

## ANUVADAKSH – An Expert English to Indian Language Machine Translation System

**ANUVADAKSH: An Expert English to Indian Languages Machine Translation System [EILMT]** is a state-of-the-art solution that allows translating the text from English to Indian languages. ANUVADAKSH MT has been developed by a consortium of 10 Indian academic and research institutes, led by CDAC-Pune with the support of “Technology Development for Indian Languages” (TDIL) Programme of the Department of Electronics & Information Technology (DeitY), Government of India.

The English to Indian Languages MT system is being developed for the 6 language pairs i.e. English to Hindi, Bengali, Marathi, Urdu, Tamil and Oriya. This system is developed to facilitate the multi-lingual community, initially in the domain-specific expressions of Tourism, and subsequently it would foray into various other domains as well, in a phased manner.

ANUVADAKSH MT Systems have been released on TDIL Data centre [www.tdil-dc.in](http://www.tdil-dc.in) for public use and feedback.

This is a collaborative effort of the consortium institutes which have brought forward the integration of four Machine Translation Technologies – TAG (Tree-Adjoining-Grammar based MT), SMT (Statistical based MT), AnalGen (Rules-Based MT) and EBMT (Example Based MT).

The screenshot displays the web interface of the ANUVADAKSH MT System. The browser address bar shows the URL: [http://www.tdil-dc.in/components/com\\_mtsystem/CommonUI/homeMT.php](http://www.tdil-dc.in/components/com_mtsystem/CommonUI/homeMT.php). The page header includes the TDIL logo and the text "Indian Language Technology Proliferation and Deployment Centre" and "ILTP-DC भारतीय भाषा प्रौद्योगिकी प्रसरण एवं विस्तारण केन्द्र". The interface features a navigation menu with "Home", "Contributors", "Feedback", and "Help". The main content area shows a text input field with the English sentence: "Qutub Minar is the pride of Delhi. The tall minaret was constructed in 1192 by Qutab-ud-din Aibak, and later completed by his successor Iltutmish. Qutub Minar is a World Heritage Site and has survived the ravages of time impressively. The Minar of Delhi is surrounded by a lush green garden, which is an ideal leisurely place for visitors. Qutub Minar is the favourite destination of tourists." Below the input field is a "Translate" button. The translated output is displayed in Hindi: "कुतुब मीनार दिल्ली का गौरव है लम्बी मस्जिद की मीनार कुतुब उड कोलाहल ऐबक द्वारा 1192 में निर्मित की गई थी और उसके क्रमानुयायी इल्तुमिश द्वारा बाद में पूरा की कुतुब मीनार वर्ल्ड हैरिटेज साइट है और समय के विनाश प्रभावशाली ढंग से अस्तित्व में हो गया है दिल्ली की मीनार घना हरा भरा उद्यान द्वारा चारों ओर से घिरा है जो पर्यटकों के लिए आदर्श धीरे धीरा स्थान है कुतुब मीनार पर्यटकों का प्रिय गंतव्य है". The output is presented in a table with "more" buttons for each line. At the bottom, there is a footer with copyright information: "© 2009-10 Department of Information Technology, MCIT, Govt of India." and logos for W3C XHTML 1.0 and W3C CSS.

A common GUI for the ease of users is provided. The screens shows a glimpse of English-Hindi MT System

## ANUVADAKSH E-IL MT System - Technical Features

- Various Communication Protocols as
  - Transaction via TDIL web portal
  - User friendly Graphical User Interface
  - The user can either upload and edit a file/document or type directly into the text area of the system, for translation
- Different formats i.e. .rtf, .xls, .txt and .html are supported - Pre-Processing module, prepares input text into engine suitable form.
- The System is designed to use four-translation engines working in parallel namely AnalGen, EBMT, SMT & TAG. The Collation & Ranking Module brings the best translation.
- Post processing module provides additional features for EILMT Translation engine like
  - Morph Synthesizer for smoothening the translated output
  - Multiple translation options
  - Synonym selection option
  - Typing facility in Target Language for all the six languages
  - Transliteration Facility
  - Retaining the original format of English text
- W3C compliant System
- Cross Browser compatibility for IE, Mozilla Firefox, Google Chrome, Apple Safari and Opera

## Possible Use

The ANUVADAKSH E-IL MT Systems have reached a stage of development where some of these can be made available to the users. There are various deployment possibilities.

- A) The systems could be made available on the web as a service for general purpose use. User can copy–paste a text and get instant translation on demand.
- B) MT as a service may be provided on some popular websites and could get the web page's content translated into a language of their choice.
- C) The systems can be used by making it available as a translator's help desk to the professional translators both individuals as well organizations.

### Consortia Leader:

Dr. Hemant Darbari, C-DAC Pune,

### Consortia Members:

C-DAC - Mumbai, IIIT - Hyderabad, IISc - Bangalore, IIT - Mumbai, Jadavpur University - Kolkata, , IIIT - Allahabad  
Amrita University - Coimbatore, Banasthali Vidyapeeth – Banasthali, Utkal University – Bhubaneswar

### Support:

Technology Development for Indian Languages Programme  
Department of Electronics & Information Technology  
Government of India