RESEARCH DOCUMENT

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

- Submit a new title for verification.
- Receive feedback on title similarity, restricted words, and compliance issues.
- View the probability score of title verification.
- Modify and resubmit a rejected title.

2. PRGI Official

- Review and approve/reject submitted titles manually if needed.
- Update the list of restricted words and disallowed prefixes/suffixes.
- Monitor the system's verification accuracy and flag issues.

3. System Administrator

- Manage user roles and access permissions.
- Oversee system performance, uptime, and security.
- Update and optimize verification algorithms and database indexing.
- Troubleshoot technical issues and implement necessary fixes.

Functional Requirements:

The functional requirements of the software include,

1. Similarity Check:

a. Implement a mechanism to check for similar-sounding names using phonetic similarity algorithms (e.g., Soundex, Metaphone).

b. Identify titles that have common prefixes or suffixes (e.g., The, India, Samachar, News).

c. Ensure that variations in spelling or slight modifications do not bypass the similarity check (e.g., Namaskar vs. Namascar).

d. Calculate a similarity percentage for each title comparison.

2. Prefix/Suffix Handling:

a. Maintain a list of disallowed prefixes and suffixes.

b. Reject any new titles that include these disallowed prefixes or suffixes if they cause the new title to resemble an existing title closely.

3. Guideline Enforcement:

a. Maintain a list of disallowed words (e.g., Police, Crime, Corruption, CBI, CID, Army).

b. Ensure that titles containing these disallowed words are rejected.

c. Prevent the creation of new titles by combining existing ones (e.g., if "Hindu" and "Indian Express" exist, "Hindu Indian Express is not allowed").

d. Check for titles with similar meanings in other languages and reject them (e.g., "Daily Evening" and "Pratidin Sandhya").

e. Disallow adding periodicity (e.g., daily, weekly, monthly) to existing titles to form new ones.

4. Verification Probability:

a. Provide a probability score indicating the likelihood of a title being verified. For instance, if a title has a similarity score of 80%, the verification probability shall not be more than 100%-80%=20%

5. User Feedback:

a. Provide clear feedback to the user if their submitted title is too similar to an existing title, contains disallowed prefixes/suffixes, violates guidelines, or is created by combining existing titles.

b. Display the verification probability to the user.

c. Allow the user to modify their title and resubmit it for verification.

Non – Functional Requirements:

The non-functional requirements of the software include,

1. Database Interaction:

a. Efficiently search and compare new titles against the database of 160,000 titles.

b. Track current applications and use them for future reference to reject similar titles submitted later.

c. Use indexing and optimised search techniques to handle the large dataset and ensure quick responses.

2. Scalability:

a. Design the system to handle an increasing number of titles and user submissions.

b. Ensure that the system remains performant as the database grows.

3. Accuracy:

a. The system provides consistent results.

b. The system provides an accurate verification probability score.

4. Performance:

a. Title verification is completed within a reasonable time frame (e.g., under 2 seconds per title).

b. The system can handle multiple title verification requests simultaneously without significant performance degradation.

5. User Experience:

a. Users receive clear and actionable feedback on why their title was rejected.

b. Users see a probability score indicating the likelihood of their title being verified.

c. The interface for title submission and feedback is user-friendly and intuitive.

6. Robustness:

a. The system handles edge cases and variations in spelling effectively.

b. The system is resilient to errors and provides meaningful error messages when issues occur.

Software Features:

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

1. Title Applicant:

- User Account Management
  - Sign-up, login, logout, and password reset.
- Title Submission & Verification
  - Submit a new title for verification.
  - Receive feedback on title status (approved/rejected with reasons).
  - View similarity percentage and verification probability.
  - Flag rejected titles for review if mistakenly rejected
  - Modify and resubmit rejected titles.
  - Submit the final title after successful verification
- User Dashboard
  - Search and view existing titles
  - Track past submissions and their statuses.
  - View detailed rejection reasons and suggestions.

- Help & Support
    - Access FAQs and guidelines for title creation.
    - Contact support for queries.
- Settings & Customization
    - Allow users to set preferences.

2. PRGI Official (Regulatory Authority):

- Login & Role-based Access
    - Secure authentication with role-based permissions.
    - Sign-up, login, logout, and password reset.
- Restricted Word & Affix Management
    - Add, update, or remove disallowed words.
    - Manage disallowed prefixes and suffixes.
- Title Review & Oversight
    - Review flagged title submissions for possible manual intervention.
    - Override system decisions when necessary.
- Monitoring & Logs
    - View Existing Titles
    - View submission trends and history.
    - Access reports on rejected and accepted titles.
- Help & Support
    - Access FAQs and guidelines for title creation.
    - Contact support for queries.
- Settings & Customization
    - Allow users to set preferences.

3. System Administrator:

- User & Role Management
    - Log In, Reset Password, Log out
    - Add/remove PRGI officials.
    - Manage user access and permissions.
- System Configuration & Optimization
    - Maintain and optimize the title verification engine.
    - Manage database indexing and performance.
- Audit & Security Monitoring
    - Track all user activities and changes.
    - Ensure data security and compliance with regulations.
- Settings & Customization
    - Allow users to set preferences.

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
- Receive feedback on rejected titles
- View Verification Result
- Modify and Resubmit Title
- Register Title
- Flag rejected titles
- Search and view existing titles
- Track Submission History
- Manage preferences in settings
- Access Help & Guidelines
- Contact Support

2. PRGI Official:

- Sign Up
- Log In
- Log Out
- Reset Password
- Manage Disallowed Words
- Manage Disallowed Affixes
- Set/Modify Acceptance Probability
- Search and view existing titles
- Review Flagged Title Submissions
- Override System Decisions
- View Submission Reports & Trends
- Manage preferences in settings

- Access Help & Guidelines
- Contact Support

3. System Administrator:

- Log In
- Log Out
- Reset Password
- Manage User Accounts
- Configure System Settings
- Monitor & Optimize Database Performance
- Audit User Activity & Logs
- Manage User Support Requests
- Ensure Security & Compliance
- Manage preferences in settings
