CONTENT DOCUMENT

ROUGH SRS

Overview:

The Press Registrar General of India (PRGI) is responsible for maintaining a comprehensive database containing approximately 160,000 registered titles. When a user submits a new title for verification, it is essential to ensure that the proposed title does not duplicate or closely resemble any existing title. This process is critical in preventing confusion, maintaining the uniqueness of registered publications, and ensuring adherence to specific regulatory guidelines. However, the current method of title verification presents several challenges that necessitate the development of an automated system.

The Actual Problem:

The actual problem is that PRGI lacks an efficient and accurate system for verifying the uniqueness of new title submissions against its extensive database of 160,000 registered titles. The current manual verification process is slow, inconsistent, and prone to human errors, leading to cases where duplicate or misleading titles are either mistakenly approved or wrongly rejected. Basic keyword searches fail to detect phonetically similar words, synonymous phrases, or minor modifications that alter the structure but not the meaning of a title. Additionally, PRGI guidelines restricting certain words and title combinations are not systematically enforced, resulting in regulatory violations. The growing volume of submissions further exacerbates these challenges, making manual verification increasingly unsustainable. Without an automated system, ensuring the uniqueness and compliance of new titles remains a complex and inefficient task.

Drawbacks of Existing Methods:

The existing approach primarily relies on manual verification or basic keyword searches. This method is highly inefficient due to the sheer volume of registered titles and the continuous influx of new applications. Manually comparing new submissions against such a vast database is time-consuming, making the process slow and cumbersome. Additionally, human intervention introduces the risk of inconsistencies and errors, as different individuals may interpret title similarities subjectively. As a result, some titles that should be rejected might get approved, while others that are sufficiently distinct may be incorrectly flagged as duplicates.

A significant limitation of keyword-based searching is its inability to detect phonetically similar words, synonymous phrases, or minor modifications that alter the structure of a title without changing its essence. Titles such as "Times Journal" and "Journal of Times" may bypass simple keyword matching systems despite their striking resemblance. Furthermore, issues such as spelling variations, spacing inconsistencies, and punctuation differences add another layer of complexity, making manual verification an unreliable approach for ensuring uniqueness.

Another pressing issue in the current system is the enforcement of PRGI guidelines. Certain words and phrases are explicitly restricted from being used in titles, and combinations of existing titles should not be permitted. However, due to the manual nature of the verification process, these guidelines are not systematically enforced, leading to occasional breaches of regulatory policies. Without a robust mechanism to detect and reject such violations consistently, the integrity of the database is compromised.

Scalability is another concern with the current manual verification process. As the number of registered titles continues to grow, the system becomes increasingly unsustainable. The demand for title verification is only expected to rise, necessitating additional manpower and leading to escalating operational costs. A manual system, no matter how well-structured, cannot efficiently handle an ever-expanding database while maintaining accuracy and efficiency.

Root Cause of the Problem:

The root cause of this problem lies in the inefficiency and limitations of the manual verification process currently used by PRGI. The vast number of registered titles and continuous new submissions make it impractical to manually compare and validate each entry. Human-based verification is slow, inconsistent, and prone to errors, leading to misjudgements in title uniqueness. Furthermore, basic keyword search methods fail to detect phonetically similar words, synonymous phrases, or minor modifications, allowing duplicate or misleading titles to pass through. Additionally, the lack of systematic enforcement of PRGI guidelines contributes to regulatory breaches, making the process unreliable. The inability of the manual system to scale efficiently with the growing database further exacerbates the problem, making automation a necessary solution.

Impacts of the Problem:

The existing manual system for title verification creates significant inefficiencies for all stakeholders involved.

For PRGI, the process is slow and labour-intensive, requiring human reviewers to manually compare new submissions against a database of 160,000 titles. This not only delays approvals but also increases operational costs. The reliance on subjective judgment leads to inconsistencies, where duplicate titles may be approved while unique ones are wrongly rejected. Additionally, enforcing PRGI guidelines is difficult, resulting in occasional regulatory breaches.

For reviewers, the burden of manually verifying numerous titles daily leads to fatigue and errors. The limitations of keyword-based searches make it challenging to detect phonetically similar words, minor structural variations, and synonym-based duplications. This adds complexity and increases the risk of inaccurate decisions.

For applicants, delays in verification create frustration, impacting the timely registration of their titles. The lack of clear rejection reasons makes it difficult for them to modify and resubmit their proposals, leading to repeated attempts and wasted effort. Furthermore, the risk of approval for misleadingly similar titles creates confusion and potential disputes among publishers.

For the public and media industry, the approval of closely resembling titles undermines uniqueness, making it harder to distinguish between different publications. This affects credibility and creates confusion in the marketplace.

The inefficiencies of the current system highlight the urgent need for automation. A technology-driven solution would streamline operations, improve accuracy, enforce guidelines consistently, and enhance user experience for all involved.

Proposed solution:

To address the challenges associated with manual title verification, an automated system will be developed to efficiently analyse and validate new title submissions. This system will leverage advanced similarity detection techniques, enforce regulatory guidelines, and provide users with actionable feedback.

The core of the solution involves implementing phonetic similarity algorithms (such as Soundex and Metaphone) to detect closely resembling titles, ensuring that minor spelling variations do not bypass the verification process. Additionally, the system will maintain a structured list of disallowed prefixes, suffixes, and words, rejecting titles that violate these predefined rules.

To enhance accuracy, a probability-based verification model will be introduced, calculating a similarity percentage and determining the likelihood of approval. This ensures transparency in the decision-making process. The system will also track current applications, preventing duplicate or overly similar titles from being submitted in the future.

A high-performance database architecture with optimized search techniques will enable rapid comparisons against the extensive repository of 160,000 existing titles. Scalability considerations will ensure the system remains efficient as the database expands.

For user interaction, the system will provide real-time feedback, highlighting reasons for rejection and offering users the option to refine and resubmit their titles. This ensures a smooth and intuitive verification process while maintaining the integrity and uniqueness of registered publications.

By integrating these components, the proposed system will significantly reduce manual effort, enhance accuracy, improve processing speed, and ensure regulatory compliance, ultimately streamlining the title verification workflow.

Importance of Automation:

An automated web-based system is imperative to streamline the title verification process. Such a system would significantly enhance the speed and accuracy of verification by leveraging advanced algorithms, including fuzzy matching, natural language processing (NLP), and machine learning techniques. These technologies would enable the system to detect not only exact matches but also phonetically similar words, structural rearrangements, and semantically identical phrases. By automating the comparison process, the likelihood of errors and inconsistencies would be drastically reduced.

An automated system would also ensure the standardized enforcement of PRGI guidelines. It would automatically flag and reject titles containing restricted words or impermissible combinations, thereby upholding regulatory compliance. Furthermore, the system could be designed to provide detailed explanations for rejected titles, offering suggestions for modifications to assist applicants in choosing unique and compliant titles.

Efficiency in handling large-scale data is another advantage of an automated approach. With optimized database indexing and search algorithms, millions of titles could be processed within seconds, making the verification process both seamless and scalable. Unlike manual verification, which requires extensive human effort, an automated system could accommodate increasing data volumes with minimal resource allocation.

Additionally, a web-based platform would provide a user-friendly interface, allowing applicants to submit their titles online and receive instant feedback on their validity. This real-time validation mechanism would eliminate the need for prolonged waiting periods, reducing administrative overhead and enhancing user experience. The ability to access verification results promptly would benefit both PRGI and the applicants by expediting the registration process.

The implementation of an automated title verification system for PRGI is not just a technological advancement but a necessity for modernizing the process. By addressing the inefficiencies of the current manual system, ensuring compliance with regulatory standards, and providing a scalable solution for future growth, the development of such a system represents a crucial step toward enhancing transparency, accuracy, and efficiency in title registration.

Benefits of Automation:

An automated title verification system streamlines operations, ensuring faster processing, reduced workload, and lower costs for PRGI. By enforcing guidelines consistently, it eliminates errors and regulatory breaches, maintaining the integrity of registrations.

For reviewers, automation reduces manual effort, allowing focus on complex cases. Advanced algorithms detect phonetic similarities, synonym-based duplications, and minor modifications more accurately, improving decision-making and preventing misleading approvals.

For applicants, instant feedback replaces long waiting periods. Clear rejection reasons and modification suggestions simplify resubmissions, reducing repeated attempts and enhancing user experience.

For the public and media industry, unique titles preserve originality, preventing confusion and strengthening credibility. A standardized, transparent process fosters trust in PRGI's registration system.

By improving speed, accuracy, and accessibility, automation is essential for modernizing title verification.

Challenges in Implementation:

Implementing the automatic title verification system presents several complex challenges. One of the most critical difficulties lies in ensuring the accuracy of similarity detection. The system must correctly identify phonetic and textual similarities without generating false positives or false negatives, which requires sophisticated algorithms capable of handling variations in spelling, transliteration, and minor modifications.

Another major challenge is optimizing performance while handling a vast database of over 160,000 registered titles. Efficiently searching and comparing new submissions in real time demands advanced indexing and search techniques to prevent latency issues as the system scales.

Managing multilingual title verification adds further complexity, as the system must recognize similar meanings across different languages, accounting for nuances in translation and local naming conventions. Additionally, developing a reliable verification probability score that provides meaningful and consistent evaluations based on similarity metrics remains a critical hurdle.

Finally, ensuring security and integrity within the system is paramount. Preventing unauthorized modifications to restricted word lists, avoiding title manipulation workarounds, and integrating seamlessly with existing PRGI workflows while maintaining a user-friendly experience require careful architectural design and ongoing refinement.

Stakeholders:

The key stakeholders involved in this system include,

1. Press Registrar General of India (PRGI) Officials

- Oversee and manage the registration of titles.

- Ensure compliance with regulatory guidelines.

- Utilize the system for efficient title verification.

2. Applicants (Publishers, Media Houses, Individuals)

- Submit new titles for verification.

- Receive feedback and modify submissions if required.

- Rely on the system for a transparent approval process.

3. System Administrators & Developers

- Maintain, update, and optimize the automated verification system.

- Ensure database integrity and performance.

- Address technical issues and implement new features.

4. Government & Regulatory Bodies

- Define and enforce title registration guidelines.

- Monitor compliance with national media and publication laws.

5. Legal & Compliance Teams

- Handle disputes related to title rejections.

- Ensure that the system aligns with copyright and trademark regulations.

6. General Public & Readers

- Benefit indirectly from clear, unique, and distinguishable publication titles.

- Avoid confusion due to duplicate or misleading publication names.

Each stakeholder plays a crucial role in ensuring the system functions efficiently and maintains the integrity of the title verification process.

Users and their Actions:

The following are the actions provided for each user,

1. Title Applicant:

- Can register, log in, and manage their account

- Can submit titles for automated verification

- Can view verification results and modify/register title, if needed

- Can search and view existing title records and personal history

- Can access help resources and contact support

2. PRGI Official:

- Can log in and manage their official profile

- Can manage restricted words and banned affixes

- Can view real-time title registrations

- Can approve/reject title registrations after other legal processes

- Can browse, filter, and view existing title records

3. System Administrator:

- Can manage all user accounts and access controls

- Can approve PRGI official account registrations

- Can respond to support queries from users

- Can access and monitor all user activities

Users and their Actions: (Extended)

The following are the actions provided for each user,

1. Title Applicant:

- Can register and log in to the site

- Can submit a title for verification

- Can view similarity results and guideline violations

- Can view verification probability and final status (accepted/rejected)

- Can modify and resubmit rejected titles

- Can register accepted titles by submitting additional details

- Can search and view existing registered titles

- Can access their title verification and registration history

- Can read help and guidelines

- Can contact support for queries

2. PRGI Official:

- Can register and log in after admin approval

- Can view dashboard with real-time title registrations history

- Can manage restricted words and disallowed affixes

- Can approve/reject title registrations after other legal processes

- Can browse, filter, and view existing title records

- Can access help and contact support

3. System Administrator:

- Can access admin-only dashboard (single fixed account)

- Can view and respond to support queries submitted by users

- Can inspect user profiles and activity logs

- Can approve or reject new official account requests

- Can view, freeze or even delete user accounts (applicants and officials)

- Can monitor system health, database checks, and ensure platform integrity

SYSTEM REQUIREMENTS:

Functional Requirements:

1. **Similarity Detection System:**
   Detect phonetic, spelling-variant, and semantic similarities between submitted titles and existing ones. Uses algorithms like Soundex, Metaphone, Levenshtein distance, and optional NLP-based embeddings.

   Example: A user submits "Namaskar Bharat". The system finds it 88% similar to "Namascar Bharat Times" using Soundex + Levenshtein, and flags it for review.

2. **Guideline Rule Enforcement:**
   Automatically checks titles against PRGI rules such as banned words, restricted prefixes/suffixes, prohibition of combination of titles and addition of periodicities to other titles.

   Example: A user submits "Crime Reporter Daily". The system flags it due to the banned word "Crime" and improper suffix usage.

3. **Multilingual Title Handling:**
   Detects semantic duplicates across different languages. Involves using multilingual synonym dictionaries or translation APIs to find meaning overlaps.

   Example: A Hindi title "Dainik Samachar" is flagged for being semantically equivalent to the English title "Daily News".

4. **User Feedback and Guidance:**
   Gives clear, actionable reasons for title rejections to help users identify the problem in the submitted title.

   Example: A title like "Indian Express Chronicle" is rejected with explanation: "Combination of two existing titles: Indian Express + Chronicle." Suggestions are provided to modify the name.

5. **Title Submission and Tracking:**
   Tracks submissions, prevents duplicate applications, and maintains user-specific title history.

   Example: If "Economic Post" is already submitted by one user and under review, another user submitting the same title will be notified of the conflict.

6. **Role-Based Access System:**
   Different interfaces and permissions for Title Applicants, PRGI Officials, and System Administrators.

   Example: Only PRGI Officials can manage banned words or approve title registrations. Admin responds to queries.

Non-Functional Requirements:

1. **Scalability:**
   The system must efficiently handle a large volume of data and users.

   Example: Even with 200,000 titles and 2,000 daily submissions, the system should not crash or lag.

2. **Performance:**
   Fast response time for similarity checks and feedback.

   Example: Title similarity and rule validation completes under 2 seconds even during peak usage.

3. **Usability:**
   The interface should be intuitive and assist users throughout the process.

   Example: A user gets suggestions as they type, like "Avoid banned words such as 'CBI', 'Crime', etc."

4. **Accuracy:**

High correctness in similarity detection and rule enforcement.

Example: Two visually different titles like "Desh ki Awaaz" and "Voice of Nation" are correctly flagged as similar.

5. **Maintainability:**

Easy to update rules and logic without full system redeployment.

Example: Official adds "Forces" to the disallowed list from their dashboard, no code changes required.

6. **Security:**

Secure authentication and protection of title data and user accounts.

Example: Titles submitted by one applicant can't be viewed or edited by another. Passwords are hashed and stored securely.

Software Features:

The list of software features categorized based on the 3 types of users is given below,

Title Applicant:

1. **Title Submission & Analysis**
   - Users can submit a new title through their dashboard input field.
   - Upon submission, the system redirects them to an analysis page.
   - The **left sidebar displays the top 50 most similar titles** from the database for user comparison.
     - *Example:* If a user submits "India Morning Express," they might see "India Evening Express" at the top with 93% similarity.
   - The **top match's similarity score is shown**, and the **verification probability** is calculated as 100 - similarity score.
   - If a disallowed word or rule violation is detected, **verification probability becomes 0%** immediately.
     - *Example:* If the title contains the word "CBI," even if the similarity score is low, it will still get a 0% probability.
   - Based on the PRGI-set threshold (e.g., 30% minimum probability), the title gets:
     - Status: Accepted if above threshold.
     - Status: Rejected if below.

2. **Feedback for Rejected Titles**
   - If a title is rejected, the user is shown a **feedback explanation** based on the failed rules.
     - *Example:* "Rejected due to disallowed word: 'Police'" or "Too similar to existing title: 'India Police Weekly'."

3. **Modify & Resubmit Option**
   - Users can edit the rejected title and try again using a **"Modify and Resubmit"** button.

4. **Title Registration (Post-Approval)**
   - If a title is accepted, users can **fill out required registration details** to officially register the title.
   - The title is registered once an official approves the registration, after all other legal processes.
     - *Example:* They may need to enter publisher name, location, language, etc.

5. **Search and View Existing Titles**
   - Users can **search for and explore existing verified titles** in the database.
   - Filter and sort options (by language, region, etc.) may be available.

6. **Verification & Registration History**
   - Full **history of submissions and registration attempts** with status shown in table.
     - *Example:* "My Submissions: India Morning Express – Rejected – Submitted on 03/02/2025."

7. **Notification System**
   - Real-time **notifications** for:
     - Admin responses to contact queries.
     - Official approvals for title registrations.

8. **Help & Guidelines Section**
   - A **user help centre** that includes:
     - Guidelines for creating valid titles.
     - Visual examples of accepted vs. rejected titles.
   - The document includes explanations of the features available for any Title Applicant.

9. **Contact Support**
   - Users can **submit a query or report a problem** to the system administrator.
   - Responses are shown as **notifications** within the user's panel.

10. **Account Settings**
    - Modify profile details like email, password, and personal info.
    - *Example:* Update registered email address from user1@mail.com to user2@mail.com.

PRGI Official:

1. **Dashboard Overview**
   - Monitor trends such as title registrations, along with the timestamp of registration.

2. **Approval of Title registrations**
   - Approve or reject any registration manually based on other legal documents.
   - All other activities except verification and registration of titles are done externally.

3. **Restricted Vocabulary Management**
   - Add or edit restricted words and banned affixes (prefixes/suffixes).
   - All changes directly affect automated guideline enforcement.

4. **Existing Titles Explorer**
   - Search and filter the database of registered titles by language, status, publisher, etc.
   - View full metadata of each title (e.g., submitter, acceptance date, approving official).

5. **Help and Guidance Access**
   - Access the guideline documentation to understand the system and the process of title verification.
   - The document includes explanations of the features available for any PRGI official.

6. **Support Communication**
   - Send queries to admins through an internal messaging system.

7. **Account and Settings Management**
   - Update login details, change password, and configure system preferences.

System Administrator:

1. **Support Query Management**
   - View all support queries submitted via the "Contact Support" feature.
   - Respond to user queries, with replies sent as in-app notifications.

2. **User Activity Monitoring**
   - View detailed profiles of all users (applicants and officials).
   - View activity history such as title submissions, registrations, etc.
   - Monitor behaviour for misuse, inactivity, or anomalies.

3. **Official Account Verification**
   - Access a list of all official signup requests.
   - Manually verify the legitimacy of official credentials.
   - Approve or reject official registrations.
   - Only upon approval can an official log in.

4. **User Account Management**
   - View a list of all registered users (both applicants and officials).
   - Temporarily freeze user accounts to restrict access.
   - Permanently remove users from the system.

5. **Exclusive Admin Access**
   - Only one admin account exists.
   - No public sign-up route for admin.
   - Admin credentials are securely maintained and cannot be changed by other user roles.

6. **Platform Integrity Oversight**
   - Periodically check the functional behaviour of modules (title verification, registration, etc…).
   - Verify that the database reflects the correct and expected data entries.
   - Although most non-functional checks are done outside the app (e.g., server tools), the admin ensures feature integrity.

Use Cases:

Here are the use cases for each of the three users,

1. Title Applicant:
- Sign Up
- Log In
- Log Out
- Reset Password
- Submit Title for Verification
- View Similarity Score
- View Verification Probability
- View Verification Result
- Receive feedback on Rejection
- Modify and Resubmit Title
- Register Title
- Search and view existing titles
- View Submission History
- Manage account settings
- Access Help & Guidelines
- Contact Support

2. PRGI Official:
- Sign Up
- Log In
- Log Out
- Reset Password
- Monitor Title Registration Trends
- Approve/Reject Title registrations
- Manage Restricted Words and Affixes
- Search and View Existing Titles
- Manage account settings
- Access Help & Guidelines
- Contact Support

3. System Administrator:

- Log In
- Log Out
- Reset Password
- Monitor User Activity
- Approve/Reject Official accounts
- Manage User Accounts
- Respond to Support Queries
- Manage account settings

END