DOI Creation and Citation

Team NanChan (11)
Mukta Chanda 2021101071
Nandini Maroo 2021111012

Project Description

The "DOI Creation and Citation" system is a robust platform designed to facilitate the seamless management of datasets, licenses, and citations within the realm of research papers. Offering a suite of functionalities, the system ensures efficient DOI generation, customizable templates for data and license use agreements, citation tracking, and more. The primary objective is to provide researchers with an intuitive and standardized platform for managing and referencing datasets in research papers.

Features

- Authenticated entry into the portal.
- Can add datasets into the portal. A unique DOI is generated for every dataset to track its citation.
- Can choose a license for the dataset from the available options or create a custom template.
- Privileged admin role that accepts or denies the custom template to be added into the portal for future purposes.
- Citations for datasets generated in Vancouver, BibTeX, and American and are copied to clipboard for faster citation.
- Can add research papers into the portal. Adobe Extract PDF API is used to extract text from pdf uploads.
- Can download the datasets and view the research papers on the portal. Each license for the dataset shows more information about the license terms.

Requirements

Functional Requirement

The "DOI Creation and Citation" system provides end users with a range of functionalities to
manage data, licenses, generate citations, and assign DOIs to datasets referenced in research
papers. Here is an overview of the key functionalities provided to end users:

1. **Template Management:**
   Functionality: Users can create and manage data/license use agreement templates.
   Description: End users can either select from pre-existing templates or create custom templates.
   Custom templates undergo an admin approval process to ensure quality and standardization.
   Approved templates become available for general use.

2. **DOI Creation:**
   Functionality: Automatically generate DOIs for datasets.
   Description: The system automatically assigns a DOI to each dataset, ensuring a unique identifier
   for easy referencing in research papers.

3. **Data/License Use Agreement:**
   Functionality: Create and manage data/license use agreements.
   Description: Users can create data/license use agreements based on available templates or
   customize them according to their specific needs. Admins review and approve custom
   agreements, making them available as standard templates for all users.

4. **Citation Generation:**
   Functionality: Generate citations in various formats.
   Description: Users can generate citations for datasets in multiple formats (e.g., BibTeX,
   American, Vancouver, etc.) for use in research papers.

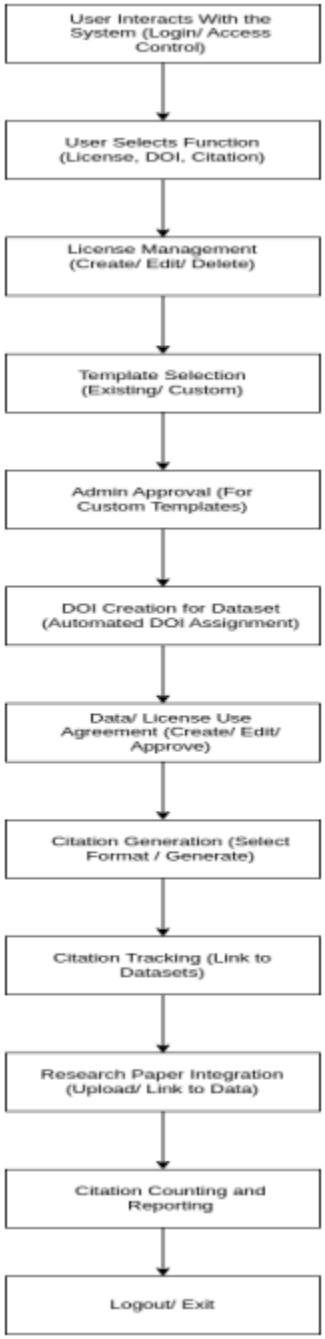
5. **Citation Tracking:**
   Functionality: Track and store citations made in research papers.
   Description: The system records citations made in research papers, linking them to

the
referenced datasets and research papers. This functionality aids in keeping track of dataset usage
and references.

6. **Research Paper Integration:**
   Functionality: Integrate with the research portal.
   Description: The system seamlessly integrates with the research portal, allowing the monitoring
   and capture of research papers uploaded by users. This integration ensures that citations are
   recorded accurately.
7. **Citation Counting and Reporting:**
   Functionality: Count and report dataset citations.
   Description: Users can view citation statistics, allowing them to assess the impact and usage of
   datasets in research papers. This functionality helps in monitoring and measuring the value of
   datasets.
8. **User Interaction:**
   Functionality: Interact with the system for license, citation, and template management.
   Description: End users can create, modify, and manage licenses, citations, and templates. They
   can also access citation statistics and generate citations easily through the user-friendly interface.
9. **Admin Approval Workflow:**
   Functionality: Admins review and approve custom templates and agreements.
   Description: Admins play a crucial role in maintaining the quality and consistency of templates
   and agreements. They approve or reject custom templates and agreements, ensuring they meet
   the required standards.
10. **Standard Template Integration:**
    Functionality: Approved custom agreements become standard templates.

Description: Once custom agreements are approved by admins, they become available as
standard templates for all users, promoting consistency and efficiency.

Functional Flow Diagram



Non-functional Requirements

1. **Reliability:**
    - **Availability:** The system should be available 24/7, with minimal downtime, to ensure that DOIs can be created and cited at any time.
    - **Data Integrity:** The DOI database must maintain the integrity of DOI assignments and citations.
2. **Performance:**
    - **Response Time:** The system should generate DOIs and citations quickly, with low latency, to prevent delays for users.
    - **Scalability:** The system should handle a growing number of DOI requests and citations efficiently. It must scale horizontally to accommodate increased demand.
3. **Security:**
    - **Data Encryption:** All communication with the system should be encrypted to protect sensitive DOI-related data.
    - **Authentication and Authorization:** Access to the DOI creation and citation feature should be restricted and controlled based on user roles and permissions.
    - **Compliance:** Ensure compliance with DOI registration standards and security best practices.
4. **Usability:**
    - **User Interface:** The user interface for creating and citing DOIs should be intuitive and user-friendly.
    - **Accessibility:** The system should be accessible to users with disabilities, ensuring that all users can create and cite DOIs.
    - **Documentation:** Provide comprehensive documentation and support materials for users to understand and utilize the feature effectively.
5. **Auditability:**

- **Logging:** The system should maintain detailed logs of DOI creation and citation activities for auditing and tracking purposes.

6. **Scalability:**
    - **Horizontal Scalability:** The system should be designed to handle a large number of DOI creation and citation requests, with the ability to scale horizontally by adding more servers if needed.
7. **Maintenance:**
    - **Code Maintainability:** Ensure that the codebase is well-structured and documented for easy maintenance and future enhancements.
    - **Modifiability:** The system should be designed to accommodate changes and updates, with minimal impact on existing functionality.
8. **Load Handling:**
    - **Load Testing:** Conduct load testing to determine the system's performance under various levels of load and ensure it can handle peak demand without degradation in service.
9. **Backup and Recovery:**
    - **Data Backup:** Implement regular data backups to prevent data loss and facilitate recovery in case of system failures.
    - **Disaster Recovery Plan:** Have a plan in place for recovering the DOI database and associated services in the event of a disaster.
10. **Compliance:**
    - **DOI Registration Standards:** Ensure compliance with relevant DOI registration standards and guidelines, as well as any legal or industry-specific requirements.
    - **Data Retention:** Define data retention policies to comply with data protection and privacy regulations.
11. **Data Migration:**
    - If migrating an existing database to include DOIs, ensure a smooth transition and data integrity during the migration process.
12. **Third-Party Integration:**

- If the system integrates with external DOI registration services or citation management tools, ensure seamless integration and compatibility.

Stakeholder buy-ins

1. **Researchers:** Those who will use the DOI feature for citing their work.
2. **Database Administrators:** Responsible for managing and maintaining the database.
3. **Institutional Leadership:** University or organizational administrators who provide support and funding.
4. **Legal and Compliance Teams:** Ensure compliance with policies and legal requirements.
5. **Financial Stakeholders:** Those responsible for budget allocation and funding.
6. **DOI Registry Representatives:** External partners for DOI integration.
7. **End Users of Cited Works:** Those whose work is cited using DOIs.
8. **Project Sponsors:** Grant providers or project sponsors who initiated the project.
9. **Database and IT Teams:** Responsible for implementing and managing the DOI feature.
10. **User Experience (UX) Designers:** Collaborate on user-friendly design and user expectations.

Design

Architecture Compliance:

The system architecture is designed for compliance with industry best practices and technology standards. Key components and technologies contributing to architecture compliance include:

- **Microservices Architecture:**
  - The system follows a microservices architecture, dividing functionalities into independent, loosely coupled microservices.

    - Each microservice handles specific tasks, such as template management, DOI creation, citation generation, and research paper integration.
    - This architectural choice enhances scalability, maintainability, and flexibility.
- **MERN Stack:**
    - **MongoDB:** The choice of MongoDB as the database ensures a scalable and flexible data storage solution, fitting the dynamic nature of the application.
    - **Express.js:** The backend utilizes Express.js as a web application framework for Node.js, providing a robust foundation for building RESTful APIs.
    - **React.js:** React.js is employed for building a responsive and interactive user interface on the frontend.
    - **Node.js:** The backend runtime environment is powered by Node.js, offering a non-blocking, event-driven architecture.
- **Dockerization:**
    - The system leverages Docker containers to encapsulate each microservice and its dependencies.
    - Dockerization ensures consistency in deployment across various environments, easing scalability and facilitating efficient resource utilization.
- **Adobe Extract PDF API:**
    - The Adobe Extract PDF API is integrated into the system for extracting text from PDF uploads.
    - This API enhances the system's capability to process and analyze content from research papers, contributing to seamless data integration.

Meet Functional Requirements:

- **Template Management:**

  Users can create and manage data/license use agreement templates. Custom templates undergo an admin approval process, promoting quality and standardization.
- **DOI Creation:**

Automatic generation of DOIs for datasets ensures a unique identifier for easy referencing in research papers.

- **Data/License Use Agreement:**

  Users can create, customize, and manage data/license use agreements. Admins review and approve custom agreements, making them available as standard templates.

- **Citation Generation:**

  Users can generate citations for datasets in various formats (e.g., BibTeX, Vancouver, American) for use in research papers.

- **Citation Tracking:**

  The system records citations made in research papers, linking them to referenced datasets. This aids in tracking dataset usage and references.

- **Research Paper Integration:**

  Seamless integration with the research portal ensures accurate recording of research papers uploaded by users.

- **Citation Counting and Reporting:**

  Users can view citation statistics, allowing assessment of dataset impact and usage in research papers.

- **User Interaction:**

  End users can easily manage licenses, citations, and templates through the user-friendly interface.

- **Admin Approval Workflow:**

  Admins play a vital role in maintaining quality and consistency, reviewing and approving custom templates and agreements.

- **Standard Template Integration:**

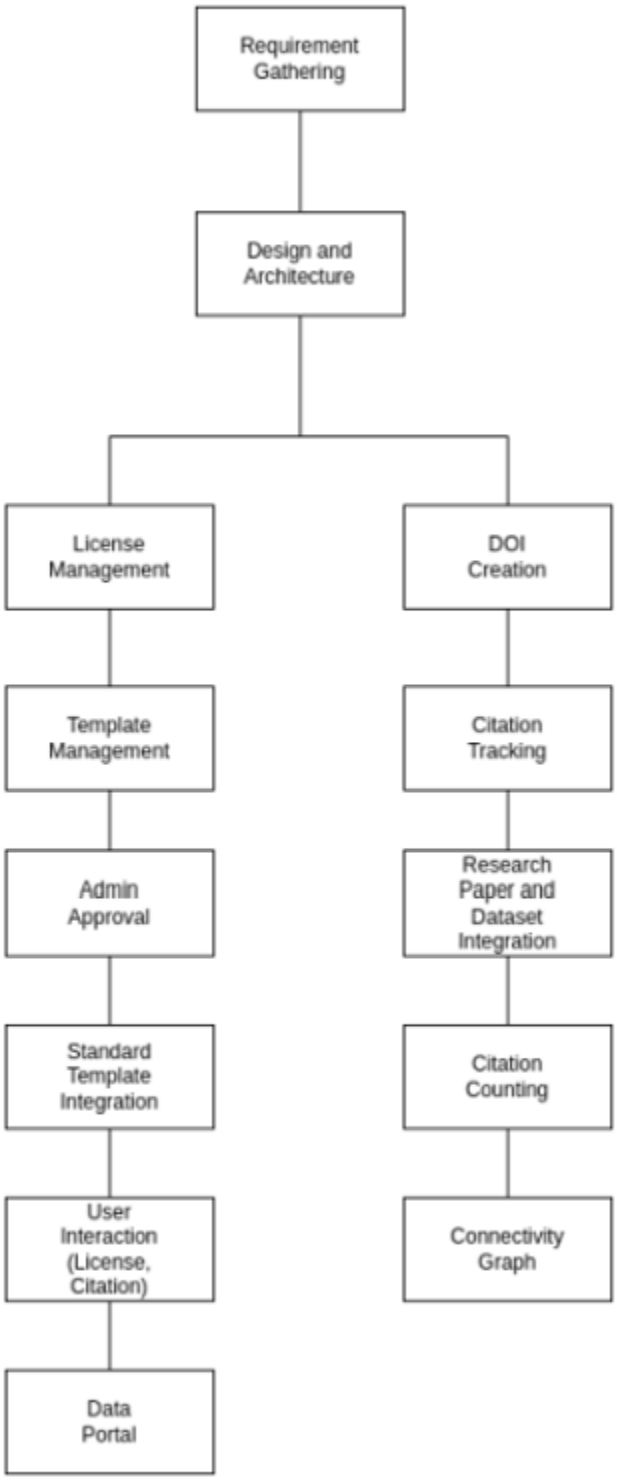
  Approved custom agreements become standard templates for all users, promoting consistency.

Database Schema:

- **User:**

    - unique_id
    - email
    - username
    - password
- **License:**
    - name
    - info
    - pending
- **Dataset:**
    - name
    - author
    - publicationYear
    - count
    - fileLink
    - license
    - doi
- **Research Paper:**
    - name
    - author
    - fileLink

Logical Flow Diagram:



Meet Non-functional Requirements

- **Reliability:** The system achieves reliability through 24/7 availability and robust data integrity mechanisms. Continuous monitoring and automated recovery processes contribute to high reliability.
- **Performance:** The system meets performance requirements with low-latency responses for DOI and citation generation. Horizontal scalability ensures efficient handling of increased requests.
- **Security:** Security measures include data encryption, strong authentication, and authorization mechanisms. Compliance with DOI registration standards and best practices enhances overall system security.
- **Usability:** The user interface is designed for ease of use, with intuitive navigation and accessibility considerations. Comprehensive documentation supports users in understanding and utilizing the system effectively.
- **Auditability:** Detailed logging mechanisms capture DOI creation and citation activities for auditing purposes. Logs provide transparency and traceability.
- **Scalability:** The system's microservices architecture allows for horizontal scalability, ensuring it can handle a growing number of requests by adding more servers.
- **Maintenance:** The codebase is structured for maintainability, with clear documentation and modularity. Future enhancements can be seamlessly integrated with minimal impact on existing functionality.
- **Load Handling:** Load testing is conducted to assess the system's performance under varying loads, ensuring it can handle peak demand without compromising service quality.
- **Backup and Recovery:** Regular data backups are implemented to prevent data loss, and a disaster recovery plan is in place to facilitate quick system recovery in case of failures.
- **Compliance:** The system adheres to DOI registration standards, legal requirements, and industry-specific regulations. Data retention policies are defined to comply with privacy regulations.
- **Data Migration:** If applicable, a structured data migration plan ensures a smooth transition, minimizing the impact on data integrity during the migration process.

- **Third-Party Integration:** For systems integrating with external services, compatibility and seamless integration are ensured, meeting the requirements of third-party DOI registration services or citation management tools.

Planning and Execution

Schedule and Plan

The planning and execution phase of the "DOI Creation and Citation" system followed a meticulously crafted schedule and plan to ensure the successful completion of the project. The timeline was structured to accommodate various development stages, testing, and deployment.

Task Completed

1. **System Architecture Design:** Defined a robust architecture compliant with functional requirements, ensuring scalability and efficiency.
2. **Implementation of Functional Requirements:** Developed and implemented all specified functionalities, including template management, DOI creation, data/license use agreements, citation generation and tracking, research paper integration, and more.
3. **Admin Approval Workflow:** Implemented a streamlined workflow for admin approval of custom templates and agreements, maintaining quality and standardization.
4. **Integration with Research Portal:** Successfully integrated the system with the research portal, enabling accurate monitoring and capture of uploaded research papers.
5. **Citation Counting and Reporting:** Implemented functionality for counting and reporting dataset citations, providing users with valuable insights into dataset usage.
6. **User-Friendly Interface:** Designed and implemented an intuitive user interface for easy interaction with the system, including license, citation, and template management.
7. **Standard Template Integration:** Established a seamless process for integrating approved custom agreements as standard templates, promoting consistency and efficiency.

Key Achievements

- Successful implementation of all specified features and functionalities.
- Seamless integration with the research portal for enhanced user experience.
- Robust admin approval workflow ensuring quality and standardization.
- User-friendly interface facilitating easy management of licenses, citations, and templates.

Challenges

Throughout the project, challenges were encountered and effectively addressed, contributing to the overall improvement of the system. Challenges included:

- **Integration Complexity:** Overcoming complexities during the integration phase, particularly with the Adobe Extract PDF API for research paper uploads.
- **Customization Balance:** Striking a balance between providing customization options for users and maintaining standardization for consistent data management.

Work Distribution

- **Development Team:** Responsible for the implementation of core functionalities, system architecture, and user interface design. (Mukta+Nandini)
- **Testing Team:** Conducted comprehensive unit testing, integration testing, and system testing to ensure the robustness of the system. (Mukta)
- **Admin Approval Team:** Focused on refining and optimizing the admin approval workflow, ensuring efficient processing of custom templates and agreements. (Mukta)
- **Documentaion** (Nandini)

Pending Tasks

As of the project completion phase, all planned tasks and functionalities have been successfully implemented. The system is ready for deployment, and any additional tasks or optimizations will be addressed in future updates based on user feedback and evolving requirements.