experiments implemented by other researchers, thereby encouraging openness and reliability in the scientific validation process (Konkol et al. 2019).

Reproducibility and reproducible research have recently gained focus and emphasis in geoscientific publications (Koukouraki & Kray, 2023). Like other fields, adopting the practice of reproducibility in geoscientific publications improves transparency and validity among the scientific community (Nüst et al., 2017). However, geoscientific research involves a huge volume of datasets and requires high computational costs (Kedron et al., 2021), which makes it challenging to adopt the practice of reproducibility. However, efforts have been made to address this issue by assessing the computational reproducibility of geoscientific research through re-running the analysis of the research using the original code embedded in the paper (Konkol et al., 2019). Adopting a practice like executable research compendiums (ERC) involves encapsulating user interface configurations, code, data, and methods within a single package with the objective of enhancing the reusability, reproducibility, and transparency of computational research (Nüst et al., 2017). Furthermore, Konkol & Kray (2019) conducted detailed examinations of reproducing figures in geoscientific publications to promote reproducible research. Additionally, Koukouraki & Kray (2023) discussed efforts to ensure map reproducibility and the assessment process, contributing significantly to advancing the efforts towards achieving reproducibility in geoscientific results.

In geoscientific publications, figures such as maps are often used to communicate and visualize critical geographic scientific results, and being capable of reproducing them utilizing the code, method, and datasets is therefore also an important aspect of reproducible research (Konkol & Kray, 2019). Recognizing this importance, Koukouraki & Kray (2023) investigated the reproducibility of maps and their rate of reproducibility through assessing the difference between maps visually without using additional assessment tools. This approach is always prone to error and demands time and effort. In this case, designing and developing a tool for facilitating this assessment of map reproducibility contributes to the overall map reproducibility decision.

1.2 Motivation and Problem Statement

Giraud & Lambert (2017) state that figures in geoscientific publications, particularly maps, often showcase crucial results in geoscientific publications and thus should be reproducible. In an effort to reproduce maps, an assessment, comparison, and evaluation between the original map retrieved from the original paper and the reproduced map re-created using the codes and methods from the original paper are mandatory (Konkol et al., 2019). However, due to the absence of standardized assessment metrics, tools, and criteria, the map reproduction assessment process has become challenging. In an effort to assess the success of map reproduction, Koukouraki & Kray (2023) used an image-based(pixel-by-pixel) comparison tool called compare from ImageMagick[1] . However, this tool exhibits limitations in handling diverse file types, sizes, and image resolutions. Consequently, significant variations arise in the reproducibility assessment results, even when the map images appear visually and semantically similar. This discrepancy occurs due to the inherent differences in the embedded source code and data within the original paper, which often result in variations in file type, size, and resolution. To address this gap, there is a need for developing a map reproducibility assessment tool that is efficient, interoperable, and independent of file types, sizes, and image resolutions between the original and reproduced maps, relying on their textual features.

Comparing geographical maps visually involves the assessment of their components, such as spatial accuracy, map legend interpretation, and understanding the textual elements (Bill et al., 2022). Nevertheless, visual comparison is often susceptible to errors and subjective outcomes, demanding significant time and effort. The challenge intensifies when dealing with dense textual features and a mix of heterogeneous fonts, making it difficult to visually recognize and interpret text labels. Thus, there is need for advanced tools that utilize optical character recognition (OCR) technology, which is vital for extracting these textual elements.

1.3 Objectives and Research Questions

The primary goal of this thesis is to develop and implement a methodological framework and approaches aimed at facilitating the assessment of map reproducibility, with a specific focus on textual elements within the map. This approach is crucial because textual features on map elements play a significant role in representing key features of geoscientific results and are integral components of maps. Additionally, texts integrated into maps aid in understanding and interpreting aspects that cannot be conveyed or replaced by other visual elements, such as colour or intensity. The study aims to address the following research questions:

1) To what degree do web-based text extraction, parsing, and comparison of maps facilitate the reproducibility assessment of maps in geoscience publications?
2) To what extent can thes latest OCR technology accurately extract text elements from map images?

To address the research questions, the following objectives need to be implemented:

- Designing and implementing a web-based application that integrates OCR technology for text extraction, parsing, and comparison of geographic maps to facilitate the assessment of map reproducibility.

[1] https://imagemagick.org/

- Designing a comprehensive conceptual framework for facilitating the assessment of map reproducibility, incorporating text analysis, comparison, and bounding box comparison.
- Evaluating the usability and usefulness of the web-based application.

1.4 Significance of the research

The study implements a novel tool for the simplified comparison of original and reproduced maps based on their textual features, thereby facilitating the overall map reproduction assessment process. This tool serves a dual purpose: firstly, it investigates the applicability of optical character recognition (OCR) technology on maps, shedding light on the challenges and potential limitations associated with extracting textual elements from maps. Secondly, the tool is designed with consideration towards integration into web-based applications, aligning itself with improving accessibility, interoperability, and enhancing the visualization of extracted texts and annotated maps to facilitate the map reproducibility assessment process smoothly. Additionally, by addressing challenges related to textual elements in map reproducibility assessments, the tool offers valuable insights for cartographers and mapmakers, serving as a creative resource that informs and empowers cartographic workflows and practices.

Furthermore, the research aligns with the principles of open reproducible research in geoscience, contributing to the transparency and reliability of geospatial research. The open-source nature of the tool not only promotes collaboration within the scientific community but also promotes a collective effort to enhance the accuracy of geographical information.

1.5 Thesis outline

The thesis is organized into seven chapters. Chapter 1 provides an overview of the research, including its rationale and problem statement. Chapter 2 focuses on reviewing the literature, encompassing existing research and theoretical frameworks. Chapter 3 elaborates on suggested conceptual ideas and solutions. The implementation process of these methods and solutions is detailed in Chapter 4. Chapter 5 presents and discusses the overall system evaluation and results. Chapter 6 contains the discussion, while Chapter 7 concludes the thesis.

Remark: In this thesis, the term "original map image" refers to the map retrieved directly from the original publication paper, whereas the term "reproduced map image" refers to a map that has been recreated using computer code, data, and methods from the original paper.

2 Related Work

The related work discussed in this paper has mainly concentrated on four pillars: reproducible research in geoscience, map production and reproduction, map text extraction using optical character recognition (OCR), and the applicability of fuzzy string-matching techniques to geographical texts.

2.1 Reproducible Research

2.1.1 Reproducibility vs Replicability

The idea of reproducibility in research publications was first introduced by Jon Claerbout, a geoscientist at Stanford University (Claerbout & Karrenbach 1992). After that, several studies (Donoho 2010; Peng 2011), among them, have investigated the reproducibility of scientific publications. Peng (2011) emphasizes the importance of reproducibility as a minimum standard for evaluating scientific claims. In literature, there are ongoing debates and confusions about the definition and usage of the terminology's "reproducibility" and "replicability." In certain instances, the term reproducibility is treated as the repeatability, robustness, reliability, or generalizability of scientific findings (Editorial, 2016). To solve this problem, Barba (2018) introduced the Claerbout/Donoho/Peng convention following the (Claerbout & Karrenbach 1992; Donoho 2010; Peng 2011) idea of reproducibility and replication to define distinctly reproducible research as ''*Authors provide all the necessary data and the computer codes to run the analysis again, re-creating the results*", whereas replication is collecting new data (sometimes using different methodologies) and conducting new analysis, yet arriving at the same scientific conclusions as another work.

Conversely, the Association for Computing Machinery (ACM) adopted the different definitions of reproducibility and replicability in 2016 and defined reproducible research as measurement that can be obtained with stated precision by a different team using different experimental setups or measuring systems, whereas replicability of research means measurement can be obtained with stated precision by a different team using the same experimental setup (Barba, 2018). Moreover, Feger & Woźniak (2022) suggest that the definition of reproducibility becomes better and more accurate when it is defined based on a research-centric approach, as this approach prioritizes the needs and perspectives of researchers, recognizing their central role in the scientific process.

2.1.2 Reproducible Research in Geoscience

Recently, numerous geoscientific publications have been assessed and reviewed at the AGILE conference[2] to experiment with reproducibility. The principle of adopting research reproducibility practices in geoscience publications is key to promoting transparency,

[2] The Association of Geographic Information Laboratories in Europe Conference proceedings(https://agile-gi.eu/)

openness, and robustness in the field (Nüst & Pebesma, 2021). However, there are challenges and limitations when it comes to making geoscience research reproducible. Pebesma et al. (2012) outline the challenges for the limitation of actual reproducibility of geospatial research, noting that geospatial research involves a large volume of datasets, data conversions, and the utilization of geospatial and non-geospatial software tools for data processing and result visualization. Furthermore, Nüst & Pebesma (2021) uncovered that the diverse nature of the field and the diverse academic backgrounds of researchers within it pose limitations to the practice of reproducibility. Additionally, Kedron et al. (2021) argue that geoscientific publications, being computationally and data-intensive, present significant challenges in achieving reproducibility. Moreover, Nüst et al. (2018) found that motivation and understanding towards the practice of reproducibility are limited within the scientific community, which challenges the adoption of reproducibility in the field. Similarly, Konkol et al. (2019) argue that the lack of understanding of reproducibility concepts among the scientific community limits the adoption of reproducibility practices.

In addressing the challenges that limit the practice of reproducibility, Nüst & Pebesma (2021) and Nüst et al. (2017) develop innovative approaches to enhance computational reproducibility by encouraging literate programming, embedding code and reports simultaneously, and leveraging containerization of applications through Docker[3]. In addition, the Executable Research Compendium (ERC) platform proposed by Nüst et al. (2017) encapsulates the entire research environment, facilitating seamless execution without concerns related to operating system type and software versions. Moreover, the use of open-source tools and programming languages like R[4] in geospatial analysis is better for reproducing geoscientific research results because they promote accessibility, transparency, collaboration, and flexibility (Pebesma et al., 2012).

2.2 Map Production and Reproduction in Geosciences

2.2.1 Map Production

According to the International Cartography Association (ICA), cartography is the science, art, and technology of making and using maps, while a map is an abstract visual representation of the geo-governments (Menno-Jan Kraak, 2019). In Geographic Information Systems (GIS), Taylor (1991) characterizes cartography as the process of arranging, presenting, communicating, and making use of geo-information in visual, digital, or tactile formats. This encompasses all phases, starting from data preparation and extending to the ultimate incorporation into the development of maps and associated spatial information products. Mocnik (2023) describes maps as a medium of communication like books, magazines, CDs, and actual painting. This highlights the importance of maps not only as informative tools but

[3] https://www.docker.com/

[4] https://www.r-project.org/about.html

also as effective means of conveying complex spatial information, demonstrating their unique role in facilitating, understanding, and interpreting complex geographic data.

Recently, the digital revolution, coupled with geographic information systems (GIS) has revolutionized the map-making process. The process has been popularized due to open-source datasets and geographic information system (GIS) mapping software, which has enabled a wider audience to participate in cartography (Kraak & Fabrikant, 2017).

Jobst & Gartner (2019) discussed the significant impact of geospatial artificial intelligence (GeoAI), the application of artificial intelligence fused with geospatial data on map production, and they emphasized the importance of integrating artificial intelligence in extracting information, connecting knowledge, and enhancing the capabilities of traditional cartography. In terms of reproducibility, understanding this evolution is crucial to handling the complexities of reproducing maps in the contemporary scientific environment. Making maps and understanding them as well is a complex process that requires the processing of various steps (Mocnik, 2023). So, while considering the reproducibility of maps, the method of map production has its own impact.

2.2.2 Map Reproducibility and Assessment Process

In the literature, there's a scarcity of publications on map reproducibility and assessment methodologies, despite growing concerns about reproducibility in science. Konkol & Kray (2019) highlighted the importance of map reproducibility as a core aspect of reproducible research, emphasizing the re-creation of figures such as maps through the re-computation of codes and methods from the original paper.

Konkol et al. (2019) examine the reproducibility of figures, including maps, in geographic research and attempt to reproduce results from 41 papers by revealing technical and non-technical issues and proposing custom guidelines to enhance the computational reproducibility of figures in geographic papers.

Giraud & Lambert (2017) introduced a spectrum to assess map reproducibility, ranging from not reproducible (if only map layout information is included) to fully reproducible (if associated code, data, and linked metadata are provided). They highlighted the challenge of achieving reproducibility due to human intervention in map-making and the lack of standard metrics.

Addressing this, Koukouraki & Kray (2023) proposed a custom guideline for map reproduction assessments, categorizing differences as aesthetic (visual expressions like color or font choice) and semantic (changes affecting the substance of the map, such as missing elements or factual alterations). According to their guidelines, successful reproduction is determined by the presence of aesthetic differences; otherwise, the reproduction is considered unsuccessful.

2.3 Map Comparison

For successful map reproduction assessment, the original map retrieved from the original paper and the reproduced map created using the code, data, and metadata from the original paper must be compared. In this case, Konkol et al. (2019) also suggested that for assessing the reproducibility of maps, one part of the task is that maps in the scientific publication must be recreated and compared to the original map that was published on the original paper.

Koukouraki & Kray (2023) employed a pixel-by-pixel map image comparison for map reproduction assessment; however, they found out that the approach was inadequate due to disparities in image resolution, file types, and file sizes. In this case, an automated comparison tool that is independent of the file size, type, and image resolution is necessary.

Map comparison is a common practice in most geoscientific publications (Foody, 2007), and it's also the most important topic in geoscience, such as cartography, spatial analysis, and geo-simulation modelling (Smith & Dragićević, 2022). Figures such as maps show socio-economic, time-series analysis of events, hotspot detection, and uncertainty analysis, and for areas that have an influence on environmental research, the need for comparison between these maps is essential (De Jong & Hagen-Zanker, 2004). Although map comparison may seem straightforward, its execution is very complicated, as highlighted by Dvorský et al. (2010).

De Jong & Hagen-Zanker (2004) discussed that visually comparing maps is usually prone to error, leads to subjective results, and demands time and effort. They overcame this obstacle by creating Map Comparison Kit (MCK), a piece of software that combines established methods of comparing maps cell-by-cell (such as the Kappa statistic and variants on it) with more modern methods of comparing maps using fuzzy sets (the Kappa fuzzy).

Hagen (2003) proposed a method for comparing categorical raster maps using fuzzy set theory. This approach considers both the fuzziness of location and the fuzziness category, allowing for a more complex assessment of similarity. The evaluation procedure is based on the output that has a value for each cell from 0 to 1 to determine the degree of similarity in each cell.

Loran et al. (2018) contribute a conceptual framework that facilitates the comparison of topographic maps across different time periods. They highlighted the need for a systematic approach to evaluating map comparability to avoid or minimize the detection of spurious landscape changes resulting from incompatible map series that arise due to context, distortion, and cartographic generalization across different time periods of maps. Annanias et al. (2023) address the need for a computational approach to efficiently compare old maps, which are valuable for historical research and urban development assessments. It introduces a method leveraging semantic segmentation to systematically analyse digitized old maps. They planned to create a process that could automate map comparisons, facilitating various applications. Future research aims to enhance automation and efficiency in comparisons.

To further facilitate the process of map comparison, the conversion of map textual elements into a machine-readable format or digitized version using optical character recognition (OCR) proves essential.

2.4 Map Text Extraction using Optical Character Recognition

Barget (2022) highlighted the importance of maps as they serve as the primary source of information, yet making them available for analysis poses a challenge. Optical character recognition (OCR) is a computer vision technique used for the conversion of texts into machine-readable format from handwritten documents, scanned or printed images, and natural scene images (Sabu & Das, 2018). OCR has been applied to a variety of areas, including scanned documents (Mittal & Garg, 2020), office automation, factory automation, benefiting map production efficiency, and labelling the points of interest of street view images (Du et al., 2020).

Chaudhuri Arindamand Mandaviya (2017) pointed out that employing OCR is still a challenging task in automatic cartography, the process of automating text extraction from maps. This challenge is a significant area of concern when assessing map reproducibility using the extracted texts from the map. Chiang & Knoblock (2015) noted that, even in the field of computer vision and image processing, maps are treated as a special type of input for OCR tools. Chiang (2017) mentioned that extracting text labels from maps is a challenging task, and these difficulties arise from the lack of fixed orientation in the textual elements coupled with the complex layout and irregular fonts employed on maps (Y. Y. Chiang & Knoblock, 2015).

To extract textual features from maps, Cao & Lim Tan (2002) present a map text extraction technique by integrating line continuation with line width to interpret the connection of text and graphics for the text detection process, and a commercial OCR engine is utilized for text recognition. Pouderoux et al. (2007) suggest an approach that helps to align text elements on a raster map by analysing the geometric properties of individual connected components for the text detection task, rotating the detected text string horizontally, and feeding it into the OCR engine for text recognition. Similarly, Chiang & Knoblock (2010) employ a recognition approach that involves locating individual text labels on a map, determining their orientation, and rotating them to a horizontal position. This enhances the performance since the OCR engine performs well on horizontally aligned texts. Chiang & Knoblock (2011) propose a methodology to detect multi-oriented and curved texts that don't have homogenous text by applying character grouping criteria based on the text sizes and maximum desired curvature of the text. These make the process of recognizing texts from maps using OCR more efficient. (Y. Y. Chiang & Knoblock, 2015) investigate the recognition of text elements on raster maps by reducing the user interference of manual input for detecting heterogeneous raster map text labels. Their two-step approach involves text layer extraction, identifying colors representing the map and managing text layers, and text label recognition, integrating techniques from text orientation identification (Y. Y. Chiang & Knoblock, 2010), and heterogenous and multi-sized text recognition (Y. Y. Chiang & Knoblock, 2011). Finally, they utilize the power of the commercial OCR engine, ABBYY FineReader[5], to recognize the textual elements. However, the thesis aims to shift the focus towards open-source OCR engines for extracting texts from maps to enhance map reproducibility. This transition from commercial to open-source OCR technologies introduces a novel aspect to the research, emphasizing the exploration of cost-effective and accessible solutions for achieving accurate map text extraction, comparison, and assessment.

Y. Y. Chiang & Knoblock (2013) introduce a general approach for text detection and recognition of raster map images. They implement the map text extractor tool called Strabo[6], which extracts texts by identifying text colors, automatically detecting label orientations, and rotating them horizontally to feed into the open-source OCR engine, Tesseract-OCR [7], for text label recognition. Nazari et al. (2016) improved the Strabo map text extraction tool with the introduction of the Overlapping Text Detection (OTD) algorithm. OTD classifies the areas in a map as text and non-text regions based on their shape descriptions, considering factors such as foreground pixel ratio, area size, connected components, and number of holes. This enhances the challenges related to overlapping characters and non-text symbols, contributing to accurate text extraction from maps.

Li et al. (2020) present a tool called mapKurator, focusing on automated geo-metadata generation from historical maps by integrating OCR technology. Schlegel (2021) proposes an effective methodology for automated text extraction from large-scale historical maps, addressing challenges in label extraction with tools such as Strabo and Tesseract OCR. Barget (2022) presents Transkribus Lite, a browser-based OCR tool for historical map texts. limitations in map text label extraction are addressed through manual transcription. While these studies contribute mainly to automated text extraction, it's important to note that this paper specifically targets contemporary maps. Unlike historical maps, contemporary maps may experience fewer limitations related to compression, noise, and text intersections with non-text artifacts.

Recently, adaptive open-source OCR engines have been increasingly employed, allowing systems to dynamically adjust to diverse font styles, sizes, and symbols encountered in map images. In addition, preprocessing methods play a crucial role in enhancing OCR accuracy by reducing noise and improving overall image quality (Brisinello et al., 2017). Furthermore, the integration of a large dataset of natural scene images in the training of OCR models offers a promising effect and contributes to improving the recognition and extraction of textual elements from maps (Zhang et al., 2013). These advancements collectively contribute to enhancing OCR performance, enabling more accurate and efficient information extraction from map images. In this case, investigating the performance of OCR technology in extracting textual elements from map images sheds light on the application of OCR technology to

[5] https://www.abbyy.com/
[6] https://github.com/spatial-computing/strabo-text-recognition-deep-learning
[7] https://github.com/tesseract-ocr/tesseract

geoscientific maps in addition to the comparison process for assessing the reproducibility of maps.

In terms of improving the accuracy of detecting and recognizing textual elements from a map image, H. Li et al. (2018) developed an intelligent map reader based on a deep learning approach for text detection, graph-based segmentation for the separation of texts, and an OCR engine for recognition purposes. This offers insights into improving the accuracy of the map text extraction process.

2.5 Fuzzy Based Text Matching

Fuzzy string-matching techniques have been instrumental in various fields for comparing and analyzing textual data, particularly in handling variations and errors in text inputs (Lan & Longley, 2019; Lokeshkumar et al., 2020). These techniques, based on methods such as the Levenshtein distance, offer effective means of assessing textual similarity and standardizing data across different sources (Zohar, 2021; Vukatana, 2022).

In the context of map text analysis, fuzzy-based text matching holds significant potential for improving the accuracy of comparisons and assessments. However, while existing research has explored its application in fields such as historical census data geocoding and social media text analysis, its integration into map reproducibility assessment remains underexplored (Guo et al., 2015; Zohar, 2021). This presents a notable gap in the literature, as the challenges of assessing and comparing extracted texts from maps demand innovative solutions to ensure reproducibility and accuracy.

To address this gap, this thesis aims to leverage fuzzy-based text matching techniques in the context of map reproducibility assessment. By applying these techniques to OCR-processed map texts, it also seeks to enhance the accuracy of text comparisons and minimize variations and errors inherent in textual elements. This approach offers a novel method for improving map reproducibility assessment, ultimately contributing to greater transparency and reliability in geoscientific research.

3 Conceptual Analysis and Approach

In this chapter, the conceptual approaches and solutions used for facilitating the assessment of map reproducibility using the textual elements of maps are presented and discussed. It begins by outlining the approach used for actual text extraction from maps, followed by a detailed analysis of the extracted texts. Next, map text comparison metrics are introduced for comparative assessment of textual elements, and then the logical comparison approach is designed based on the defined attributes. Finally, the applicability of OCR technology and its performance in extracting text from maps are discussed.

The proposed approaches and solutions in this thesis are supported by developing a web-based application. Basically, the developed application works as illustrated in Figure 1. Original map images and reproduced map images are processed through the text extractor module, generating extracted textual data. The extracted textual elements are then parsed, prepared for the comparative assessment process, and fed into the assessment strategy. The strategy performs computations of text similarity, bounding box overlap ratio, and distance between the bounding boxes accordingly for the processed data and generates reproducibility assessment results based on the logical comparison of the assessment attributes under the defined threshold (cutoff) values for each attribute.

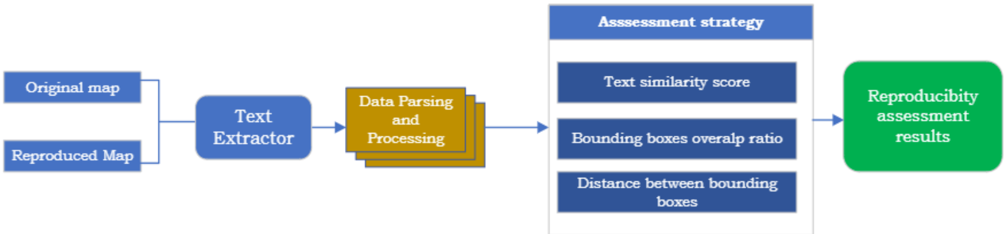

Figure 1. Simplified workflow of the developed application.

3.1 Map Textual Element Extraction

In this study, the first phase of the map reproducibility assessment approach begins with the actual extraction of texts from the input map images. Figure 2 shows the workflow for the extraction and parsing of textual elements from maps, which involves input data pre-processing, the text extraction process using an optical character recognition (OCR) engine, and output data parsing and processing.

Figure 2. Map text extraction process workflow.

i. Input and Data Pre-processing

Geographical maps can be designed and produced in a variety of ways and contain different features to convey information within the corresponding area (Elzakker, 2004). Normally, maps are composed of title, body, legend, label, ancillary text, scale bar, north arrow, and other additional metadata elements; however, it's not mandatory that all the maps include these elements, and it also differs from map types and the purpose of the maps that intensify the specific data (Dastrup et al., 2022). Proper placement of these map elements is important for accurately representing geoscientific findings and results. However, any misplacement, alteration, or absence of these elements can lead to a significant variation in the map's interpretation and unsuccessful reproduction of maps (Koukouraki & Kray, 2023).

In this study, the input maps comprise the original and reproduced map images. Considering the sources of these maps, it's more likely that these two images could be varied by image file type, image size and resolution, and overall image quality. The scope of the study limits input map images to only contemporary digital maps based on the fact that older papers typically lack the integration of source codes and detailed methods. Even though raster digital maps utilize up-to-date GIS tools or modern mapping techniques such as R packages, they could still suffer from compression or noise and lose their overall image quality when they are integrated into publications, which limits the performance of the OCR engine.

For the scope of this experimental study, JPEG, PNG, WebP, and BMP have been selected as the supported file formats for input maps. Even though conflicts with dependencies limit development options for adding additional file formats, it can be customized for the future in accordance with OCR engine supported file formats. During the pre-processing phase, the compatibility of input maps with supported map image file formats is verified, ensuring seamless integration. And the language of the input map textual elements was also ensured. In this thesis, English was chosen in alignment with the widespread use of the language in geoscience publications.

ii. Text Extraction Process

Extracting textual elements from a map produces geospatial knowledge and encourages understanding the point of interest in the map. This process not only facilitates comprehension of points of interest on a map but also plays a pivotal role during the assessment of map reproducibility based on the textual elements of the map.

In this study, the proposed approach to extracting the textual features of the map is independent of the differences in map image resolution, file sizes, and file types between the original and reproduced maps. The approach for the actual text extraction process can be handled in two ways. These are running the OCR using Python scripts with command-line options or integrating the OCR engine into a web-based application. The script with a command-line approach gives simplicity and control over the web-based application; however, the requirement to put hardcoded image paths or pass them as arguments limited the flexibility, making it less user-friendly. On the other hand, the approach to integrating the OCR engine into a web-based application implemented in this paper introduces user-friendly and attractive dimensions to the extraction process. And it facilitates real-time user interaction, concurrent processing of multiple maps, and handling the extracted textual elements easily.

iii. Output Parsing and Processing

Following the completion of the text extraction process, the output from the OCR engine becomes available for processing and parsing. The output structure is depicted in Figure 3; each entry follows a consistent format. The first list in each entry represents the bounding box coordinates, defining the spatial extent of the detected text label within the map. Following the bounding box coordinates, a tuple is provided. This tuple contains two elements: the actual text content extracted from the map and its corresponding confidence score.

```
[[
  [[142.0, 11.0], [154.0, 11.0], [154.0, 18.0], [142.0, 18.0]],
  ('DK', 0.6918708682060242)
],
[
  [[13.0, 50.0], [92.0, 31.0], [96.0, 47.0], [17.0, 67.0]],
  ('NORTH SEA', 0.9354690909385681)
]]
```

Figure 3. Sample output from the OCR engine.

Bounding boxes denote the position (spatial context) of text elements, preserving the positional information within the given maps. This spatial context is used for assessing the placement, overlap, and alignment of text elements, contributing significantly to the comparison and assessment of map reproducibility. The confident score describes how confident the OCR model is in detecting the text label.

To utilize the output (Figure 3) for the assessment process, the texts along with their corresponding bounding boxes and confidence scores needs to be parsed and processed. This processing of outputs is important because there is a possibility of getting the same extracted text elements from different positions on the original and reproduced maps. This makes it challenging to identify which text element is detected at which position. To address this issue, the study leverages the bounding boxes associated with each text as a means of unique identification. These bounding boxes serve as unique identifiers for each text element, representing the areas detected by the OCR within the map image. This approach aims to facilitate the accurate comparison and assessment of corresponding text elements between the original and reproduced maps by providing distinct identities across all lists of texts.

3.2 Text Analysis

Analyzing the extracted textual elements from maps that will be utilized in the comparison process is essential for gaining insight and understanding of the texts. Geographic maps contain a variety of text types (Dastrup et al., 2022). In general, text and label default tasks are mainly for naming geographic features and important components of a map (Pun-Cheng, 2000). Labels help to confirm the features of a location, represent the shape and characteristics of the location, and represent the relationship between the location features. Text, in the form of descriptions, is also used for very important map design features such as title, data sources or meta data, map projection, map labels, axis labels, legends, and scale bars (Dastrup et al., 2022). Maps can have different characteristics and features in different contexts; however, texts and labels increase the understanding of a location or place in a way that cannot be replaced by other visual elements such as color or intensity of a map (Axis Maps, 2020).

The text analysis approach mainly centers on leveraging the extracted textual elements from both original and reproduced maps. These elements constitute collections of texts found in both input maps, denoting the scientific findings. These textual elements consist of strings, numerical values, and special characters. The accuracy of these texts is dependent on the input map image quality, font size and muti-sized texts on the map, and text orientation, as it makes the text inaccurately detected and recognized by the OCR engine (Y. Y. Chiang & Knoblock, 2010), presenting a potential error and discrepancies that can also lead to misinterpretations of the results. For example, textual elements could be correctly detected and recognized in the original map image but inaccurately detected and recognized in the reproduced map image, introducing bias and misjudgement during the comparison process for assessment of map reproducibility. To evaluate the accuracy of the texts, a confidence score, the certainty of OCR in recognizing individual characters or words within the identified text, is assigned to each text element. The confidence scores range from 0 to 1, with higher values denoting greater confidence in the accurate recognition of texts (Neudecker et al., 2021). However, a specific cutoff value of the confidence score for distinguishing OCR failures was not utilized.

In addition, for the scope of this thesis, geocoding is not performed to verify the accuracy of the extracted texts. This decision was made because certain map elements, as outlined in Table 1, did not contain embedded location information. Therefore, relying on geocoding services would not have been beneficial, given the absence of necessary data within these elements. This decision aims to avoid potential bias during assessment, ensuring a fair and unbiased evaluation between extracted lists. By excluding geocoding, the thesis maintains its focus on a neutral and objective analysis of textual elements.

Table 1. Texts extracted from north arrow, axes, and legend of the map.

Map element	**Detected text**
North arrow	N
Axes	60E
legend	0-100

Moreover, challenges and limitations also arise when dealing with special characters and irregular font styles in the texts. The OCR engine could extract characters as different characters, such as ‘β’, could be detected as ‘B’, introducing variations in the texts. This kind of discrepancy introduces a potential bias in the assessment of texts. This holds for the overall comparison and assessment of texts since post correction methods are not applied in this study. However, the use of fuzzy-string matching techniques in this study solves the issue to some extent.

3.3 Comparative Assessment of Map Textual Elements

This section presents the similarity measures utilized in this study to compare the extracted text from both the original and reproduced maps. The goal is to facilitate the assessment of map reproducibility by evaluating how well the textual elements match between the two versions of the map. This approach consists of the following similarity measures: text similarity, bounding box overlap, and distance between bounding boxes, and they are carefully constructed to enhance the overall comparison of textual elements.

3.3.1 Text Similarity

The first metric of the comparative assessment approach is to evaluate the similarity of the input textual elements. The aim is to distinguish how this text similarity contributes to determining the reproducibility of textual features on maps. This involves conducting pairwise comparisons of extracted textual elements. In this process, the texts extracted from the original map serve as a reference point for the corresponding text elements in the reproduced map during the comparison process. This supports the reproduced textual features of the map to be assessed based on the texts extracted from the original map. In addition, any possible variations in the length of the list of extracted original and reproduced text elements are considered.

The text similarity metrics lie in the concern of whether the text elements in the original map change their value or disappear in the reproduced map. The approach incorporates a fuzzy-based string matching technique for computing text similarity between these texts. This technique is chosen for its ability to handle text elements that may not exhibit exact matches (Lokeshkumar et al., 2020). The decision is also due to the nature of the OCR-processed input text elements. The OCR process, while powerful, is not perfect and may lead to the detection of incorrect text elements(Nguyen et al., 2022). Fuzzy matching techniques excel at accommodating variations and errors in OCR outputs, including wrongly detected texts. However, it's important to note that while these techniques enhance the comparison process, they may not fully address potential inaccuracies in the OCR-processed text elements. In some instances, fuzzy matching could unintentionally enhance inaccuracies by emphasizing similarities in incorrectly recognized text elements (Vukatana, 2022).

The proposed text similarity approach results in a text similarity score, which denotes the degree of resemblance between the textual elements. The maximum value of the text similarity score is 100 when two input texts are exactly matched, and the score is 0 when there is no similarity between the texts. The increase in the similarity score correlates with an increased probability that the compared texts are matching, and vice versa. So, there exists a direct proportional relationship between the obtained similarity score and the probability of textual similarity. In this study, this text similarity criterion is classified as a benefit criterion, indicating that a higher similarity score benefits the assessment process by representing a greater likelihood of successful textual matching. This criterion plays a significant role in the design of the equation for the final score computation in Section 3.3.4.

3.3.2 Bounding Box Overlap

Relying solely on text similarity to facilitate the assessment of map reproducibility is not a guaranteed approach. This is because the texts on the original map could be displaced from their original position in the reproduced map, which affects the comparison and evaluation of the input's texts for the assessment of map reproducibility. To handle this issue, a bounding box overlap metric is introduced to determine the spatial positional accuracy of the extracted texts. In general, identifying the position of text elements in maps is important, particularly for geographical map labels. However, the approach used in this thesis does not differentiate the implementation between geographical labels and other non-geographical text elements such as titles, legends, and scale bars because all the textual elements equally contribute to the overall map reproducibility assessment. This approach reflects the real-world scenario where all text elements contribute to the overall visual and informational context of the map.

Moreover, it's essential to recognize that text elements on the map may share similar textual content while representing different meanings and occupying different positions. For example, the text '6%','80E' could be repeated several times in different positions or locations on the map and make it too difficult to differentiate after they are extracted from the maps. To tackle the issue, this approach aims to identify the spatial alignment of texts by providing an overlap ratio between the input bounding box coordinates utilized as a unique identification method for each extracted text from the input original and reproduced maps.

The **Jaccard Index**, also known as the Jaccard Similarity Coefficient, is a statistical tool for determining how diverse and similar two sample sets are (Verma & Aggarwal, 2020). When applied to bounding boxes in this scenario, the Jaccard similarity coefficient (Figure 4) assesses the similarity between two bounding boxes by calculating the ratio of the intersection area to the union area of the given bounding boxes. The Intersection over Union (IoU) value ranges from 0, indicating no overlap between the bounding boxes, to 1, representing that the bounding boxes perfectly align and overlap(Ogwok & Ehlers, 2022). This contributes to determining the spatial accuracy of the extracted elements from the maps.

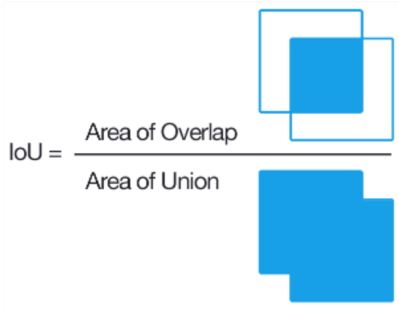

Figure 4. Visual Intersection over Union (IoU) equation, Image Adapted from "Intersection over Union (Jaccard Index) visual equation" by Adrian Rosebrock (2016, November 7). Licensed under CC BY-SA 4.0. Retrieved from http://www.pyimagesearch.com/2016/11/07/intersection-over-uni.

The study leverages the Jaccard index (intersection over union) compared to other metrics such as intersection over minimum (IoM) and bounding box regression loss due to its widespread use, simplicity, straightforward interpretation, suitability for thresholding, direct measure of localization accuracy, and alignment with training objectives in many object detection frameworks (Guan et al., 2022).

However, the bounding box comparison approach poses challenges for input map images that have different dimensions, and they introduce potential errors. This is because the bounding box coordinates, extracted from images with different dimensions of maps, result in inaccurate overlap ratio values due to inherent variations in map image scales. To address this issue, an approach for comparing bounding boxes in a normalized space is developed. This process involves dividing the coordinates of each bounding box by the respective dimensions of the image from which it was extracted, generating relative coordinates. This normalization process enables meaningful comparisons across map images with differing dimensions.

The general formula for normalizing the coordinates of a bounding box (BBox) with respect to its image dimensions (width and height) is expressed as follows:

$$\text{Normalized coordinates} = \left(\frac{BBox\_x}{width}, \frac{BBox\_y}{height}, \frac{BBox\_width}{width}, \frac{BBox\_height}{height}\right)$$

Here, *BBox_x* and *BBox_y* represent the top-left corner coordinates of the bounding box, while *BBox_width* and *BBox_height* denote its width and height, respectively. Dividing each of these parameters by the corresponding image dimensions (width and height) results in a normalized bounding box that facilitates consistent and accurate overlap ratio comparisons across diverse map images.

In this study, the aim of the bounding box overlap ratio calculated using the intersection over union (IoU) algorithm is to assess the spatial positional accuracy of textual elements. The overlap ratio, ranging from 0 (no overlap) to 1 (bounding boxes perfectly aligned), serves as a quantitative measure of spatial positional accuracy for the input text. The highest values indicate that the bounding boxes associated with the texts have the highest spatial positional accuracy, whereas the lower values indicate the lowest spatial positional accuracy of the given bounding boxes (Agarwal, 2023). There is direct proportionality between the overlap ratio and spatial positional accuracy. The higher the values of the overlap ratio, the more favourable the spatial positional accuracy of the bounding box, which is thereby classified as a benefit criterion. Finally, this attribute contributes to the final score computation in Section 3.3.4.

3.3.3 Distance Between Bounding Boxes

The approach for comparing texts using the text similarity score and spatial alignment or the bounding box overlap ratio, is not always perfect for facilitating the assessment of map reproduction. This is because in some cases, when comparing original and reproduced textual elements, even though the input texts are similar and convey the same message, the text might not align due to small differences in their spatial positioning. This makes it challenging to accurately measure how similar the text content is between the two maps because the bounding box overlap ratio of these input texts will be zero (no overlap). To tackle this issue, the study introduced the Euclidean distance metrics between the normalized bounding box coordinates on the normalized space. This metric involves determining the centroid of the bounding box coordinates. The Euclidean distance between the centers of the bounding boxes is calculated for the purpose of determining the spatial separation between the bounding boxes associated with the texts. If (x1, y1) and (x2, y2) are the center coordinates of the bounding boxes, the Euclidean distance can be calculated as follows (Euclidean Distance, 2024).

$$\text{Euclidean Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The calculated distance is expressed in normalized pixel units. The advantages of using Euclidean distance in this thesis are that it captures both the direction and magnitude of the spatial separation between the bounding boxes. The maximum distance between two bounding boxes in a normalized pixel space is the diagonal distance of the normalized space. Since the normalized space ranges from 0 to 1 in both the x and y directions, the maximum distance between two points occurs when one point is at (0,0) and the other is at (1,1). For the diagonal distance in normalized pixel space:

Maximum Normalized distance= √ (1−0)2+(1−0)2= √2

The maximum distance in normalized pixel space (√2 or approximately 1.414) acts as the pivotal threshold. This helps in evaluating the degree of displacement between text elements, determining changes in the original position of texts in the reproduced map, and contributing to the overall map reproducibility assessment. Distances above this value indicate variations or biases in bounding box coordinate normalization.

The Euclidean distance between the bounding boxes assists in determining the space between the input texts, which facilitates the detection of spatial variation. In this case, the calculated distance between the bounding box coordinates on the normalized space ranges from 0 to the highest possible distance of (1.414) pixel units. The highest values indicate the lowest positional similarity since they represent high spatial separation between bounding boxes, whereas the lowest values represent the highest positional similarity as they represent minimal spatial separation. In this case, there is an inverse relationship between distance and positional accuracy. Therefore, the criterion is classified as a cost criterion and assists in developing the final score equation metrics in Section 3.3.4.

3.3.4 Final Score Calculation

Leveraging the metrics discussed above: text similarity score, bounding overlap ratio, and the Euclidean distance between the bounding boxes associated with the texts, the study introduced a final score metrics, which is designed for reliable comparative assessment tasks. For the purpose of this thesis, these metrics are classified in such a way that text similarity score and overlap ratio are considered as benefit criteria and distance as cost criteria for determining the similarity of texts in assessing the reproducibility of maps. The final score equation is formulated for the calculation of the final score as follows:

$$\textit{Final Score} = \textit{Text Similarity Score} + \textit{Overlap Ratio} - \textit{Distance}$$

According to this equation, the maximum value of the final score is 101. When the texts are exactly similar (100), the bounding boxes associated with the texts are aligned perfectly (1), and there is no distance between the bounding boxes or spatial variation between the texts (0). The equation notably prioritizes the text similarity score, which can reach a maximum value of 100. This underscores the significant emphasis placed on text similarity within the assessment framework. Conversely, the other metrics, such as the overlap ratio and distance, contribute comparatively less to the final score, with their values limited to 1 and 1.414, respectively. To enhance clarity, the final score is converted into a percentage using the following formula:

$$\textit{Final Score percentage} = (\textit{Text Similarity Score} + \textit{Overlap Ratio} - \textit{Distance})/101 * 100$$

In terms of map reproduction assessment, this value represents a measure of how well the information in the assessed text aligns with the expected content, considering both common elements and differences. The highest value indicates a match between the texts in both original and reproduced maps, thereby assisting in improving the map reproduction process.

3.4 Logical Text Comparison Process

In this section, the comprehensive logical comparison approach is designed to determine the final match status of the texts from both maps. The process consists of all the attribute text similarity scores, the overlap ratio between the bounding boxes, and the Euclidean distance between the bounding boxes discussed above. In this experimental study, the process of defining the threshold values for each attribute to determine the matching status between the input texts is conducted heuristically, guided by practical judgment, experience, and insights gained from testing the values on different input map images.

Figure 5 shows the overall logical text comparison flowchart. According to the designed logical text comparison process, the input texts are called **matched** if the text similarity score is greater than or equal to 70%, the distance between bounding boxes in normalized space is less than or equal to 0.05 pixels, and the final score is greater than or equal to 50%. Conversely, the input texts are called **not matched** if the text similarity score is less than 70%, the distance between bounding boxes in a normalized space exceeds 0.05, and the final score is less than 50%.

Text Similarity Threshold (70%): This threshold maintains a balance between capturing meaningful textual similarities while accommodating potential variations due to OCR errors. Through experimentation and validation, it was determined that a 70% similarity threshold effectively distinguished between text elements that represent the same content versus those that differ significantly.

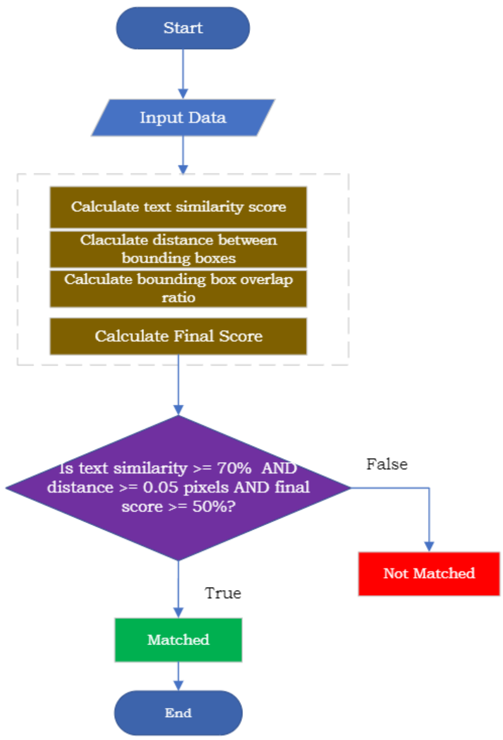

Figure 5. Logical text comparison flowchart.

Bounding Box Overlap Ratio: This criterion evaluates the spatial alignment between text elements from the original and reproduced texts. In this study, no specific threshold or cutoff value is defined for the bounding box overlap ratio in the logical comparison process. Instead, this criterion is used in the calculation of the final score, part of the comparison process between the maps. While this criterion may not directly contribute to the logical comparison process for determining match status, its significance lies in facilitating the overall assessment process.

Distance Between Bounding Boxes (0.05 pixels): The cutoff value for the distance between bounding boxes was set to a minimum of 0.05 pixels to account for potential variations in the positioning of text elements within the map. This threshold allows for a small margin of error in the spatial alignment of text elements while still maintaining sufficient proximity to be considered.

Final Score Threshold (50%): The final score of 50% or higher effectively distinguishes between text elements that exhibit sufficient similarity and spatial alignment to be considered matches and those that do not. This threshold ensures that only text elements meeting a certain level of similarity and spatial proximity are classified as matches, enhancing the accuracy and reliability of the matching process.

3.5 OCR Performance Investigation

In addition to conceptual approaches for facilitating map reproducibility assessment, the study develops an approach to investigate the performance of the integrated OCR engine to extract textual elements from geographical maps. This approach utilizes a custom-made dataset of maps that contains 198 maps with different resolutions and font sizes created using ArcMap 10.8 and available in Appendix A. For this thesis, only yellow (#FFFF00) and white (#FFFFFF) backgrounds were chosen for the map images. So, 99 of the maps have the same white background, while the remaining 99 have a yellow background. These maps have different font sizes, starting from the small font that is supported by the GIS software (5 pt) to the large font size (14 pt) that can be easily detected visually for both white and yellow backgrounds. In addition, the maps have different image resolutions, starting from the very lowest resolutions (20 dpi) to the highest resolution (200 dpi) for both white and yellow backgrounds. The texts on the test dataset of the maps are written in English. The total number of texts in the maps is 22, which are not individual characters but sets of words with a consistent black font colour for each text and Ariel standard font type.

The choice of yellow and white backgrounds for OCR performance evaluation is rooted in empirical evidence and aims to investigate the influence of background colour on reproducibility assessment. By examining OCR performance under different colour conditions, the study seeks to enhance our understanding of the reproducibility challenges inherent in geoscientific maps. In addition, Mello & Lins (1999) shed light on the significant impact of background colour on OCR accuracy, with yellow backgrounds yielding lower hit rates compared to blue, green, and pink colours. This finding underscores the importance of strategically selecting this background colour to investigate and compare with the white background, which is the most common background colour for geographical maps.

4 System Implementation

In this chapter, the proposed solutions for developing web-based automatic text extraction, parsing, and comparison of map textual elements to facilitate the map reproducibility assessment are implemented and discussed.

4.1 Web Application Development Platforms

The study uses the React.JS framework for the implementation of the web-based user interface for uploading maps and visualizing reproducibility assessment results, as it's known for its flexibility and efficiency in building interactive and responsive single-page web applications (Rawat & Mahajan, 2020). In addition, HTML5 and CSS are also used for designing and styling the web page. The node module (NPM Version 8.19.2) is used for wrapping and installing all the front-end libraries.

For uploading the input map images, local file system storage is used. The local file system is relatively cheaper and easier to implement than developing database systems in terms of cost and effort (Singh, 2022). The study uses the Python programming language (Python 3) for the development of backend environments, as it has an open-source nature and efficient data processing capabilities. This helps with the efficient processing of extracted texts from large datasets of map images. Most importantly, the selection of this programming language depends on the compatibility of the selected OCR model, which helps ease integration with the developed application. Flask Python web framework is used for its lightweight and minimalist design, which proved to be a fitting solution for the backend server implementation (Relan, 2019).

The full version of the source code for the application is written in JavaScript(React.JS), Python(Flask), HTML5, and CSS and is available at GitHub. All the links for the libraries and dependencies are available in Appendix B.

4.2 OCR Model Integration

The integration of the OCR engine into the web-based application for proper text extraction from map images requires the selection of a suitable OCR engine. In this study, the selection of the optimal OCR engine was based solely on open-source engines. This integration of an open-source OCR engine aligns with the aim of promoting open reproducible research.

In this study, PaddleOCR was chosen as the main OCR engine for the purpose of text extraction from map textual elements. The selection of this open-source OCR model is based on its ability to detect texts rotated by 0 and 180 degrees, model compatibility with the backend server, and ease of integration with web applications. In addition, PaddleOCR represents a recent advancement in OCR technology and known for its versatility in extracting textual elements across both scanned documents and natural scene images, its flexibility, and its multilingual capabilities, as highlighted by Mageshwaran R (2021), it is a fitting choice for the goal outlined in this thesis.

The PaddleOCR engine offers a comprehensive text recognition process with its collection of high-quality, pre-trained OCR models. In this paper, the latest version of PaddleOCR (PaddleOCRv3) is used. These models are freely available as open-source software, and the code can be found in the GitHub repository named PaddleOCR, designed by PaddlePaddle (C. Li et al., 2022). The PP-OCR model specifications contain the text detection model (DB) (Liao et al., 2019), text angle classification (CLS) (Howard et al., 2019), and text recognition (Single Visual Model for Scene Text Recognition (SVTR) model) (Du et al., 2022). The workflow of the PP-OCRv3 pipeline is illustrated in Figure 6.

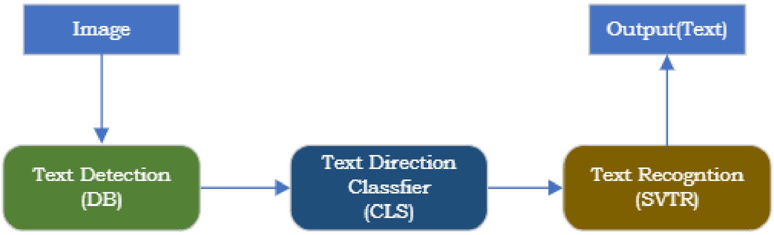

Figure 6. PP-OCRv3 simplified architecture adapted from C. Li et al., (2022).

The text detection step in PP-OCRv3 is used to identify the bounding box of the text within the input image. The OCR engine uses Differentiable Binarization (DB) for the text detector algorithm. Differential binarization (DB) simplifies the text identification process by optimizing the binarization step within segmentation networks. By producing high-quality binary images directly, DB reduces the need for complex post-processing algorithms. This simplification results in cleaner and more accurate text identification, significantly reducing post-processing complexity (Liao et al., 2019). Text angle classification (CLS) is a simple technique for an image classification model (Howard et al., 2019). Its main function is to specify the orientation of the text and correct the text box orientation if it is determined to be reversed. This classification is particularly useful for scenarios where the image is not aligned at 0 degrees. In such cases, correcting the orientation of the detected text lines becomes essential(Lo & Chou, 2022). In the final stage, the SVTR is responsible for recognizing the text within the identified bounding box and producing string-based results(Du et al., 2022).

For proper integration and utilization of the PaddleOCR engine, two installation strategies are available, each serving different resource environments. The first strategy leverages the machine's GPU for enhanced efficiency, while the second strategy uses CPU installation, prioritizing resource availability and compatibility. For this study, the second strategy is implemented. Further settings are customized during OCR model instantiation, such as the ability to use an angle classifier for rotated texts and language options. However, due to resource limitations and development issues with the integration of the OCR language model, only English characters are supported in this study.

4.3 System Development Architecture

The system development architecture (Figure 7) acts as the technical infrastructure supporting the conceptual framework outlined in Chapter 3. This section describes in detail how each architectural component serves to implement and quantify the conceptual approach towards map reproducibility assessment.

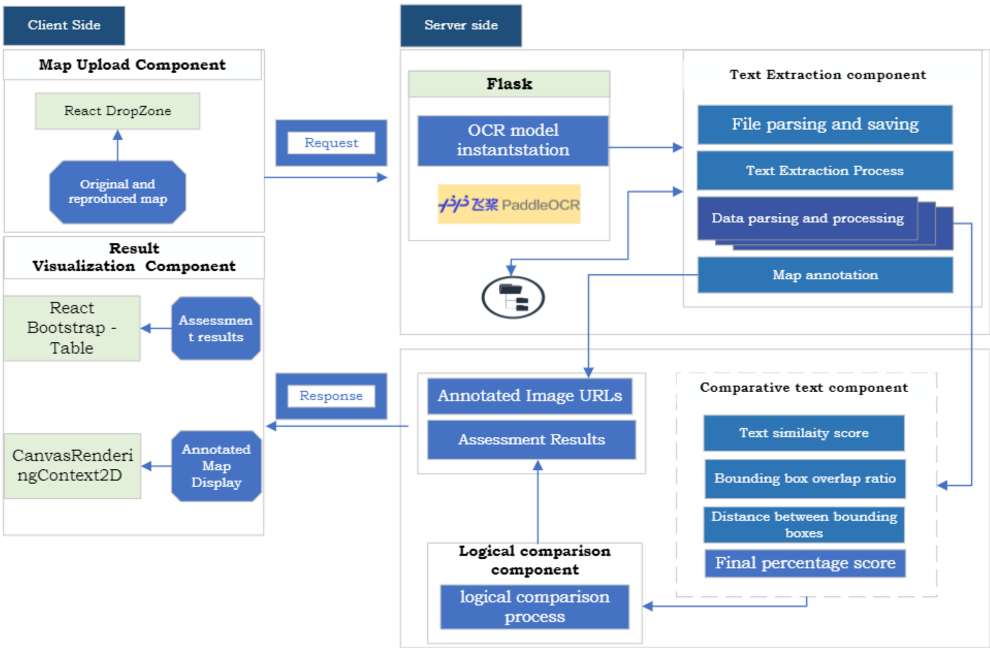

Figure 7. Web application development system architecture.

I. Client-side Architecture

The client-side architecture encompasses the application interfaces, facilitating user interaction with the map reproducibility assessment application, enhancing the user experience, and enabling seamless map image input and result visualization.

The map upload component enables system users to seamlessly upload both the original and reproduced map images. This component integrates the react-dropzone library for efficient and user-friendly file handling, and React-bootstrap-icons enhances the user interface with scalable icons. It handles components such as file type validation, asynchronous file uploads, automatic file preview, and a visually appealing interface.

The results visualization component uses Reacts Bootstrap Table to display the key metrics of reproducibility assessment results and utilizes CanvasRenderingContext2D for side-by-side visualization of annotated original and reproduced maps with additional interactivity, such as users selecting specific rows in the table to highlight corresponding text elements on both maps, aiding the overall assessment process. To ensure the accuracy of the visual highlighting of the text on both map images, a bounding box normalization process occurred during the canvas drawing process. Scaling factors are calculated based on the input map image dimensions. This normalization guarantees a consistent and proportional drawing of bounding boxes, regardless of variations in the original image dimensions.

II. Server-side Architecture

The server-side architecture is built to support the conceptual objectives of the map reproducibility assessment framework and is critical for text extraction, processing, and parsing. A virtual environment [8] is created to provide the foundational infrastructure for server-side operations and wrap all libraries, platforms, and dependencies required.

Upon uploading both maps, the component uses Axios to make an HTTP request (REST API) to the Flask server. The web server receives the maps and stores them in local file system storage for the purpose of text extraction and ensuring system scalability and efficiency. The text extraction component leverages the integrated PaddleOCR engine to facilitate the extraction of textual elements within map images and parsing extracted text. OpenCV is used for reading and writing map image files during the map annotation process. And the comparative text assessment component motivates the quantitative evaluation of map reproducibility using the similarity metrics discussed in Section 3.3, ensuring the accuracy and reliability of assessment outcomes.

4.4 Implementation Details

This section provides a comprehensive discussion of the implementation process for the application's fundamental components. Specifically, it describes the practical aspects of text extraction from map elements and explains the main modules managing the text comparison process.

4.4.1 Map Textual Element Extraction

One module was implemented for the map text extraction process in accordance with the approach discussed in Section 3.1. The text extraction module (Figure 8) involves accepting map images originating from the requested files on the Flask server. This process starts with saving the map images to file system storage for efficient file handling and retrieval. After the map image files are saved, the text extraction process begins using the OCR model object previously instantiated when the Flask server starts. After the completion of the extraction process the OCR returns the results in the form of nested lists and tuples (Figure 3).

Subsequently, the module parses and process the results, computes the dimensions of original and reproduced maps, and annotate the map images by their bounding boxes Finally, the output of this module includes structured data with annotated map image URLs, extracted texts with their respective bounding box coordinates and confidence score, and map image dimensions and used as input for the text comparison component.

[8] https://docs.python.org/3/library/venv.html

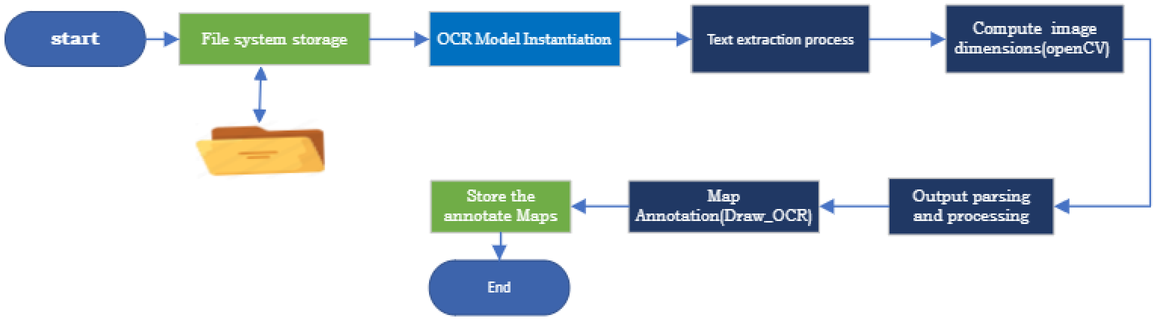

Figure 8. Text extraction module workflow.

4.4.2 Comparative Assessment of Textual Elements

Four modules were implemented for the comparative assessment of extracted textual elements for the overall assessment of map reproducibility. These modules are implemented in accordance with the proposed approach presented in Section 3.3.

Text similarity module:

In this module, **theFuzz**[9] fuzzy string-matching Python library is used for calculating the text similarity between the lists of texts, which was previously known as **fuzzy wuzzy**[10]. It uses Levenshtein distance to calculate the difference between sequences in a simple to use package and works by measuring the difference between two string sequences (Levenshtein & et al., 1966). The rationale to choose a library based on Levenshtein distance, which is an ideal algorithm for short, non-tokenized strings and helps to capture spellings, word order and choice differences effectively compared to Cosine and Jaccard similarity algorithms (Kysela, 2018). This algorithm gives a measure of the number of single character insertions, deletions, or substitutions required to change one string into another. Levenshtein distance is calculated mathematically as follows (Figure 9):

The Levenshtein distance between two strings a, b (of length $|a|$ and $|b|$ respectively) is given by $\operatorname{lev}(a, b)$ where

$$\operatorname{lev}(a,b) = \begin{cases} |a| & \text{if } |b| = 0, \\ |b| & \text{if } |a| = 0, \\ \operatorname{lev}\big(\operatorname{tail}(a), \operatorname{tail}(b)\big) & \text{if } \operatorname{head}(a) = \operatorname{head}(b), \\ 1 + \min \begin{cases} \operatorname{lev}\big(\operatorname{tail}(a), b\big) \\ \operatorname{lev}\big(a, \operatorname{tail}(b)\big) \\ \operatorname{lev}\big(\operatorname{tail}(a), \operatorname{tail}(b)\big) \end{cases} & \text{otherwise} \end{cases}$$

where the tail of some string x is a string of all but the first character of x, and $\operatorname{head}(x)$ is the first character of x. Either the notation $x[n]$ or x_n is used to refer the nth character of the string x, counting from 0, thus $\operatorname{head}(x) = x_0 = x[0]$.

Figure 9. The Levenstein distance formula from Levenshtein Distance, (2023)

The **theFuzz** has the following similarity scoring functions: the basic functions such as simple and partial ratios, advanced functions such as token sort ratio and token set ratio, drawn from

[9] https://github.com/seatgeek/thefuzz
[10] https://github.com/seatgeek/fuzzywuzzy

the Rapidfuzz[11] library, and combination API scoring functions such as QRatio (Quick Ratio), WRatio (Weighted Ratio), and UWRatio (Unicode Weighted Ratio) (Seatgeek, 2021).

This study implemented the UWRatio (Unicode Weighted Ratio) method from the similarity scoring functions for its ability to compute the similarity score as a weighting score leveraging all the scoring functions, and the method preserves Unicode characters from the input texts compared to the other functions. This method takes the input texts and the 'full process' parameter by default, which is responsible for a pre-processing of texts such as removing all the characters except letters and numbers and trimming the white spaces in the texts. This process introduces potential variations during text similarity score calculation, particularly when dealing with text elements containing special characters. So, the study implements a customized setting to escape pre-processing tasks and allow flexible computation.

In addition, the study introduced a confidence score as a weighting criterion for similarity score calculation, as the input texts are subjected to potential inaccuracies during text extraction from both original and reproduced maps. This integration ensures that texts with lower confidence scores contribute less to the similarity score. However, in scenarios where the extracted texts are exactly similar, the approach neglects the calculation of confidence score weighting to prevent a biased weighted score, as sometimes these texts could receive lower confidence scores from the OCR engine. Finally, the text comparison module returns the computed text similarity score (Figure 10).

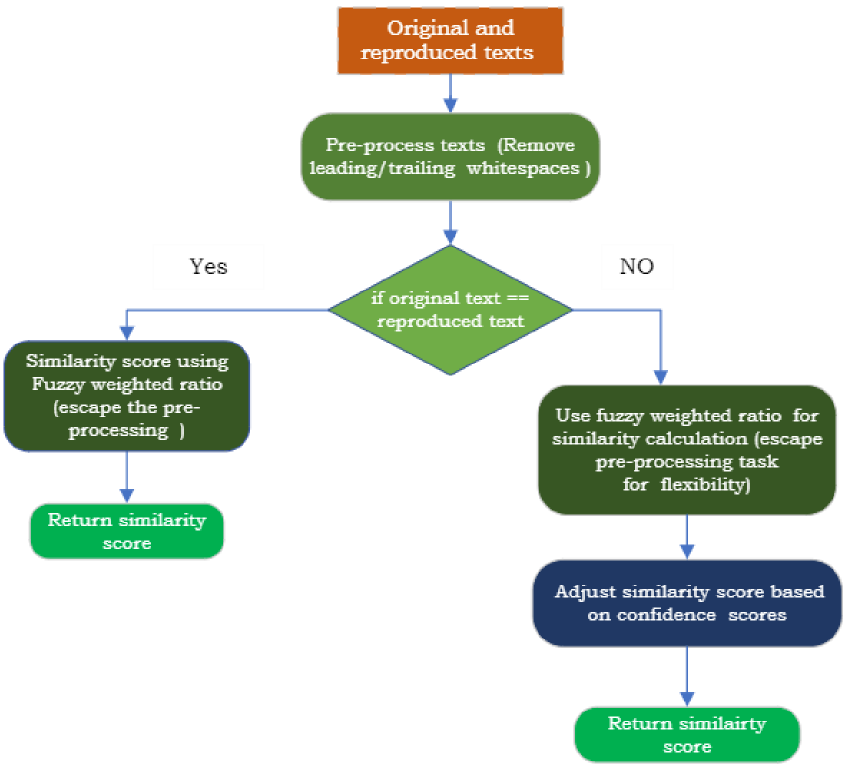

Figure 10. Text similarity score module workflow.

[11] https://rapidfuzz.github.io/RapidFuzz/

Bounding box comparison module: This module has four parameters: the bounding boxes of the original and reproduced texts and the image dimensions of both original and reproduced maps. The image dimensions are responsible for normalizing the input bounding box coordinates. The bounding box comparison module begins by normalizing bounding box coordinates and then applies the intersection over union (IoU) algorithm by calculating the area of overlap and area of union and dividing the area of overlap by the area of union. Finally, the bounding box comparison box module returns the overlap ratio (Figure 11).

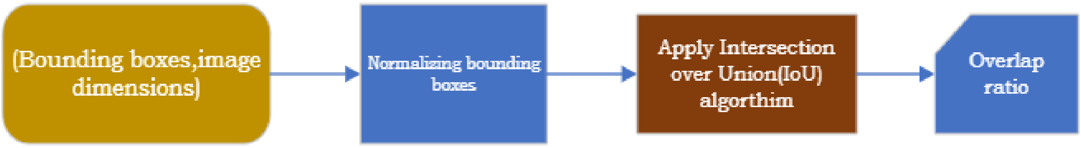

Figure 11. Bounding box comparison module workflow.

Euclidean distance between bounding box module: The module takes bounding boxes coordinates and their respective image dimensions of the maps, used for normalizing the coordinates and ensuring accurate distance calculations in normalized space. The computation begins by calculating the centroid of each input bounding box, and then the distance between these centroid points is computed.

Final score module: The method has the following three parameters: text similarity score from the text similarity module, bounding box overlap ratio calculated in the bounding box comparison module, and distance between bounding boxes from the Euclidean distance module, and applies the equation presented in Section 3.3.4. The module provides the percentage value of the final score.

5 Results and Evaluation

5.1 Results

The study presents two primary results. The first result assesses the web application in facilitating the assessment of map reproducibility through investigating web-based text extraction, parsing, and comparison, addressing Research Question One. This evaluation involves examining the degree to which the application aids in accurately assessing the reproduction of maps. The second part explores the performance of the optical character recognition (OCR) engine in extracting textual elements from map images, addressing Research Question Two. This investigation involves analyzing the accuracy and capability of OCR in identifying and extracting text from a map image.

5.1.1 Web Based Map Reproducibility Assessment

In this study, a web-based application prototype was developed to support the assessment of map reproducibility, in alignment with the conceptual framework outlined in Chapter 3. The project was developed under the MIT license, which is available as open source under the GitHub repository deployed on the Render cloud server. To evaluate the developed prototype, a map image dataset is collected from a repository of reproducibility reviews of geoscience publications during AGILE conferences (https://osf.io/phmce/) and custom-made map images are utilized. Furthermore, user evaluations were conducted to assess the efficiency and usability of the web-based application, as detailed in Section 5.2. These evaluations provided valuable insights into its practicality and user-friendliness, informing refinements and improvement.

Upon uploading both the original and reproduced maps into the application, the system generated reproducibility assessment results. These results, along with annotated maps, were then visualized side by side, providing a comprehensive view of the evaluation outcomes.

As an integral part of the web application task, the result visualization component (Figure 12) shows the side-by-side comparison of reproducibility assessment results coupled with annotated original and reproduced maps to enhance the comparison and assessment process. The reproducibility assessment results contain the extracted texts (original and reproduced), the computed comparative assessment of textual element attributes discussed in Section 3.3, and a match status that denotes the final matching between the input texts based on the designed logical comparison process of the attributes outlined in Section 3.4. The side-by-side visualization of the map images (original and reproduced) involves the annotation of texts using the bounding boxes to specify the exact position of the texts and differentiate between detected and non-detected texts on the map. Overall, this combination of assessment processes helps researchers or professionals get a more comprehensive view of how well reproduced maps convey the text of the original map. In addition, reproduced assessment results can be downloaded for further analysis and assessment purposes.

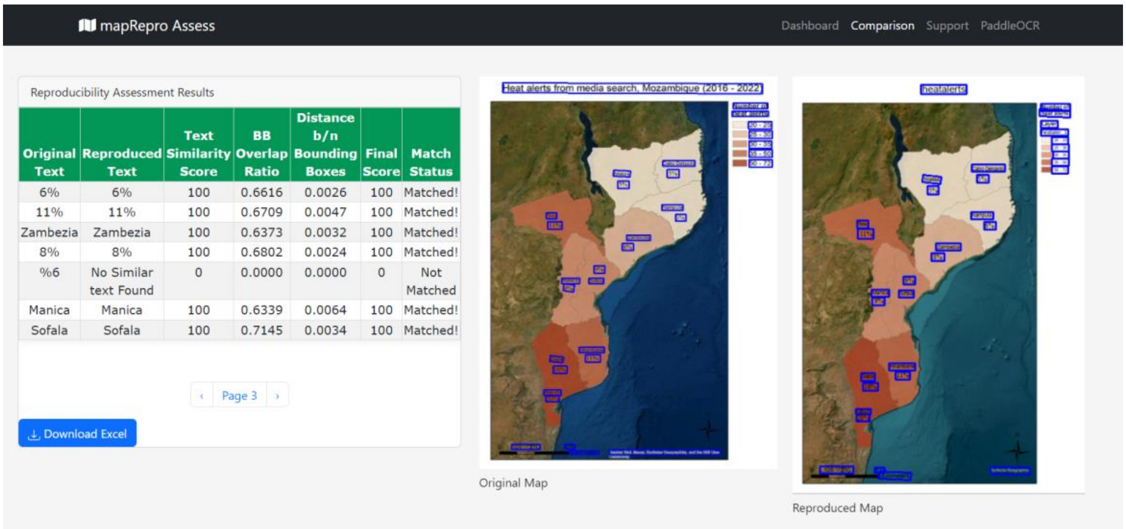

Figure 12. Side-by-side comparison of reproducibility assessment results and annotated maps *a. reproduced map image on the right: Adapted from Koukouraki, 2023, used under CC-BY Attribution 4.0 International License b. original map image on the left: Adapted from* Pereira Marghidan et al., 2023, *page 5, used under CC-BY Attribution 4.0 International License.*

To enhance the comparison and assessment process, the result visualization component (Figure 13) involves adding an interactivity feature to support the communication flow between the reproducibility assessment results and the annotated original and reproduced maps. When a specific row is selected from the assessment results table (1), the corresponding texts on both the original and reproduced maps are highlighted with a red rectangle (3)(4). Additionally, the confidence scores (5) retrieved from the OCR engine are displayed alongside the highlighted rectangle, providing insights into the OCR's confidence level for detecting the specific text. To further improve the user experience, tooltip information is incorporated into the columns of the assessment results table. Users can now hover over each column to access additional contextual information. The tooltips (2) offer brief, informative descriptions, guiding users through the meaning and significance of each column (assessment attributes).

In summary, the web-based nature of the application significantly enhances accessibility, interoperability, and ease of use, allowing geoscience researchers and GIS professionals to perform map text extractions and assessment process without the need of for specialized software. Furthermore, the design principle of the web application make it ease to compare the extracted texts using the comparative assessment tasks and adds visual comparison of the texts on both maps on real time for each extracted texts.

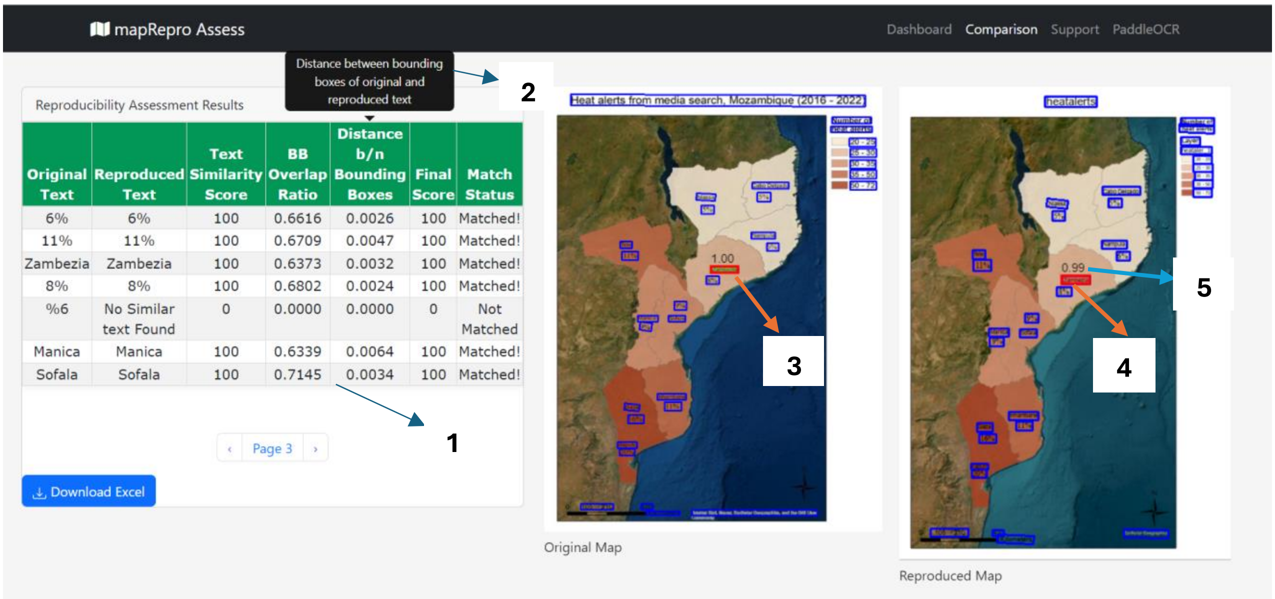

Figure 13. Interactive comparison method for facilitating the assessment of map reproducibility. *a. reproduced map image on the right: Adapted from Koukouraki, 2023, used under CC-BY Attribution 4.0 International License b. original map image on the left: Adapted from* Pereira Marghidan et al., 2023, *page 5, used under CC-BY Attribution 4.0 International License.*

5.1.2 OCR Performance Results

The OCR performance investigation conducted in this study utilized the approach and data outlined in Section 3.5. The evaluation task was executed on a laptop computer equipped with an AMD Ryzen 5 5600H Radeon Graphics 3.30 GHz processor and 8GB of RAM, running the Windows operating system. The evaluation environment utilized the Visual Studio code editor and Chrome web browser.

The results of testing the open-source OCR engine (PaddleOCR) on a set of maps with a white background are shown in Table 3. The rows show the different font sizes (5 pt - 14 pt), and the columns show the different map image resolutions in dots per inch (20dpi – 200dpi). In the results table, the "E" in brackets signifies that the extraction process encountered minor errors, resulting in certain portions of the provided texts being left undetected. The results suggest that recognizing very small text (5 pt) on maps is challenging for OCR, regardless of image resolution. Even on high resolution maps, small text remains difficult to recognize. The OCR performance decreases as map image resolution increases for certain font sizes. This could be due to factors such as increased noise and unnecessary details in high-resolution images, which limit the OCR accuracy. High image resolution also leads to longer processing times and high resource usage, likely influenced by the OCR model's training data (Borisyuk et al., 2018). Most importantly, the OCR engine's maximum long size constraint of 960 pixels affects performance. When images exceed this limit, resizing them reduces OCR performance by shrinking text elements. Moreover, it is important to note that having a large font size does not ensure accurate results at low image resolution. Factors such as the loss of text details and the shrinking of character edges at lower resolutions can contribute to potential inaccuracies in the results.

Table 2. Number of detected texts on maps with white background, (E) denotes detection with error.

	20dpi	30dpi	40dpi	50dpi	60dpi	70dpi	80dpi	90dpi	100dpi	150dpi	200dpi
5pt	0	0	0	0	0	0	0	0	0	0	0
6pt	0	0	0	0	5 (E)	4 (E)	4 (E)	15	16	11(E)	15
7pt	0	0	0	0	10 (E)	7 (E)	16	15	19	15	15
8pt	0	0	0	6 (E)	8 (E)	15	19	20	21	20	21
9pt	0	0	6 (E)	11(E)	16	16	21	21	21	21	21
10pt	0	0	5 (E)	8 (E)	18	20	21	21	21	21	21
11pt	0	0	7 (E)	20	21	21	21	21	21	21	21
12pt	0	12 (E)	5 (E)	21	20	21	22	21	22	21	21
14pt	0	4 (E)	17 (E)	21	21	21	22	22	22	21	21

In Figure 14, which depicts line chat for the OCR performance results from Table 3, it is demonstrated that the detected number of texts is increasing as the font size of the text and the image resolution of the map increase. And it indicates that the OCR detects all the given texts in the map specifically at 80 dpi (720 x 960 pixels) image resolution for font sizes ranging from 9 pt to 14 pt. The result shows a promising trend for certain font sizes and certain resolutions to utilize the OCR technology for the purpose of map reproducibility assessment and enhance reproducible research.

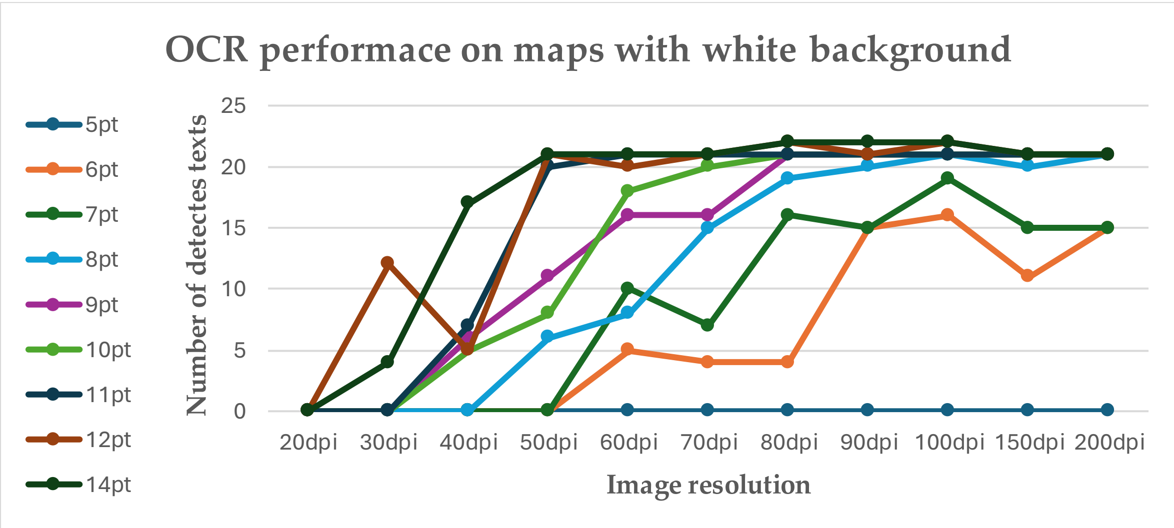

Figure 14. OCR performance results on maps with white background.

Table 4 shows a matrix of how well the OCR engine performs on the same test set of map images with a yellow background. Each row shows a different font size (5 pt to 14 pt), and each column shows a different map image resolution in dots per inch (dpi).

The result showed that, compared to the white background, texts with a small font size (5 pt) are detected under certain map image resolutions (80 dpi – 200 dpi). This shows that yellow background images increase the OCR performance due to the highest contrast ratio of black texts on yellow background (Deque Systems,2024), which is also less reflective and provides a more uniform surface for capturing small text details without inference from reflections.

Table 3. Number of detected texts on maps with yellow background. (E) denotes detection with error.

	20dpi	**30dpi**	**40dpi**	**50dpi**	**60dpi**	**70dpi**	**80dpi**	**90dpi**	**100dpi**	**150dpi**	**200dpi**
5pt	0	0	0	0	0	0	8 (E)	4 (E)	1 (E)	2 (E)	1 (E)
6pt	0	0	0	0	8(E)	6 (E)	2 (E)	15	16	8	15
7pt	0	0	0	0	11 (E)	9 (E)	14	17	20	15	16
8pt	0	0	0	3 (E)	13	14	19	20	20	19	19
9pt	0	0	5 (E)	13	15	15	20	20	20	20	20
10pt	0	0	7 (E)	9 (E)	18	21	21	21	21	21	20
11pt	0	0	7 (E)	18	20	21	21	21	21	20	20
12pt	0	11 (E)	8 (E)	21	19	21	21	21	22	21	20
14pt	0	16 (E)	20	19	20	22	22	22	22	22	22

Figure 15 presents a line chart showing the performance of the integrated OCR engine on maps with a yellow background, as outlined in Table 4. Results show that as image resolution and font size are increased, the OCR ability to recognize texts also increases. Specifically, with a resolution of 70 dpi and a font size of 14 pt, the OCR successfully recognizes all the texts on the map. Conversely, the OCR ability to recognize texts with smaller fonts decreases even if the map image resolution increases. This is due to noise and unnecessary details created in the texts that come from the high image resolution. Nevertheless, the result demonstrates a promising insight on certain image resolution and font sizes for effective utilization of OCR technology on map reproducibility assessment tasks. And it's significant in the context of the thesis, which focuses on the evaluation of digital contemporary maps.

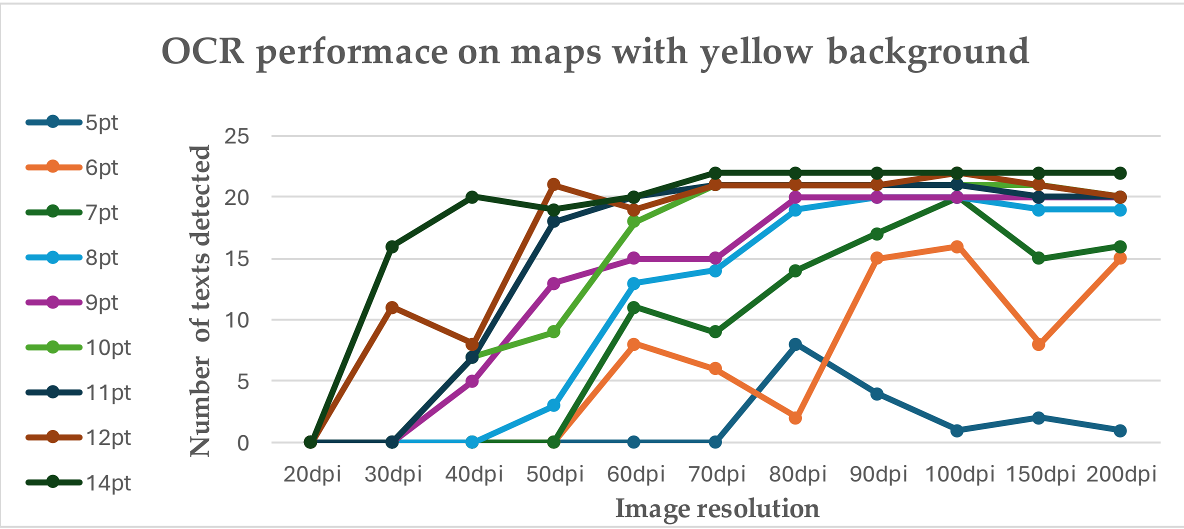

Figure 15. Performance of OCR on maps with yellow background.

5.2 Evaluation

This section outlines the objectives of the thesis for the user evaluation process of the developed web application. Following the completion of the system implementation phase, a user evaluation process is performed by deploying the application on a cloud web hosting service called Render[12] by subscribing to the standard plan that offers 1 CPU and 2 GB (Gigabyte) of RAM. The primary aim of this experiment is to comprehensively assess the application performance and functionality. Additionally, it seeks to gather valuable feedback from the users, which will significantly contribute to the overall evaluation of the application.

In this study, the chosen evaluation methods are designed to provide comprehensive insights into the usability and usefulness of the developed web application. The utilization of the System Usability Scale (SUS) method by Brooke (1995) offers several advantages. Firstly, the SUS is a widely recognized and validated tool for assessing usability, making it suitable for ensuring the robustness of the evaluation process. By employing a standardized instrument, it can benchmark the application's usability against established norms and industry standards, providing valuable context for interpretation. Moreover, the custom questionnaire created to evaluate both general and application-specific components' usefulness of the application allows for a more detailed understanding of user perceptions. By customizing questions to the specific context of the application, it can gather targeted feedback on features and functionalities, enabling the identification of areas for improvement more effectively. Additionally, the inclusion of a custom question to rank defined criteria in order of their importance enhances the depth of the evaluation and it can prioritize development efforts and focus on aspects that matter most to the assessment process.

Overall, there were a total of 15 responses from the survey participants. To get the basic insights of the evaluators, four basic questions are prepared and asked to the participants at the beginning of the evaluation. According to the survey, 80% of the participants are GIS professionals, and the rest are interdisciplinary researchers and students. In terms of their computer skills, 66.7 % of the participants have advanced computer skills, and the rest define them as intermediate skills. Participants also responded to the knowledge of OCR technology; from their responses, 66.7 % of the participants have never used OCR and 26.7 % of the participants have used OCR before, and the rest of the participants answered this question as may be. In the end, participants answered about the concept of reproducible research and particularly map reproducibility; only 26.7% of the participants are very familiar with the concept, 46.7% of the participants are somewhat familiar with it, and 26.7 % of the participants are not familiar with the concept at all.

[12] https://render.com/

5.2.1 Usability Evaluation

The primary objective of usability evaluation in this study is to assess the ease of use and adaptability of the developed system. Additionally, it is crucial to measure user satisfaction to determine if the system effectively meets the needs and expectations of its users. In addition, as per ISO 9241-11, usability refers to the extent to which an application allows targeted users to achieve desired goals. This includes the ability to complete tasks, performance, learning time, user reactions post-task completion, and the resources consumed during task execution (Bevan et al., 2016). To achieve this, the system usability scale (SUS) standard system assessment tool is utilized to evaluate the usability of the developed system. The SUS method contains 10 questions on which users indicate their level of agreement or disagreement (Brooke, 1995). Of the questions, half (even numbers) represent negative statements, and the rest (odd numbers) represent positive statements. The statement has a scale range of 1 (strongly disagree) to 5 (strongly agree). The result is a single number that represents a combined measure of the system's general usability.

The system usability scale (SUS) is a single result representation that wraps the overall usability of the application, and it leverages the Likertscale[13]. Its simplicity and reliability make it a widely adopted method for usability assessment across diverse domains (Brooke, 1995). To calculate this score, the following methodology is implemented:

- Sum the scores assigned to each statement, where each score is within the range of 0 to 4.
- For odd number questions (1, 3, 5, 7, and 9), the score value is derived by subtracting 1 from the scale position.
- For even number questions (2, 4, 6, 8, and 10), the score value is determined by subtracting the scale position from 5.
- The sum of the scores is multiplied by 2.5 to get the SUS overall score, which results within the range of 0 to 100 (Brooke, 1995).

Here is the general equation to calculate the SUS value where Each Q1, Q2, ..., Q10 represents a user's response to a specific question in the SUS questionnaire.

$$\textit{SUS score} = ((Q1-1) + (5-Q2) + (Q3-1) + (5-Q4) + (1-Q5) + (5-Q6) + (Q7-1) + (5-Q8) + (Q9-1) + (5-Q10)) * 2.5$$

The system usability scale (SUS) has been integrated into the general survey for the assessment of map reproducibility. The responses from the participants are collected, and the SUS score method is applied to calculate the SUS score for each of the 15 responses. The calculated general average SUS score was 69.33%. According to the comparison SUS score (Figure 16), the computed SUS score lies above average which is usable with some minor improvements.

[13] https://www.simplypsychology.org/likert-scale.html

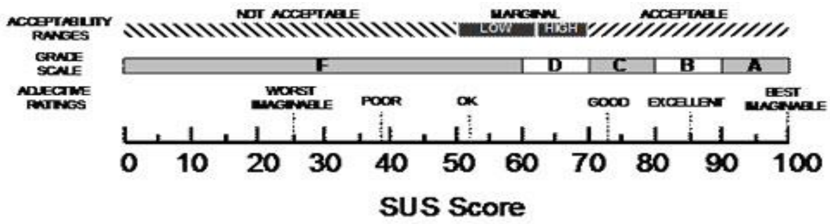

Figure 16. A comparison of SUS core adapted from (Bangor et al., 2008).

The SUS score only provides the general usability of the system, but it will not give insight into the effectiveness and usefulness of the application. To address this challenge, the general and specific application component usefulness of the application are essential.

5.2.2 Usefulness Evaluation

In this study, a custom set of nine questions has been prepared to explore the usefulness of the application. Five of these questions specifically focus on the general usefulness of the application, while the other four questions aim to get insights about the specific application component usefulness. These questions utilize a scale ranging from 1 to 10 of answering possibilities to get more insights about the distribution of the scales, where 1 denotes "not useful at all" and 10 represents "extremely useful", enabling to capture a wide range of perspectives and experiences from the participants.

Figure 17 illustrates the participants' evaluation of the general usefulness of the system. The legend in the chart represents the questions asked of the participants related to the usefulness of the application, and the horizontal axis describes the rating scale (1 - 10) for each question. The data labels inside different colours of the bar represent the percentage rate of questions corresponding to the given scale. The integration of additional important features in the application received a rating of 1 (lowest score) from 13% of participants, while 33% and 13% of participants rated the highest scores (9 and 10) respectively. The efficiency of the web-based application in text extraction, parsing, and comparison compared to manually transcribing the texts was rated 9 and 10 (highest score) by 53% and 20% of the participants respectively. This suggests that a majority found the automated process to be highly effective compared to manual transcription. And it was rated 6 (lowest score) by 7% of the participants, indicating some room for improving efficiency. The map text comparison technique integrated into the web application makes the comparison easier than visually comparing the maps rated 4 (lowest value) by 7% of the participants and 9 and 10 (highest score) by 27% and 33 % of the participants respectively. This demonstrates that this technique offers a significant role in assessing and comparing the textual elements. In terms of criteria selection, the comparative assessment workflow defined a fitting criterion for facilitating the assessment of map reproducibility was rated 4(lowest score) by 13 % of the participants and 9 and 10 (highest score) by 27% and 13% of the participants respectively. This shows that the majority found the defined criteria's perform well in facilitating the assessment process. Confidence in utilizing the web-based application for map reproduction assessment process tasks were rated 4 (lowest value) by 7% of the participants and 9 and 10 (highest scores) by 27% and 20% of the participants respectively. These demonstrate significant trust in the web-based tool for map reproduction assessment.

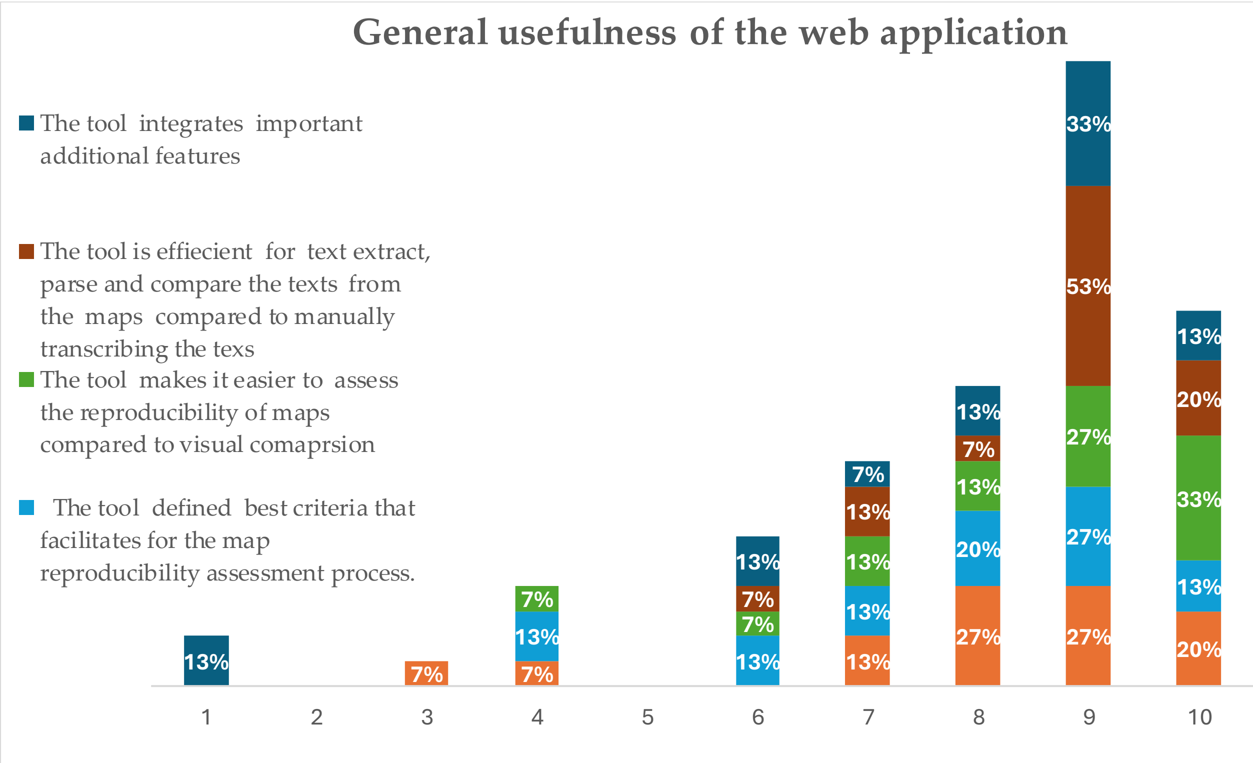

Figure 17. General usefulness of the web-based application.

Figure 18, illustrating a bar chart about the usefulness of specific application components. The survey revealed that the reproducibility assessment table, coupled with the interactive nature of the tool in enhancing the assessment of map reproducibility, was rated 1 (lowest score) by 7% of the participants and 10 (highest scores) by 40% of the participants, respectively. This shows that the majority of the participants found the interactivity function of the application to greatly enhance the comparison and assessment process. The side-by-side visualization of the annotated map images in the application offers additional support during the assessment process of map reproducibility were rated 10 (highest score) by 47% of the participants and 4 (lowest score) by 7% of the participants. It demonstrates that most of the participants found that the visualization of annotated maps enhanced the facilitation process. The fairness of the calculated text similarity score using the fuzzy string matching technique for the provided text similarity was rated 3 (low score) by 13% of the participants and 10 (highest score) by 13% of the participants. The criteria used in the application are sufficient for facilitating the assessment of map reproducibility were rated 3 (lowest score) by 7% of participants and 10 (highest score) by 33% of participants. This result demonstrates that the majority of participants agreed that the metrics defined in the comparative text assessment workflow are enough to assess the reproducibility of maps.

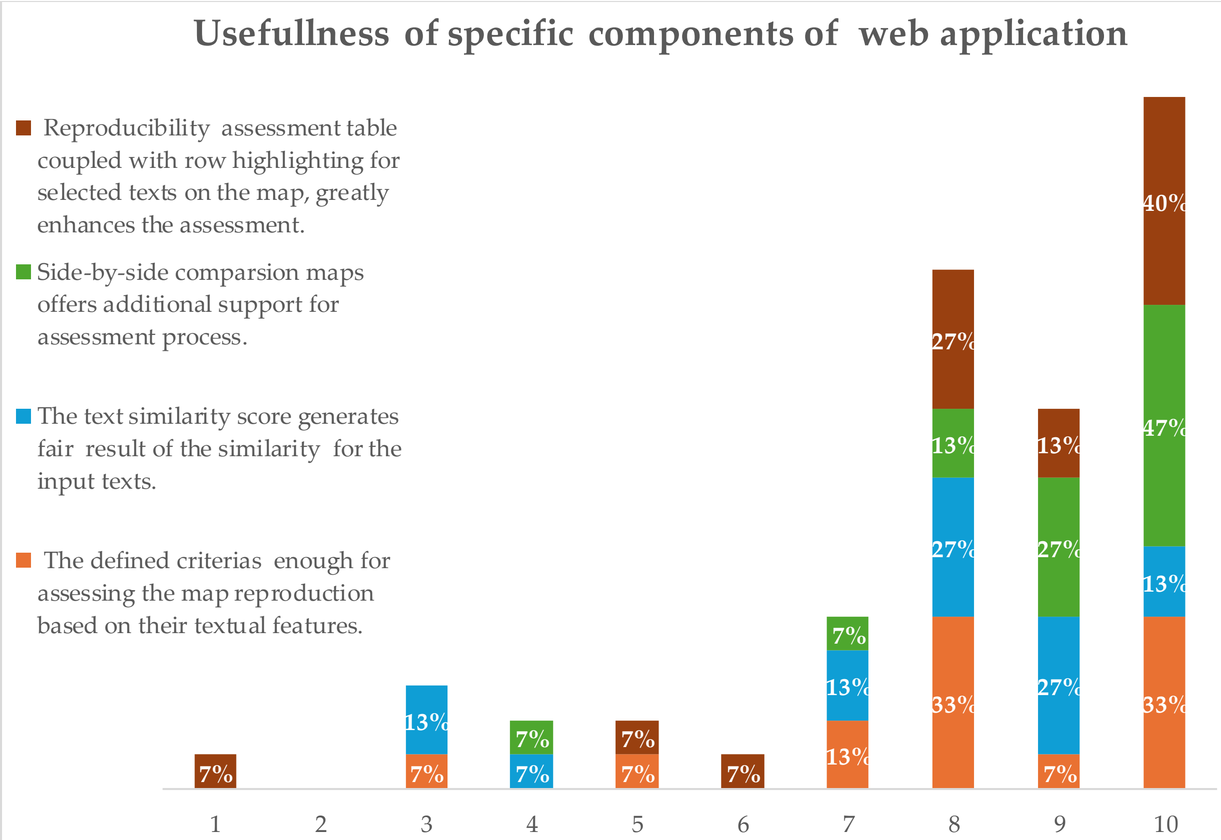

Figure 18. Usefulness of specific components of the web-based application.

The wide distribution of values across all scales makes it challenging to analyse the overall score and determine the usefulness of the web-based application. This distribution could be associated to participants' limited understanding of the reproducible research and map reproducibility assessment task, as depicted in Figure 19. When asked about their familiarity and understanding of reproducible research, particularly map reproducibility, only 26.7% of participants reported being very familiar. However, the highest percentage (46.7%) of participants indicated being somewhat familiar, while 26.7% expressed a lack of familiarity with reproducibility.

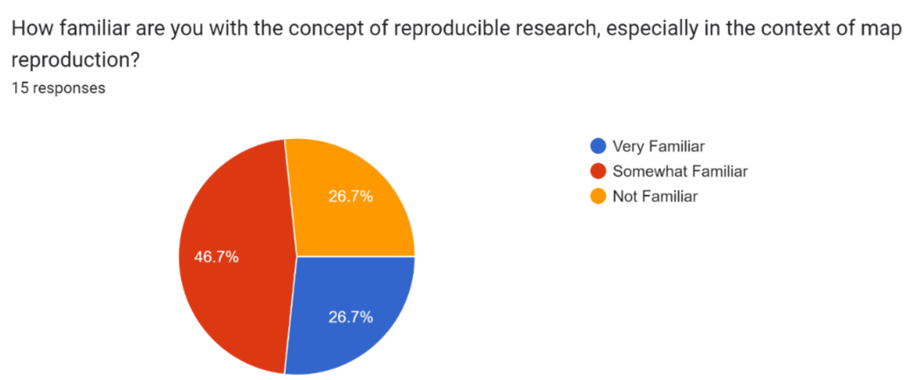

Figure 19. Reproducibility and reproducible research familiarity among the participants.

To evaluate the role of defined criteria in facilitating the assessment of map reproducibility, participants were tasked with ranking the importance of each criterion within the workflow, as illustrated in Figure 20. They assigned numerical rankings, with 1 signifying the least importance and 5 indicating the highest level of importance, to aid in assessing map reproducibility. This structured approach aimed to gather participant insights and facilitate a comprehensive understanding of the criteria's relative significance in the assessment process.

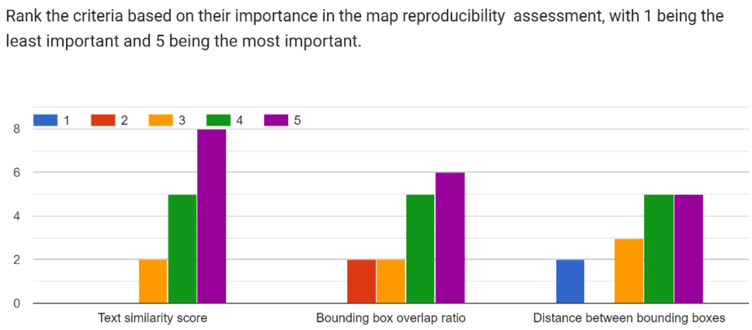

Figure 20. Criteria ranking from the participants.

Figure 20 illustrates participants' ranking for the criteria defined in the comparative workflow; the text similarity score criteria for comparing the texts in both maps were ranked 5 (the most important) by 8 participants, 4 (important) by five participants, and 3 (average) by two participants. The bounding box overlap ratio criteria were ranked 5 (the most important) by 6 participants, 4 (important) by 5 participants, 3 (moderate importance), and 2 (somewhat important) by 2 participants. The distance between bounding boxes associated with text criteria for assessing map reproducibility was ranked 5 (the most important) and 4 (important) by 5 participants, 3 (moderate importance) by 3 participants, and 1 (least important) by 2 participants. This demonstrates that while all three criteria are considered significant, the text similarity score criterion emerges as the most favored for assessing map reproducibility. However, the distance between bounding boxes criterion receives lower ratings, suggesting it may not reliably capture discrepancies, such as spacing variations between map texts in the original and reproduced maps, even if they convey the same message.

6 Discussion

This study implements a novel web-based automatic text extraction, parsing, and comparison tool that facilitates the assessment of map reproducibility based on textual elements of maps, and the tool integrates the latest open-source OCR technology for text extraction. The web application offers a user-friendly interface, efficiency, interoperability, accessibility, and an interactive nature, all of which are essential for facilitating the assessment of map reproducibility compared to the approach of assessing map reproducibility visually or using the pixel-by-pixel image-based comparison used by Koukouraki & Kray (2023) and assessing the map reproduction using the spectrum of map reproducibility assessment (Giraud & Lambert, 2017). The results also demonstrate that the comparison and assessment of map reproducibility could be carried out with less effort while saving time.

The comparative assessment approach of the study employs bounding boxes for accurately identifying text elements within the map images, providing a structured framework for analysis. To accommodate maps with varying dimensions, a normalization process is applied to the bounding boxes. This normalization process standardizes the spatial representation of text elements, ensuring consistency and facilitating accurate comparison across different maps. Moreover, the integration of confidence scores further enhances the accuracy assessment of the extracted texts. By assigning confidence scores to each extracted text element, the study evaluates the reliability and accuracy of the extraction process. This approach offers valuable insights into the confidence levels associated with the extracted texts and is used for adjusting the text similarity score for texts that do not exactly match.

The study examined the performance of the OCR engine across maps with different backgrounds, font sizes, and image resolutions. The results demonstrate that even with high-resolution map images (200 dpi), the OCR engine cannot detect small font size texts (5 pt). This is due to the limitations associated with the selected OCR engine having a maximum long size of 960 pixels. Input map images that exceed this size are resized to 960 pixels, and the text inside the map image shrinks, reducing the performance of the OCR engine.

The application, deployed on Render cloud server utilizing the standard plan with 2 GB of RAM and 1 CPU, demonstrated promising results in terms of technical performance. Leveraging the allocated resources, it consistently showcased high accuracy in text extraction and parsing, with minimal computational overhead and efficient processing times, particularly with input map images under 5 MB in size. This underscores the importance of resource allocation in optimizing the tool's performance. Significantly, its scalability and adaptability were evident in effectively managing diverse map characteristics, including varying resolutions, font sizes, and file types. These results show that the application can facilitate map text extraction and assessment with appropriate server resource allocation.

Lastly, user evaluation is conducted, and it is found that the application is usable with an average SUS score of 69.33%, which is above average according to the comparison of the SUS score by Bangor et al. (2008) with need of some improvements. And the survey reveals that the application is useful for performing map reproducibility assessment tasks with some variations on the distribution of the scales. This is due to the low understanding of reproducibility and reproducible research among the survey participants.

6.1 Limitations

This work has several limitations. Firstly, the comparative workflow designed in this study is limited to only three criteria for the final decision on the matching status of the texts. The minimum thresholds for those criteria are heuristically developed and tested on a small dataset of maps. Additionally, limitations arise from the use of the Euclidean distance criterion for assessing the spatial variation of texts. Whenever texts on the map exhibit significant spacing, the calculated distance may be stretched, leading to a large distance, and this phenomenon challenges the robustness of the defined threshold for distance. Moreover, the input maps are limited to modern digital maps, and the language of texts are limited to only maps that are written in English characters. Additionally, due to potential conflicts with the OpenCV library during application development, the supported input file types are limited to four specific formats.

When dealing with numerical texts, limitations arise in implementing the text similarity score. The fuzzy string matching technique determines similarity based on character-level comparison, ignoring the numerical context. Consequently, it can yield incorrect similarity scores; for instance, it may rate "1.00" and "100" as highly similar. In addition, the OCR engine has a maximum long side size of 960 pixels, and the angle of text rotation that can be detected accurately by this OCR engine is limited to 0 and 180 degrees.

6.2 Future Work

Given the limitations discussed above, there is a lot to improve in future research. This involves enhancing the calculation of a text similarity score by considering the contextual values of the input texts in addition to the character level similarity for numerical input texts. Additionally, future work is to incorporate topological text comparison to optimize the evaluation of spatial text variations in maps through leveraging the arrangement and connectivity of text elements, enhancing Euclidean distance implemented on this thesis by offering detailed understanding of the structural features of textual components. In addition, conducting a survey among focus groups to ascertain the accepted values for each selected criteria and test on large dataset of maps is also future work.

Furthermore, future work involves customizing the web application to integrate multiple languages for text extraction from diverse map images. This enhancement would contribute for the map reproducibility assessment to span across multiple geosciences research published in different language other than English.

7 Conclusion

In conclusion, the practice of reproducibility is paramount in scientific research, particularly in the field of geoscience (Konkol et al., 2019). Embracing reproducibility and adopting reproducible research practices in geoscience publications are essential steps towards ensuring transparency and validation of scientific findings and results (Nüst et al., 2018). Maps, as visual representations of complex geoscientific data and results, play a pivotal role in this process. By enhancing map reproducibility, it can further promote reproducible research in the field (Konkol & Kray, 2019). This work presents a systematic tool that facilitates the assessment of map reproducibility based on the textual features of a map by utilizing OCR technology for text extraction purposes. The study focuses on textual elements, as they are an integral part of a map and serve as the primary means of communication and interpretation. The study contributes a text comparison workflow consisting of text similarity scores between texts using the fuzzy string-matching technique, bounding box overlap ratios through the integration of the Jaccard index or intersection over union algorithm, and spatial variations between the bounding boxes associated with the texts using the computation of Euclidean distance, and it provides researchers and GIS professionals with detailed map reproducibility assessment results.

The application is deployed on a cloud server to evaluate its effectiveness in facilitating the assessment of map reproducibility. The evaluation included usability and usefulness testing, incorporating the standard System Usability Scale (SUS) method and a custom questionnaire. The results indicate that while the application is usable, with an average SUS score of 69.33%, there is room for improvement. Furthermore, participants found the tool useful in achieving the aim outlined in this thesis. The logical text comparison workflow defines minimum thresholds for each criteria: a text similarity score of 70%, a Euclidean distance of 0.05 pixel units, and a final score of 50%. These defined cutoff values accommodate a robust comparison and assessment process. However, the fuzzy string matching technique used to calculate the text similarity score may yield biased results, especially when comparing numerical input texts. As a result, it's essential to conduct a comprehensive examination of the results prior to applying them in the actual map reproducibility assessment process.

Still, the study demonstrates that the approach of using textual elements for map reproducibility assessment plays an important role as it is independent of file type, size, and image resolution between the input map images. However, it's important to acknowledge that variations in image resolution, particularly when one input image has significantly lower or higher resolution than the other, may slightly affect the results. Considering all the different aspects of map reproducibility assessment, it is essential to focus on comparison methods between map textual elements; the comparative assessment of textual elements in this experimental study consequently reveals a possible approach that could facilitate the assessment of map reproducibility by focusing on the textual elements of the map.

Bibliography

Agarwal, A. (2023). IOU score and its variants for deep learning. Retrieved from https://medium.com/@ankuraga/iou-score-and-its-variants-for-deep-learningbd09ecf14832.

Annanias, Y., Wiegreffe, D., Niekler, A., Kuźma, M., & Harvey, F. (2023). *Development of a Semantic Segmentation Approach to Old-Map Comparison*. *277*, 14:1-14:0. https://doi.org/10.4230/LIPIcs.GIScience.2023.14

Axis Maps. (2020). Labeling and text hierarchy in cartography. https://www.axismaps.com/guide/labeling.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Barba, L. A. (2018). *Terminologies for Reproducible Research*. http://arxiv.org/abs/1802.03311

Barget, M. (2022, February 1). *Reading historical maps with optical character recognition (OCR)*. INSULAE. https://insulae.hypotheses.org/485

Bevan, N., Carter, J., Earthy, J., Geis, T., & Harker, S. (2016). *New ISO Standards for Usability, Usability Reports and Usability Measures* (pp. 268–278). https://doi.org/10.1007/978-3-319-39510-4_25

Bill, R., Blankenbach, J., Breunig, M., Haunert, J.-H., Heipke, C., Herle, S., Maas, H.-G., Mayer, H., Meng, L., Rottensteiner, F., Schiewe, J., Sester, M., Sörgel, U., & Werner, M. (2022). Geospatial Information Research: State of the Art, Case Studies and Future Perspectives. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, *90*(4), 349–389. https://doi.org/10.1007/s41064-022-00217-9

Borisyuk, F., Gordo, A., & Sivakumar, V. (2018). Rosetta. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. https://doi.org/10.1145/3219819.3219861

Brisinello, M., Grbic, R., Pul, M., & Andelic, T. (2017). Improving optical character recognition performance for low quality images. *2017 International Symposium ELMAR*, 167–171. https://doi.org/10.23919/ELMAR.2017.8124460

Brooke, J. (1995). *SUS: A quick and dirty usability scale*. https://www.researchgate.net/publication/228593520

Cao, R., & Lim Tan, C. (2002). *Separation of Overlapping Text from Graphics*. Proceedings of Sixth International Conference on Document Analysis and Recognition. https://doi.org/10.1109/icdar.2001.953752

Chaudhuri Arindam and Mandaviya, K. and B. P. and G. S. K. (2017). Optical Character Recognition Systems. In *Optical Character Recognition Systems for Different Languages with*

Soft Computing (pp. 9–41). Springer International Publishing. https://doi.org/10.1007/978-3-319-50252-6_2

Chiang, Y. Y. (2017). Unlocking textual content from historical maps - Potentials and applications, trends, and outlooks. *Communications in Computer and Information Science*, *709*, 111–124. https://doi.org/10.1007/978-981-10-4859-3_11

Chiang, Y. Y., & Knoblock, C. A. (2010). An approach for recognizing text labels in raster maps. *Proceedings - International Conference on Pattern Recognition*, 3199–3202. https://doi.org/10.1109/ICPR.2010.783

Chiang, Y. Y., & Knoblock, C. A. (2011). Recognition of multi-oriented, multi-sized, and curved text. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 1399–1403. https://doi.org/10.1109/ICDAR.2011.281

Chiang, Y. Y., & Knoblock, C. A. (2013). A general approach for extracting road vector data from raster maps. *International Journal on Document Analysis and Recognition*, *16*(1), 55–81. https://doi.org/10.1007/s10032-011-0177-1

Chiang, Y. Y., & Knoblock, C. A. (2015). Recognizing text in raster maps. *GeoInformatica*, *19*(1), 1–27. https://doi.org/10.1007/s10707-014-0203-9

Claerbout, J. F., & Karrenbach, M. (1992). Electronic documents give reproducible research a new meaning. *SEG Technical Program Expanded Abstracts 1992*, 601–604. https://doi.org/10.1190/1.1822162

Dastrup, R. A., Ma, Gisp, & Lisa C. Young, Ph. D. (2022, February 1). *4.2 Map Design and map elements*. Geographic Information Technologies. https://open.maricopa.edu/gist/chapter/4-2-map-design-and-map-elements/

De Jong, K., & Hagen-Zanker, A. (2004). *The MAP COMPARISON KIT: methods, software and applications Coupled Innovative Training Network View project TEEB Quantitative Assessment View project*. www.rivm.nl

Deque Systems. (2024). Check Text and Background for Sufficient Color Contrast. Retrieved February 28, 2024, from https://dequeuniversity.com/tips/color-contrast.

Donoho, D. L. (2010). An invitation to reproducible computational research. *Biostatistics*, *11*(3), 385–388. https://doi.org/10.1093/biostatistics/kxq028

Du, Y., Chen, Z., Jia, C., Yin, X., Zheng, T., Li, C., Du, Y., & Jiang, Y.-G. (2022). *SVTR: Scene Text Recognition with a Single Visual Model*. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. https://doi.org/10.24963/ijcai.2022/124

Du, Y., Li, C., Guo, R., Yin, X., Liu, W., Zhou, J., Bai, Y., Yu, Z., Yang, Y., Dang, Q., & Wang, H. (2020). *PP-OCR: A Practical Ultra Lightweight OCR System*. http://arxiv.org/abs/2009.09941

Dvorský, J., Snášel, V., & Voženílek, V. (2010). On maps comparison methods. *2010 International Conference on Computer Information Systems and Industrial Management Applications, CISIM 2010*, 557–562. https://doi.org/10.1109/CISIM.2010.5643513

Euclidean distance. (2024, January 17). Wikipedia. https://en.wikipedia.org/wiki/Euclidean_distance

Editorial. (2016). Reality check on reproducibility. *Nature*, *533*(7604), 437–437. https://doi.org/10.1038/533437a

Elzakker, V. (2004). *The use of maps in the exploration of geographic data*. Koninklijk Nederlands Aardrijkskundig Genootschap.

Feger, S. S., & Woźniak, P. W. (2022). Reproducibility: A Researcher-Centered Definition. *Multimodal Technologies and Interaction*, *6*(2). https://doi.org/10.3390/mti6020017

Foody, G. M. (2007). Map comparison in GIS. *Progress in Physical Geography*, *31*(4), 439–445. https://doi.org/10.1177/0309133307081294

Giraud, T., & Lambert, N. (2017). Reproducible cartography. *Lecture Notes in Geoinformation and Cartography*, 173–183. https://doi.org/10.1007/978-3-319-57336-6_13

Guan, L., Jia, L., Xie, Z., & Yin, C. (2022). A Lightweight Framework for Obstacle Detection in the Railway Image Based on Fast Region Proposal and Improved YOLO-Tiny Network. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–16. https://doi.org/10.1109/TIM.2022.3150584

Guo, L., Li, Z., & Sun, W. (2015). *Association for Information Systems AIS Electronic Library (AISeL) Understanding Travel Destinations from Structured Tourism Blogs Recommended Citation "Understanding Travel Destinations from Structured Tourism Blogs" Understanding Travel Destinations from Structured Tourism Blogs* (Vol. 80). http://aisel.aisnet.org/whiceb2015/80

Hagen, A. (2003). Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science*, *17*(3), 235–249. https://doi.org/10.1080/13658810210157822

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). *Searching for MobileNetV3*. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). https://doi.org/10.1109/iccv.2019.00140

Jobst, M., & Gartner, G. (2019). Map Production Supporting Activities of Digitalization and AI. *Proceedings of the ICA*, *2*, 1–5. https://doi.org/10.5194/ica-proc-2-53-2019

Kedron, P., Li, W., Fotheringham, S., & Goodchild, M. (2021). Reproducibility and replicability: opportunities and challenges for geospatial research. In *International Journal of Geographical Information Science* (Vol. 35, Issue 3, pp. 427–445). Taylor and Francis Ltd. https://doi.org/10.1080/13658816.2020.1802032

Konkol, M., & Kray, C. (2019). In-depth examination of spatiotemporal figures in open reproducible research. *Cartography and Geographic Information Science*, *46*(5), 412–427. https://doi.org/10.1080/15230406.2018.1512421

Konkol, M., Kray, C., & Pfeiffer, M. (2019). Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study. *International Journal of Geographical Information Science*, *33*(2), 408–429. https://doi.org/10.1080/13658816.2018.1508687

Koukouraki, E. (2023, September 7). Reproducibility review of: Extreme heat alerts and impacts across Mozambique 2016 - 2022: gathering evidence from media articles. OSF. https://doi.org/10.17605/OSF.IO/EU8KW.

Koukouraki, E., & Kray, C. (2023). *Map Reproducibility in Geoscientific Publications: An Exploratory Study*. https://doi.org/10.4230/LIPIcs.GIScience.2023.6

Kraak, M. J., & Fabrikant, S. I. (2017). Of maps, cartography and the geography of the International Cartographic Association. *International Journal of Cartography*, *3*(sup1), 9–31. https://doi.org/10.1080/23729333.2017.1288535

Kysela, J. (2018). *A Comparison of Text String Similarity Algorithms for POI Name Harmonisation* (pp. 121–130). https://doi.org/10.1007/978-3-319-94544-6_12

Lan, T., & Longley, P. (2019). Geo-referencing and mapping 1901 census addresses for England and Wales. *ISPRS International Journal of Geo-Information*, *8*(8). https://doi.org/10.3390/ijgi8080320

Levenshtein, V. I., & others. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Levenshtein distance. (2023, December 23). Wikipedia.https://en.wikipedia.org/wiki/Levenshtein_distance

Li, C., Liu, W., Guo, R., Yin, X., Jiang, K., Du, Y., Du, Y., Zhu, L., Lai, B., Hu, X., Yu, D., & Ma, Y. (2022). *PP-OCRv3: More Attempts for the Improvement of Ultra Lightweight OCR System*. http://arxiv.org/abs/2206.03001

Li, H., Liu, J., & Zhou, X. (2018). Intelligent Map Reader: A Framework for Topographic Map Understanding with Deep Learning and Gazetteer. *IEEE Access*, *6*, 25363–25376. https://doi.org/10.1109/ACCESS.2018.2823501

Li, Z., Chiang, Y. Y., Tavakkol, S., Shbita, B., Uhl, J. H., Leyk, S., & Knoblock, C. A. (2020). An Automatic Approach for Generating Rich, Linked Geo-Metadata from Historical Map Images. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 3290–3298. https://doi.org/10.1145/3394486.3403381

Liao, M., Wan, Z., Yao, C., Chen, K., & Bai, X. (2019). *Real-time Scene Text Detection with Differentiable Binarization*. Proceedings of the AAAI Conference on Artificial Intelligence, 34(07), 11474–11481. https://doi.org/10.1609/aaai.v34i07.6812

Lo, S. W., & Chou, H. M. (2022). Evaluating and Improving Optical Character Recognition (OCR) Efficiency in Recognizing Mandarin Phrases with Phonetic Symbols. *Proceedings of the 2022 IEEE International Conference on Internet of Things and Intelligence Systems, IoTaIS 2022*, 390–394. https://doi.org/10.1109/IoTaIS56727.2022.9975969

Lokeshkumar, R., Sabnis, O. V., & Bhattacharyya, S. (2020). A Novel Approach to Extract and Analyse Trending Cuisines on Social Media. In *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 49, pp. 645–656). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-43192-1_72

Loran, C., Haegi, S., & Ginzler, C. (2018). Comparing historical and contemporary maps - a methodological framework for a cartographic map comparison applied to Swiss maps. *International Journal of Geographical Information Science*, *32*(11), 2123–2139. https://doi.org/10.1080/13658816.2018.1482553

Mageshwaran R, M. (2021a, July 11). *Review of best open-source OCR Tools*. Medium. https://medium.com/technovators/review-of-best-open-source-ocr-tools-fc839a20e61f.

Mello, C. A. B., & Lins, R. D. (1999). *A Comparative Study on OCR Tools*. https://api.semanticscholar.org/CorpusID:10107804

Menno-Jan Kraak. (2019). 22-ica_strategic_plan_2019-2027. *Https://Icaci.Org/Files/Documents/Generalassembly2019/22-Ica_strategic_plan_2019-2027.Pdf*.

Mittal, R., & Garg, A. (2020). Text extraction using OCR: A Systematic Review. *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 357–362. https://doi.org/10.1109/ICIRCA48905.2020.9183326

Mocnik, F. B. (2023). Why we can read maps. *Cartography and Geographic Information Science*, *50*(1), 1–19. https://doi.org/10.1080/15230406.2022.2127911

Nazari, N. H., Tan, T., & Chiang, Y.-Y. (2016). Integrating Text Recognition for Overlapping Text Detection in Maps. *Electronic Imaging*, *28*(17), 1–8. https://doi.org/10.2352/ISSN.2470-1173.2016.17.DRR-061

Neudecker, C., Baierer, K., Gerber, M., Clausner, C., Antonacopoulos, A., & Pletschacher, S. (2021). A survey of OCR evaluation tools and metrics. *The 6th International Workshop on Historical Document Imaging and Processing*, 13–18. https://doi.org/10.1145/3476887.3476888

Nguyen, T. T. H., Jatowt, A., Coustaty, M., & Doucet, A. (2022). Survey of Post-OCR Processing Approaches. *ACM Computing Surveys*, *54*(6), 1–37. https://doi.org/10.1145/3453476

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018). Reproducible research and GIScience: an evaluation using AGILE conference papers. *PeerJ*, *6*, e5072. https://doi.org/10.7717/peerj.5072

Nüst, D., Konkol, M., Schutzeichel, M., Pebesma, E., Kray, C., Przibytzin, H., & Lorenz, J. (2017). Opening the publication process with executable research compendia. *D-Lib Magazine*, *23*(1–2). https://doi.org/10.1045/january2017-nuest

Nüst, D., & Pebesma, E. (2021). Practical Reproducibility in Geography and Geosciences. *Annals of the American Association of Geographers*, *111*(5), 1300–1310. https://doi.org/10.1080/24694452.2020.1806028

Ogwok, D., & Ehlers, E. M. (2022). Jaccard Index in Ensemble Image Segmentation: An Approach. *ACM International Conference Proceeding Series*, 9–14. https://doi.org/10.1145/3581792.3581794

Pebesma, E., Nüst, D., & Bivand, R. (2012). The R software environment in reproducible geoscientific research. *Eos*, *93*(16), 163. https://doi.org/10.1029/2012EO160003

Peng, R. D. (2011). Reproducible research in computational science. In *Science* (Vol. 334, Issue 6060, pp. 1226–1227). American Association for the Advancement of Science. https://doi.org/10.1126/science.1213847

Pereira Marghidan, C., van Aalst, M., Blanford, J., Maure, G., & Marrufo, T. (2023). Extreme heat alerts and impacts across Mozambique 2016 – 2022: gathering evidence from media articles. *AGILE: GIScience Series*, *4*, 1–13. https://doi.org/10.5194/agile-giss-4-11-2023

Pouderoux, J., Gonzato, J.-C., Pereira, A., & Guitton, P. (2007). *Toponym Recognition in Scanned Color Topographic Maps*. Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). https://doi.org/10.1109/icdar.2007.4378766

Pun-Cheng, L. S. C. (2000). An Orthodox Approach towards Automated Bilingual Name Placement in Large Scale Maps. In *Journal of Geospatial Engineering* (Vol. 2, Issue 2).

Rawat, P., & Mahajan, A. N. (2020). ReactJS: A Modern Web Development Framework. In *International Journal of Innovative Science and Research Technology* (Vol. 5, Issue 11). www.ijisrt.com

Relan, K. (2019). Beginning with flask. Building REST APIs with Flask, 1–26. https://doi.org/10.1007/978-1-4842-5022-8_1

Sabu, A. M., & Das, A. S. (2018). A Survey on various Optical Character Recognition Techniques. *2018 Conference on Emerging Devices and Smart Systems (ICEDSS)*, 152–155. https://doi.org/10.1109/ICEDSS.2018.8544323

Seatgeek. (2021). Seatgeek/thefuzz: Fuzzy string matching in Python. GitHub. https://github.com/seatgeek/thefuzz.

Schlegel, I. (2021). Automated Extraction of Labels from Large-Scale Historical Maps. *AGILE: GIScience Series*, *2*, 1–14. https://doi.org/10.5194/agile-giss-2-12-2021

Singh, A. J. (2022). File system vs DBMS - what's the difference? Retrieved from https://www.scaler.com/topics/file-system-vs-dbms/.

Smith, A. K., & Dragićević, S. (2022). Map Comparison Methods for Three-Dimensional Space and Time Voxel Data. *Geographical Analysis*, *54*(1), 149–172. https://doi.org/10.1111/gean.12279

TAYLOR, D. R. F. (1991). *Geographic Information Systems: The Microcomputer and Modern Cartography* (pp. 1–20). https://doi.org/10.1016/B978-0-08-040277-2.50009-X

Verma, V., & Aggarwal, R. K. (2020). A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective. *Social Network Analysis and Mining*, *10*(1). https://doi.org/10.1007/s13278-020-00660-9

Vukatana, K. (2022). OCR and Levenshtein distance as a measure of image quality accuracy for identification documents. *International Conference on Electrical, Computer, and Energy Technologies, ICECET 2022*. https://doi.org/10.1109/ICECET55527.2022.9872824

Zhang, H., Zhao, K., Song, Y.-Z., & Guo, J. (2013). Text extraction from natural scene image: A survey. *Neurocomputing*, *122*, 310–323. https://doi.org/10.1016/j.neucom.2013.05.037

Zohar, M. (2021). Geolocating tweets via spatial inspection of information inferred from tweet meta-fields. *International Journal of Applied Earth Observation and Geoinformation*, *105*. https://doi.org/10.1016/j.jag.2021.102593

Appendix A: Test dataset , survey questionnaire, and demo

- Test dataset used for investigating the performance OCR engine(PaddleOCR): https://drive.google.com/drive/folders/1nLR7T7jlSgyCBmdCsa5F_xLQyjlnq05?usp=drive_link
- Survey questions used for the evaluation of the developed tool: https://forms.gle/hEZDxnAQ7cBKb2qXA
- Application Demo: https://youtu.be/n5yW-Ewaa9s

Appendix B: Links for GitHub and Libraries

- GitHub: https://github.com/Yohannes19/Map-text-extraction-OCR
- theFuzz: https://pypi.org/project/thefuzz/
- Axios: https://axios-http.com/docs/intro
- OpenCV: https://docs.opencv.org/3.4/d6/d00/tutorial_py_root.html
- React-dropzone: https://react-dropzone.js.org/
- React-bootstrap: https://react-bootstrap.netlify.app/
- Paddle OCR: https://github.com/PaddlePaddle/PaddleOCR