18 April 2025

English

United Nations Group of Experts on Geographical Names 2025 session New York, 28 April – 2 May 2025 Item 5 (c) of the provisional agenda * Technical expertise: Writing systems and pronunciation

Addressing inconsistencies in Romanization: Towards an integrated program for the Unified Arabic System for the Romanization of Names

Submitted by Sultanate of Oman **

Summary:

The Romanization of Arabic geographical names presents an ongoing challenge due to the variety of transliteration systems in use. This multiplicity has led to widespread inconsistencies in the representation of Arabic names in Latin script, causing communication issues, particularly in geographical, geospatial, and administrative contexts. To address this, the current paper proposes the development of an integrated, software-based solution grounded in the Unified Arabic System for the Romanization of Names. This system, endorsed by the United Nations, is valued for its clarity, linguistic coherence, and practical application. The proposed program incorporates algorithmic logic based on phonetic and grammatical rules from both Arabic and English, offering a reliable means of producing consistent Romanized outputs. The application allows for both manual input and bulk processing via Excel files, making it accessible to specialists across sectors. Its reverse transliteration feature and linguistic adaptability further widen its potential use cases, making it an innovative and much-needed contribution to global efforts in data standardization and linguistic technology.

^{*} GEGN.2/2025/1

^{**} The full report was prepared by Hilal Rashid Al-Rajhi National Survey and Geospatial Information Authority (NSGIA), Ministry Of Defence, Oman

Introduction

Arabic is a language with deep historical roots and wide geographic distribution. Its rich phonetic system and unique orthography make the task of transcribing its words into the Latin alphabet particularly complex. Despite this, the demand for Romanized Arabic names continues to grow in domains such as cartography, international diplomacy, transportation, tourism, and digital mapping. The crux of the problem lies in the absence of a unified, standardized method for transliteration. Over time, Multiple systems have emerged, each shaped by different linguistic, national, or institutional priorities. As a result, a single Arabic place name may appear in various Romanized forms, creating confusion and hindering consistency in official documents, databases, maps, and search engines. These inconsistencies not only affect cross-border collaboration but also challenge researchers and specialists who rely on precise geospatial data and name referencing.

This paper aims to confront this issue by presenting a technological solution based on the United Nationsendorsed Unified Arabic System for the Romanization of Names. The project proposes a user-friendly, algorithm-powered program that can transliterate Arabic geographical names into Latin script with a high degree of accuracy, transparency, and repeatability. It is designed not just as a theoretical framework but as a practical tool, adaptable and extensible to serve multilingual and multidisciplinary audiences.

Background and Context

Numerous systems have historically been developed to Romanize Arabic script, among them the systems of the American Library Association–Library of Congress (ALA-LC), the International Organization for Standardization (ISO 233), and the BGN/PCGN system used by American and British geographic authorities. While each system presents its own linguistic rationale, they differ significantly in execution. For example, the Arabic name "القاهرة" may be rendered as "al-Qāhirah," "Al-Qahira," or even "El Kahira" depending on the chosen standard. Such variation leads to multiple entries in databases, confusion in digital search results, and discrepancies in international documents or maps.

The United Nations, through its Group of Experts on Geographical Names (UNGEGN), has long advocated for the adoption of a single, consistent approach to Romanizing geographical names. The Unified Arabic System for the Romanization of Names was introduced as a means to harmonize transliteration practices across the Arabic-speaking world and beyond. While the system has been well received and endorsed, especially within academic and governmental circles, its practical implementation—particularly through digital means—remains limited. This gap between theoretical standard and applied use forms the foundation of the proposed project.

Problem Definition and Justification

The current transliteration landscape is fragmented and lacks a unified tool that can implement the UNendorsed system in a scalable, accessible manner. The consequences are not trivial: inconsistent place name representations can result in mismatched data across GIS platforms, hinder communication in transport and logistics, and cause confusion in the dissemination of official documents and international agreements. Existing tools, if available, often require specialized training or fail to adhere to a single standard.

There is a critical need for a software-based transliteration tool that integrates linguistic rules with computational efficiency. Such a program must be robust enough to handle large datasets and flexible enough to allow for both manual and automated workflows. Furthermore, it should be designed with the user in mind, offering a seamless interface for experts in geospatial sciences, public administration, international organizations, and language studies.

Research Objectives

The central objective of this project is to develop a comprehensive software program that transliterates Arabic geographical names into Latin script using the Unified Arabic System. The program is envisioned as a bridge between linguistic theory and digital utility. It will feature core capabilities such as phonetic and grammatical rule enforcement, user-driven input (both direct and file-based), and immediate display of results. In addition to forward transliteration (Arabic to Latin), the program will support reverse functionality, allowing Latin-script names to be reconstructed into Arabic—a feature that enhances its relevance in multilingual and archival settings.

The tool will also prioritize accessibility and usability. Users should be able to load an Excel spreadsheet containing Arabic names, process the entire file in seconds, and receive a cleaned, Romanized output suitable for publication or further analysis. These objectives, while technologically challenging, are made achievable by grounding the development process in both rigorous linguistic study and modern software architecture principles.

Methodology

The process began by researching existing applications that deal with the Romanization of Arabic names and thoroughly examining the document titled "Unified Arabic System for the Romanization of Names", which is endorsed by the United Nations. This document served as a foundational reference, offering clarity on the principles and standards that should guide the transliteration process.

Following this initial research, I conducted an in-depth study of the Romanized datasets available at the National Survey and Geospatial Information Authority (NSGIA), with a particular focus on understanding the transliteration dictionary employed. After gaining clarity on the existing structure, I compiled all relevant components of the Romanization process, including the fonts used and other technical factors that influence the rendering of Arabic geographical names in Latin script.

Subsequently, I evaluated possible logical approaches to address the Romanization inconsistencies and identified the most suitable algorithm that could effectively support the selected approach. For the development phase, I chose C# as the primary programming language due to its flexibility and robustness, and utilized Microsoft Access for data storage and management.

The application was then developed, tested, and its output compared against the names dataset to validate accuracy. Based on these tests, the algorithm was iteratively improved to align with any updates or changes in the transliteration dictionary.

The final result is a user-friendly application capable of accurately Romanizing Arabic geographical names. Moreover, it empowers users to update the transliteration dictionary themselves with minimal need to modify the application's core code or algorithm, ensuring adaptability and long-term usability.

Program Design and Features

At the heart of the program is an algorithm that processes Arabic script based on the unified Romanization rules, producing consistent Latin-script output. The interface will allow users to enter single Arabic names manually or upload Excel spreadsheets for batch processing. Immediate transliteration feedback will be provided, and users will be able to export results in Excel formats.

A critical innovation is the inclusion of reverse transliteration, allowing users to input Latin-script names and retrieve their most likely Arabic equivalents. While more complex due to inherent ambiguities in Romanized forms, this feature expands the program's utility for translation work, linguistic studies, and data integration.

The design will emphasize usability, ensuring that the program can be employed with minimal training. The interface will accommodate both technical and non-technical users, and special attention will be given to multilingual support, scalability, and platform compatibility.

Expected Results and Impact

Upon completion, the program is expected to provide a highly accurate, reliable, and accessible tool for the Romanization of Arabic geographical names. Its adoption will lead to improved standardization across documents, databases, and geospatial systems. In turn, this will enhance data interoperability and contribute to more cohesive international collaboration in areas such as mapping, logistics, and regional planning.

The software also has educational and research value, offering linguists and students a clear and functional demonstration of transliteration rules in action. Its flexible architecture allows it to be adapted to other languages or Romanization systems, positioning it as a foundation for future linguistic tools.

Furthermore, the reverse transliteration feature opens new possibilities for bilingual document preparation, digital archiving, and machine translation enhancement. As such, the project contributes not only to solving a technical problem but also to enriching the broader field of computational linguistics.

Conclusion

Furthermore, the project offers a practical solution to the challenges posed by varied Romanization systems and enhances information exchange in multiple domains, including geomatics, tourism, and transportation. By adhering to precise scientific principles and offering a flexible user interface, this program becomes a valuable tool for both specialists and researchers alike. Additionally, the application holds significant potential for broader linguistic utility, as the robust algorithm it employs could be adapted for use with other languages beyond Arabic. With minor adjustments to the application's internal settings, the same logic and algorithmic framework can be repurposed to support the Romanization of other languages, greatly expanding its scope and relevance. Moreover, its algorithmic foundation allows for the development of reverse transliteration capabilities, enabling the conversion of names from Latin script back into Arabic, further enhancing its usability and versatility.

Points for discussion

The Group of Experts is invited to:

- (1) Propose any ideas that may serve enhancement of the application
- (2) Ask about software development process
- (3) Ask about application's potential capabilities
- (4) Supply any kind of resources that may help developing the application
- (5) Consider this application as future toponymic research