

Romenagri Transliteration System Release 2.1 (Linux) Version 0.1.2 Alpha User Guide and Reference Manual

Copyright (C) 2006 Abhishek Choudhary

Permission is granted to copy, distribute and / or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with Invariant Section being Technical Description, no Front-Cover Texts, and no Back-Cover Texts.

ALL TRADEMARKS ACKNOWLEDGED AS BELONGING TO THEIR RESPECTIVE OWNERS

THIS DOCUMENT DOES NOT IMPLY ANY WARRANTY AS TO THE ACCURATENESS OF CONTENTS OR ANY COMMITMENT TO PROVIDE ANY SERVICE WHATSOEVER. IT IS ONLY INTENDED TO FACILITATE THE USER EXPERIENCE, HOWEVER NO FITNESS FOR ANY PURPOSE IS GUARANTEED, NOT EVEN THE FITNESS FOR MERCHANTABILITY.

रोमनागरी अनुवादन प्रणाली

~~~~~

परिचय

~~~~~

१. मूल खातानाम प्रविष्टि खण्ड

जिस खाते में अनुवादन के लिए मूल पाठ है, वह यहाँ प्रविष्टि करें। मूल निवेश बटन दबाने पर वह खाता मूल पाठ खण्ड में दिखेगा।

२. मूल खातानाम चयन बटन (...)

मूल खातानाम चयन करने के लिए यह बटन दबाएँ। चयनित खातानाम मूल खातानाम प्रविष्टि खण्ड में स्वतः प्रविष्टि हो जाएगा। मूल निवेश बटन दबाने पर वह खाता मूल पाठ खण्ड में दिखेगा।

३. मूल निवेश बटन

यह बटन दबाने पर चयनित खाता मूल पाठ खण्ड में दिखेगा।

४. मूल पाठ खण्ड

इस खण्ड में मूल पाठ लिख सकते हैं, अथवा निवेश कर सकते हैं। अनुवादन बटनों को दबाने पर अनुवादित पाठ अनुवादित पाठ खण्ड में दिखेगा।

५. देवनागरी से रोमनागरी अनुवादन बटन

यह बटन दबाने पर मूल पाठ खण्ड में प्रविष्टि देवनागरी पाठ का रोमनागरी अनुवादन अनुवादित पाठ खण्ड में दिखेगा।

६. रोमनागरी से देवनागरी अनुवादन बटन

यह बटन दबाने पर मूल पाठ खण्ड में प्रविष्टि देवनागरी पाठ का रोमनागरी अनुवादन अनुवादित पाठ खण्ड में दिखेगा।

७. अनुवादित पाठ खण्ड

अनुवादन बटनों को दबाने पर अनुवादित पाठ यहाँ में दिखेगा। इसे आप संचित कर सकते हैं, अथवा नकल कर अन्य अनुप्रयोगों में चिपका सकते हैं।

८. अनुवादित खातानाम प्रविष्टि खण्ड

जिस खाते में अनुवादित पाठ संचय करना चाहते हैं, वह यहाँ प्रविष्टि करें। अनुवाद संचय बटन दबाने पर अनुवादित पाठ उस खाते में संचित हो जाएगा।

९. अनुवादित खातानाम चयन बटन (...)

अनुवादित खातानाम चयन करने के लिए यह बटन दबाएँ। चयनित खातानाम अनुवादित खातानाम प्रविष्टि खण्ड में स्वतः प्रविष्टि हो जाएगा। अनुवाद संचय बटन दबाने पर अनुवादित पाठ उस खाते में संचित हो जाएगा।

१०. अनुवाद संचय बटन

यह बटन दबाने पर अनुवादित पाठ चयनित खाते में संचित हो जाएगा।

११. सहायता बटन

सहायता के लिए यह बटन दबाएँ।

१२. निकास बटन

निकास के लिए यह बटन दबाएँ।

Technical description

~~~~~

(Ref. Choudhary Abhishek, Karwa Sweta. "Hindawi - A Software Development System for North Indian Vernaculars", Presented at CSI YITPA (East) 2004, Science City, Kolkata)

Romenagri is a non-ambiguous invertible case and diacritic independent compiler acceptable transliteration system with the associated algorithm and has been the sole basis for much of the project's motivation. The authors have independently developed it and demonstrated it to be applicable to all languages using the North Indian composite syllabic scripts - viz. Assomiya, Bangla, Devnagri, Gujrati, Oriya and Punjabi. Romenagri utilises syllabic complements in Roman script for the symbols of the North Indian scripts. The mapping for a specific script may be a subset of the complete mapping owing to the absence of certain characters in the specific case, e.g. the wa and ba of Devnagri match a single symbol in Bangla ba. The words are formed by actively concatenating successive syllabic compliments, looked up from a table through an O(1) search achieved by using the normalised code for the Indian script symbol as an array index. The process of active concatenation uses the "de-voweling" operator carat (^), which forms an equivalent of halanta or hasanta of the Indian scripts and distinguishes the matra of the vowels by preceding the syllabic compliment of their akshara form. The syllabic compliment looked up from the mapping table is pushed onto a stack. On encountering a carat as part of a looked-up compliment, the last pushed vowel

character “a” is popped out of the stack and discarded. The remaining part of the compliment, after the carat, is then pushed onto the stack. On encountering the end of a word, the content of the stack is popped to obtain the required transliteration, after which the stack is flushed.

The process of converting Romenagri back to the Indian script representation is more complex and is achieved by using a recursive descent parser. The authors have designed the syllabic compliment so as to facilitate  $O(n)$  parsing. The parser operates at 5 levels. The word is submitted at level 1, and the initial syllabic compliments are consumed. Successive levels are entered in case of multiple possibilities with the ultimate level identifying a matra. All other symbols are identified at earlier levels. After each production the parser enters level 1 with the non-consumed part of the input.

The only phonetic modifier used in Romenagri is the underscore “\_” character, which generally forms a part of the input set of most compilers. This allows rule adherent transliteration for keywords written in Indian scripts. The underscore characters present in the original Indian script text are expanded to two underscore characters. Hence, the inversion parser treats every paired underscore as a character and every nascent underscore as a phonetic modifier. Figure-1 gives an instance of Romenagri transliteration.

क + ्र + ि + य + ा = क्रिया

ka + ^ra + ^i + ya + ^aa = kriyaa

Figure 1: An instance of Romenagri transliteration with corresponding syllabic compliments