iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

EXCEL SHEET COMPARISION AND RECONCILIATION SYSTEM

K.S.S. Soujanya Kumari (Assistant Professor)
Department of Information Computer Science and Systems Engineering Andhra University College of Engineering (A), Visakhapatnam, AP.

Vommi Tejasri (MCA Student)
Department of Information Technology and Computer Applications Andhra University College of Engineering (A), Visakhapatnam, AP.

Corresponding Author: Vommi Tejasri, Email: tejasrivommi@gmail.com

ABSTRACT
The increasing use of Excel spreadsheets for storing and exchanging structured information creates challenges in maintaining data consistency across multiple versions of the same dataset. Manual comparison of large Excel files is time-consuming, error-prone, and inefficient, especially when datasets contain numerous fields or undergo frequent updates. This paper presents an Excel Sheets Comparison and Reconciliation System, a web based solution that automates the identification and analysis of differences between two Excel spreadsheets.
The proposed system allows users to upload source and target Excel files and select a unique key field, such as Roll Number, Name, or Email, for record-level comparison. The system identifies matching and mismatched records, records missing from either file, and specific fields containing discrepancies. It also provides search and filtering facilities, summary information, persistent storage of previous comparison activities, and downloadable reconciliation reports. The system is implemented using Java, Spring Boot, Spring Data JPA, MySQL, Thymeleaf, and Apache POI. An attendance dataset is used as a representative example, although the system can be applied to various structured Excel datasets. The proposed solution reduces manual verification effort, improves discrepancy detection reliability, and supports efficient data reconciliation in educational, administrative, organizational, and business applications.

Keywords:
Excel Sheet Comparison, Data Reconciliation, Spreadsheet Data Processing, Data Validation, Data Comparison, Excel Automation.

INTRODUCTION
Excel spreadsheets are widely used for storing, organizing, exchanging, and analyzing structured information. Educational institutions, business organizations, financial departments, government offices, and administrative units use spreadsheets to maintain records such as student details, attendance, employee information, transactions, inventory, and examination data. Although Excel provides flexibility and ease of use, maintaining consistency between multiple versions of the same dataset remains a significant challenge. Data may be modified, copied, transferred, or updated by different users, leading to missing records, incorrect values, duplicate entries, inconsistent formats, and differences between source and target files.
Manual comparison and reconciliation of Excel files become difficult when datasets contain hundreds or thousands of records and multiple fields. Comparing spreadsheets cell by cell requires considerable time, effort, and concentration. It may also result in overlooked discrepancies and inaccurate conclusions, especially when the files are frequently updated. Identifying whether a record is present in both files, checking the values of individual fields, and locating missing records can be complicated when performed manually. Repeated reconciliation activities increase the workload of employees and delay decision-making. Therefore, an automated solution is required to compare spreadsheet datasets and present the differences in a clear and organized manner.
Previous research has addressed several problems related to spreadsheet errors, data cleaning, record linkage, schema matching, and data reconciliation. Studies on spreadsheet errors have emphasized the importance of verification and testing in spreadsheet-based applications. Data-cleaning research has focused on identifying missing, incorrect, and inconsistent values. Record-linkage techniques help identify corresponding records from different datasets, while schema-matching approaches establish relationships between fields in separate data

 © Author(s) | CC BY 4.0 International License.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

sources. These approaches provide important foundations for improving data quality and identifying records before comparison. However, many existing methods are designed for complex data-management environments or focus on a specific problem rather than providing a simple and user-friendly Excel-to-Excel reconciliation process.The research gap identified from these approaches is the lack of a practical web-based system that allows users to upload two Excel files, select a suitable comparison key, and automatically generate a structured reconciliation report. In real-world applications, users may need to compare records using a unique identifier such as Roll Number, Name, Email, Employee ID, or Product ID. After identifying corresponding records, each field must be compared to determine whether the values are equal or different. The system should also identify records that are available in one file but missing from the other. A unified solution combining record matching, field-level comparison, discrepancy identification, filtering, and report generation can reduce the difficulties associated with manual reconciliation.

The main objective of this study is to develop an Excel Sheets Comparison and Reconciliation System that automates the comparison of two Excel datasets. The proposed system allows users to upload source and target files, validate the uploaded files, select a comparison key, read and preprocess the data, and match corresponding records. It identifies matched records, mismatched records, records missing from either file, and the specific fields containing discrepancies. The system also provides search and filtering facilities, summary information, downloadable reconciliation reports, and persistent storage of previous comparison activities.

The system is implemented using Java, Spring Boot, Spring Data JPA, MySQL, Thymeleaf, and Apache POI. Apache POI is used to read and process Excel files, while Spring Boot supports the web application and its processing logic. MySQL is used to store comparison details and previous reconciliation activities. The system follows a structured workflow consisting of file upload, validation, data reading, preprocessing, record matching, field comparison, discrepancy identification, report generation, and report download. An attendance dataset is used as a representative example to demonstrate the working of the system; however, the proposed solution is not restricted to attendance-related data.

The significance of the proposed system lies in its ability to reduce manual verification effort, improve the reliability of discrepancy detection, and provide a systematic method for reconciling spreadsheet data. By presenting comparison results in an organized format, the system supports faster error identification and better data-quality management. It can be applied in educational, administrative, financial, organizational, and business environments where multiple Excel files must be compared regularly. The system can also be extended in the future to support additional file formats, automatic column mapping, duplicate-record detection, advanced matching techniques, and data visualization.

LITERATURE SURVEY

The comparison and reconciliation of data from different files is closely related to several areas of research, including data cleaning, record linkage, entity resolution, schema matching, spreadsheet auditing, and data integration. Several researchers have studied the problems of inconsistent data, duplicate records, spreadsheet errors, and matching records from different data sources. The following studies provide the background for the proposed Excel Sheet Comparison and Reconciliation System.

A. Spreadsheet Errors and Data Quality

Panko (1998) studied spreadsheet errors and highlighted the importance of checking and testing spreadsheets. Rahm and Do (2000) examine data-cleaning problems such as missing values, incorrect values, and inconsistencies. These studies establish the need for reliable data verification and quality improvement.

B. Record Linkage and Data Matching

Fellegi and Sunter (1969) introduced a framework for identifying whether records from different datasets match. Rahm and Bernstein (2001) studied schema matching between data sources. These studies are relevant to the proposed system because corresponding records must be identified before their fields can be compared.

C. Data Reconciliation

Open Refine demonstrates how interactive tools can support data cleaning and reconciliation. Recent research has also explored discrepancy detection in heterogeneous tabular datasets. These approaches are relevant to the proposed system, which focuses on comparing two Excel files and identifying differences between corresponding records.

D. Research Gap

Existing research has addressed spreadsheet errors, data cleaning, record linkage, and reconciliation. However, there is a practical need for a simple web-based system that allows users to upload two Excel files, select a comparison key, identify matched and mismatched records, detect missing records, and generate reconciliation reports. The proposed system addresses this gap through automated Excel comparison and organized result

 © Author(s) | CC BY 4.0 International License.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

reporting.

OBJECTIVES

1) To develop a system for automatically comparing two Excel sheets.
2) To identify matched, mismatched, and missing records.
3) To detect field-level differences between corresponding records.
4) To reduce the time and manual effort required for Excel data comparison.
5) To improve the accuracy and reliability of the reconciliation process.
6) To provide searchable and organized comparison results.
7) To generate downloadable reconciliation reports for further use.
8) To maintain a history of previous comparison activities.

METHODOLOGY

The methodology of the proposed Excel Sheet Comparison and Reconciliation System is designed to make the process of comparing two Excel files simple, systematic, and less time-consuming. The main idea of the system is to take two Excel sheets as input, identify the corresponding records using a selected key, compare their values, identify the differences, and present the results in an understandable format. The complete process is performed through a web-based application.

A. System Input

The process starts with the user providing two Excel files that need to be compared. The first file can be considered as the original or reference file, while the second file can be considered as the file that needs to be checked against the first file. The files are uploaded through the web interface of the system. The system accepts Excel data in a structured tabular format containing rows and columns.

The two files may contain the same type of information but can have differences in the records or values. The system does not restrict the comparison to a particular type of data. Different datasets such as student information, employee records, customer information, inventory details, or other structured records can be used.

B. Excel File Upload

After opening the application, the user selects the two Excel files that are required for comparison. The uploaded files are processed by the application using Apache POI, which is used to read Excel workbooks, worksheets, rows, and cell values.

The system reads the contents of the uploaded files without requiring the user to manually copy the data into another format. The column names and records are extracted from the Excel sheets and prepared for the comparison process.

C. Selection of Comparison Key

After uploading the files, the system allows the user to select a common field that can be used to identify corresponding records in both Excel sheets. This field acts as the comparison key.

For example, if the Excel sheets contain student information, the Roll Number can be selected as the key. Similarly, Email, Employee ID, Customer ID, or another unique field can be used depending on the dataset. The selected key is important because it helps the system determine which record in the first Excel file corresponds to which record in the second Excel file. Instead of comparing every row with every other row, the system first uses the selected key to find the corresponding records.

D. Record Matching

Once the comparison key is selected, the system processes the records from both Excel files. The key values from the files are compared to identify corresponding records.

If the same key value exists in both files, the system treats the records as corresponding records and proceeds with a field-level comparison. If a key value is available only in the first file, the record is identified as a record missing from the second file. Similarly, if a key value is available only in the second file, it is identified as a record that is missing from the first file.

This process helps the system classify the records into different categories instead of simply showing that the files are different.

F. Identification of Missing Records

An important part of the reconciliation process is identifying records that exist in one file but do not exist in the other file. The system checks the selected key values between both datasets to identify these records. The records can therefore be categorized as:

- Matched Records – records available in both files with the same compared values.
- Mismatched Records – records available in both files but containing one or more different values. -
Missing in File 1 – records available in the second file but not found in the first file. - Missing in File 2

 © Author(s) | CC BY 4.0 International License.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

– records available in the first file but not found in the second file. This classification gives the user a clear understanding of the differences between the two Excel sheets.

G. Reconciliation Result

After completing the comparison, the system generates a reconciliation result containing the identified matches, mismatches, and missing records. The result is presented through the web interface so that the user can easily review the comparison.

The system provides summary information about the comparison, which helps the user understand the overall result without manually checking the complete Excel files. Detailed mismatch information is also provided for records that require further verification.

H. Search and Filtering

The system provides search and filtering functionality to make the generated results easier to work with. When a large number of records are compared, displaying all the results at once can make it difficult to identify a particular record.

The user can search for a specific record using relevant information and can filter the results according to the comparison category. This allows the user to focus specifically on matched records, mismatched records, or missing records depending on the requirement.

I. Report Generation

The comparison results can be generated as reports for further use. The system provides downloadable Excel reports containing the relevant comparison and reconciliation information.

The generated reports can be used for verification, documentation, record maintenance, or further analysis. This eliminates the need for users to manually prepare a separate report after completing the comparison.

J. Database Storage

The system uses MySQL for storing the required information related to comparison activities. Spring Data JPA is used to manage the interaction between the application and the database.

The database allows relevant comparison information and history to be maintained so that previous reconciliation activities can be referred to when required. This provides better organization and makes the system more useful when multiple comparisons are performed over time.

Technologies Used

The proposed system is developed using a combination of web and data-processing technologies. Java is used as the main programming language, while Spring Boot is used to develop the backend application and manage the different components of the system. Thymeleaf is used to create the web pages and provide interaction between the user and the application.

Apache POI is used for reading and processing Excel files. It allows the application to access worksheets, rows, columns, and cell values programmatically. MySQL is used as the database, and Spring Data JPA is used for database operations.

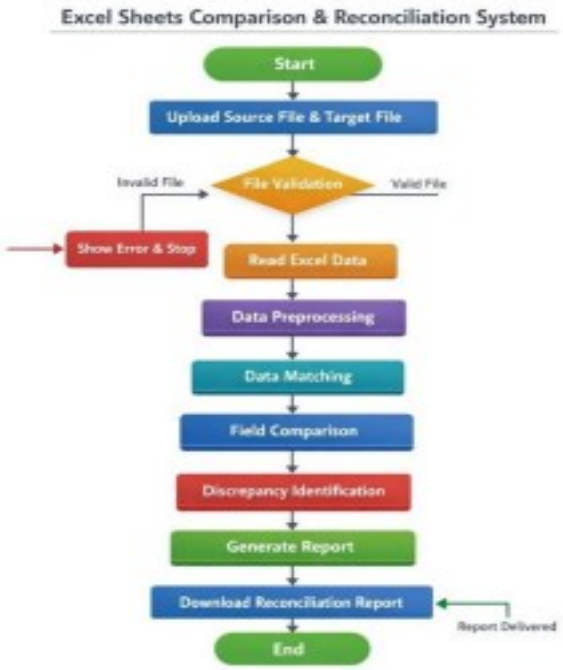
Overall Methodology Flow

The complete methodology of the proposed system can be represented as:

Upload Excel File 1 and File 2 → Select Comparison Key → Read Excel Data → Match Records → Compare Corresponding Fields → Identify Matches/Mismatches/Missing Records → Generate Reconciliation Results → Search and Filter Results → Store Comparison History → Generate and Download Reports

 © Author(s) | CC BY 4.0 International License.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/



RESULTS AND DISCUSSION

The developed Excel Sheet Comparison and Reconciliation System was tested using Excel files containing student attendance data as an example dataset. The attendance data is used only to demonstrate the working of the proposed Excel comparison and reconciliation process. The main purpose of the system is to compare two Excel sheets, identify the corresponding records using a selected key, detect differences, and provide the results in an organized form. The developed application was successfully executed through a web interface and the main outputs obtained during testing are discussed below.

A. Excel File Upload and Comparison Interface

The first output shows the main comparison interface of the system. The interface provides separate options for uploading File 1 and File 2. After selecting the required Excel files, the user can choose the field by which the records have to be compared. In the demonstrated example, RollNo is selected as the comparison key. The Compare Files button starts the comparison process and the View History option is provided for reviewing previous comparison activities. This makes the process easier because the user does not have to manually open both Excel sheets and compare the rows one by one.

Figure 1. Excel Reconciliation System – file upload and comparison interface.

 © Author(s) | CC BY 4.0 International License.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

B. Comparison Summary

After the comparison is completed, the system displays a summary of the obtained results. The shown output reports 0 matched records, 0 mismatched records, 0 records missing from File 1, and 0 records missing from File 2. It also reports a total of 100 students in the example dataset. The attendance part of the demonstration
shows 26 students as eligible and 74 students as not eligible. The zero values in the comparison categories represent the particular test run shown in the screenshot and should not be treated as a limitation of the system. The main result is that the application processes the uploaded data and presents the comparison categories and summary values automatically.

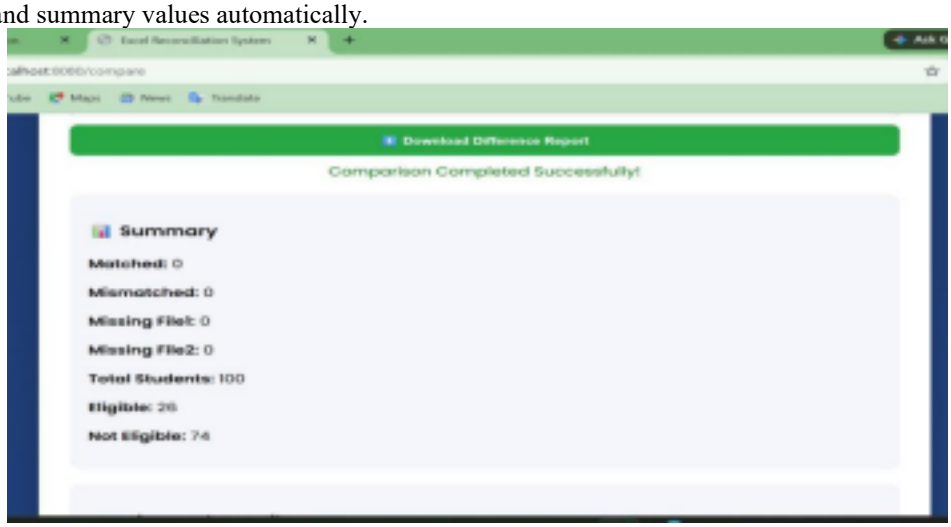


Figure 2. Comparison completed successfully and summary of the processed data.

C. Report Generation and Download

The system also provides options for generating and downloading reports after the comparison process. The report section provides separate options for File 1 and File 2 and a Download Difference Report option. This is useful because the user can save the comparison output instead of depending only on the results displayed on the web page. The generated report can be used for further checking, documentation, and record maintenance. Therefore, the system supports the process from file upload and comparison to result reporting.

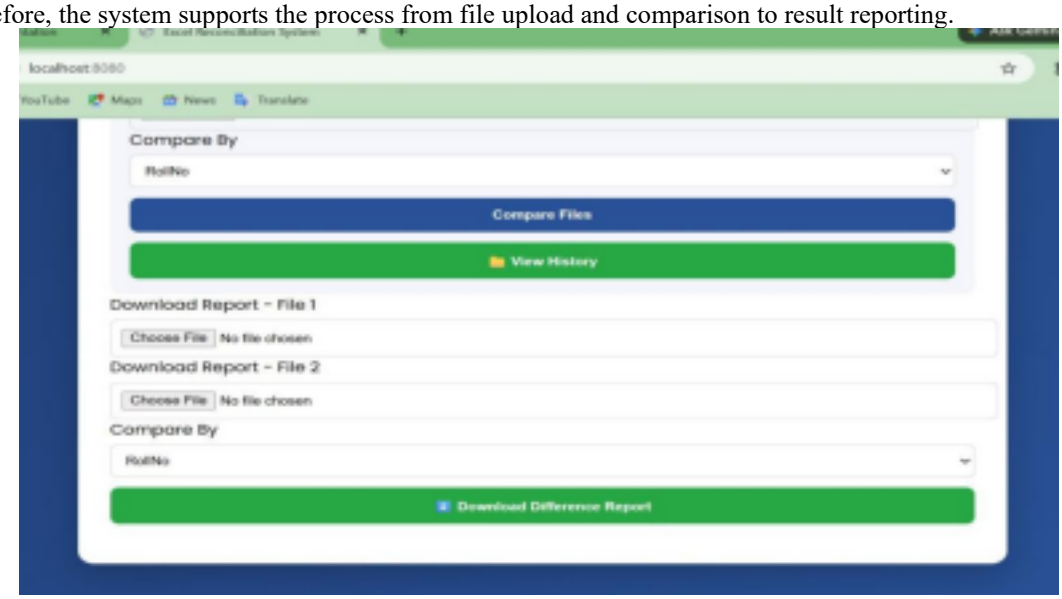


Figure 3. Report download options provided by the system.

D. Attendance Data as a Demonstration Dataset

The following outputs show the attendance records used to demonstrate the applicability of the system with a

 © Author(s) | CC BY 4.0 International License.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

practical dataset. The displayed records contain Roll Number, Name, attendance percentage, and eligibility status. The system displays the attendance percentage for each student and classifies the student as Eligible or Not Eligible. The screenshots show lower percentages such as 18.0%, 19.0%, 14.0%, and 12.0% as Not Eligible, while higher percentages such as 90.0%, 92.0%, 93.0%, 96.0%, 98.0%, and 99.0% are displayed as Eligible. This demonstrates how structured Excel data can be processed and presented through the application. However, attendance analysis is only a demonstration use case; the proposed system is primarily designed for general Excel sheet comparison and reconciliation.

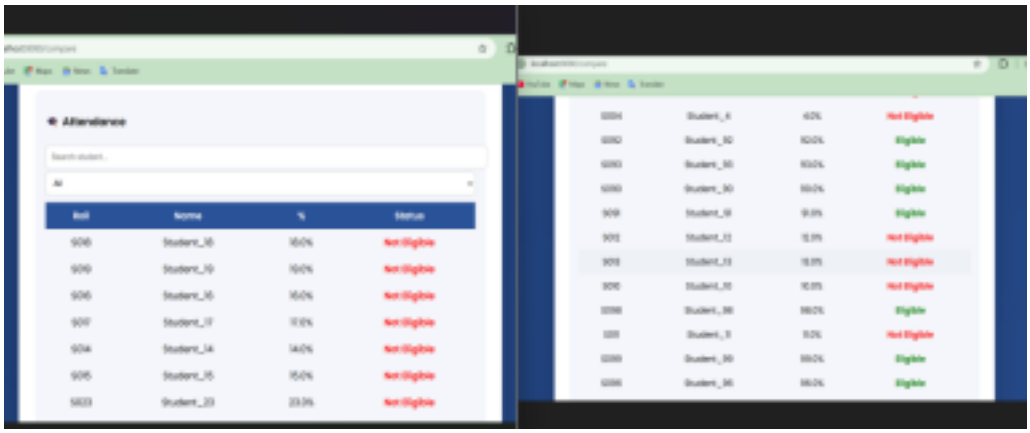


Figure 4. Attendance records displayed with percentage and eligibility status.

E. Discussion of the Results

The results obtained from the developed application show that the proposed system provides a structured way to process Excel data through a web interface. The first output demonstrates that two Excel files can be selected and a comparison key can be chosen before starting the reconciliation process. This is useful because the same comparison system can be used with different types of datasets by selecting an appropriate common field. In the example, RollNo is used as the key, but another suitable field can be selected depending on the dataset.

The summary output gives the user an immediate view of the comparison result. Instead of manually checking a large number of rows, the user can first look at the summary and then concentrate on records that require further attention. The system is intended to identify matched records, mismatched records, and records missing from either file. When differences are present, the difference report provides a convenient way to review them. This reduces the manual effort involved in spreadsheet reconciliation.

Another useful result is the report generation facility. In a manual process, after identifying differences, the user may need to prepare another document or Excel file to record the findings. The proposed system provides downloadable reports directly from the application. This makes the comparison process more organized and helps maintain a record of the results.

The attendance screens further show that the system can work with structured records containing different fields and can present processed information in a readable format. The example contains 100 students, with 26 shown as eligible and 74 shown as not eligible in the demonstrated summary. These values are specific to the sample attendance dataset and are included to demonstrate the application rather than to represent the main objective of the research.

Overall, the developed system demonstrates the basic workflow of Excel reconciliation: uploading two files, selecting a comparison key, processing the records, generating a comparison summary, viewing the results, and downloading reports. The output screens show that the processed information can be presented in a simple manner. The main advantage is that the user can avoid repetitive manual comparison of Excel records. The same approach can be extended to student records, employee records, customer information, inventory data, and other structured spreadsheet information.

CONCLUSION

The Excel Sheet Comparison and Reconciliation System was developed to make the process of comparing two Excel sheets easier, faster, and more reliable. The system reduces the need for manual row-by-row checking by allowing users to upload two Excel files and select a suitable key field for comparison. Based on the selected key, the system processes the records and identifies matched records, mismatched records, and records that are missing from either file.The developed system also provides a clear summary of the comparison results and allows users to review and download the generated reports. The use of Apache POI for Excel processing helps the system

 © Author(s) | CC BY 4.0 International License.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

handle spreadsheet data automatically, while Spring Boot, Spring Data JPA, and MySQL provide the required application and database support. The history feature also helps in maintaining information about previous comparison activities.

The attendance dataset used in the project demonstrates how the system can be applied to a practical type of structured Excel data. However, the system is not limited to attendance data and can be used for different types of Excel datasets such as student records, employee details, customer information, inventory records, and other administrative data.

Overall, the project demonstrates that automated Excel comparison can reduce manual effort, minimize the possibility of human errors, and provide a systematic approach for identifying differences between datasets. In the future, the system can be improved by adding support for more file formats, advanced matching techniques, automatic column mapping, duplicate detection, and more detailed visualization of reconciliation results.

REFERENCES

[1] R. R. Panko, “What We Know About Spreadsheet Errors,” Journal of Organizational and End User Computing, vol. 10, no. 2, pp. 15–21, 1998.
[2] E. Rahm and H. H. Do, “Data Cleaning: Problems and Current Approaches,” IEEE Data Engineering Bulletin, vol. 23, no. 4, 2000.
[3] I. P. Fellegi and A. B. Sunter, “A Theory for Record Linkage,” Journal of the American Statistical Association, vol. 64, no. 328, pp. 1183–1210, 1969.
[4] D. G. Brizan and A. U. Tansel, “A Survey of Entity Resolution and Record Linkage Methodologies,” Communications of the IIMA, vol. 6, no. 3, 2006.
[5] E. Rahm and P. A. Bernstein, “A Survey of Approaches to Automatic Schema Matching,” The VLDB Journal, vol. 10, no. 4, pp. 334–350, 2001.
[6] A. Halevy, A. Rajaraman, and J. Ordille, “Data Integration: The Teenage Years,” Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB), pp. 9–16, 2006.
[7] P. Christen, Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection, Springer, 2012.
[8] S. Aurigemma and R. R. Panko, “The Detection of Human Spreadsheet Errors by Humans versus Inspection (Auditing) Software,” 2010.
[9] OpenRefine, “Reconciliation,” OpenRefine Documentation. The documentation describes reconciliation as matching a dataset with an external source and relates it to record linkage, data matching, entity resolution, field matching, and duplicate detection.
[10] M. A. Hernández and S. J. Stolfo, “The Merge/Purge Problem for Large Databases,” Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, pp. 127–138, 1995.
[11] P. Christen, “Febrl: An Open Source Data Cleaning, Deduplication and Record Linkage System with a Graphical User Interface,” Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1065–1068, 2008.
[12] P. Christen, T. Churches, and M. Hegland, “Febrl – A Parallel Open Source Data Linkage System,” Proceedings of the 8th International Conference on Computational Science, 2004.
[13] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios, “Duplicate Record Detection: A Survey,” IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 1, pp. 1–16, 2007.
[14] Du Plessis, C., & Wa Nkongolo, M. (2026). Detecting reconciliation discrepancies in tabular data using transformers. Frontiers in Artificial Intelligence, 9, Article 1840051.
[15] Heterogeneity in entity matching: A survey and experimental analysis. (2026). Data & Knowledge Engineering, 164, Article 102575.

 © Author(s) | CC BY 4.0 International License.