

Request for Proposals for Development of Tools & technologies for Text Mining from Unstructured Natural Language Text for Indian Languages

TDIL Programme, Dept. of Electronics & Information Technology, Govt. of India invites project proposals for the development of Tools & technologies and modules for Text Mining from Unstructured Natural Language Text for Indian Languages.

1. Introduction

Department of Electronics & Information Technology , Ministry of Communications & Information Technology Govt. of India promotes e-Development of India through multi pronged strategy of e-Infrastructure creation to facilitate and promote e-governance, promotion of Electronics & Information Technology- Information Technology Enabled Services (IT-ITeS) Industry, providing support for creation of Innovation / Research & Development (R&D), building Knowledge network and securing India's cyber space. [Details at: <http://www.deity.gov.in>]

2. TDIL Programme & Its Objective

Technology Development for Indian Languages (TDIL) Programme is a Flagship Programme of Department of Electronics & Information Technology (DeitY), Ministry of Communication & Information Technology (MC&IT), Govt. of India has the objective of developing Information Processing Tools and Techniques to facilitate human-machine interaction without language barrier; creating and accessing multilingual knowledge resources; and integrating them to develop innovative user products and services. The Programme also promotes Language Technology standardization through active participation in International and national standardization bodies such as ISO, UNICODE, World-wide-Web consortium (W3C) and BIS (Bureau of Indian Standards) to ensure adequate representation of Indian languages in existing and future language technology standards. [<http://www.tdil.mit.gov.in>]

Objectives:

- Research and Development of Multilingual Computing in all 22 Constitutionally Recognized Languages
- Proliferation of Language Technology Solutions
- Development of Standards related to Language Technology

3. Problem Definition & Broad Objective

Problem Definition:

- i Text mining is "the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources".
- ii Primary component of text mining is information extraction (IE) from unstructured text (natural language text). Two important goals of text mining systems are identification of relevant objects ("entities") and extraction of events in which these entities participate and influence one another.

- iii Another important goal of the Text Mining Systems is identifying of responses such as emotions or reactions to events or artifacts.
- iv “Shared Tasks” is an important component for assessment and improvement of the text mining methods and technologies and to propose remedial measures to address gaps.

Broad Objective:

- i To develop complete Text Mining System (TMS) and/or various functional modules of a TMS which may be used for the tasks mentioned above.
- ii To develop corpora in the following domains for organizing shared tasks as mentioned in point iv above.
 - 1) sentiment analysis from Twitter feeds
 - 2) public health reports from leading newspapers
 - 3) agricultural reports from newspapers and trade journals

The availability of text and data mining systems may lead to the development of technology businesses, increase research productivity and quality and, on a wider social scale, to lead to the discovery of new treatments for serious illnesses, technical intelligence, counter-surveillance and related areas, etc.

A background document on Text Mining may be referred at: _

<http://tdil.mit.gov.in/pdf/RFP/Background%20Paper%20-%20Text%20Mining.pdf>

4. Proposal Details

Comprehensive project proposal may be submitted clearly specifying the technological problems to be addressed, expected outcome, deployment plan and budget requirements. System developers have to participate in one of the shared tasks enumerated above. The project proposal should contain the following details

Technical Details

- 1) Executive summary
- 2) Theoretical background and implementation details
- 3) Year-wise statement of deliverables
- 4) Proposed national and international collaborations (if any)
- 5) System developers should state at least one mandatory shared task they will participate in

Budgetary Details

- 1) Year-wise budget breakup
- 2) Specification of required manpower and equipment

5. Evaluation of the Project Proposal

Project proposals will undergo rigorous evaluation by a National Level Expert Committee at the first stage, followed by evaluation by the TDIL Working Group and DEITY Officials.

6. Who can Apply

An individual or Consortium of Indian University/ Institution/ R &D Organization may submit the project proposal. A Consortium Mode project may involve small, medium or large Industrial Partners having DSIR* recognized in-house R&D unit(s) and Indian Universities, Institutions and R&D Organizations.

7. Proforma: The details can be downloaded from:

<http://deity.gov.in/content/scheme-forms>

8. Last Date:

TDIL Programme invites proposal round the year. However, to be considered in the next meeting of the TDIL Working Group, the last date of submission of project proposal is 31st March 2015.

9. How to Apply

Completed Project Proposal both Soft and Hard-Copies(3) duly signed by the Chief Investigator and Head of Institution may be submitted to:

Swaran Lata, Director

TDIL Programme,

Dept. of Electronics & Information Technology, Govt. of India

Electronics Niketan, 6 CGO Complex, Lodhi Road,

New Delhi 110 003 e-mail: slata@deity.gov.in

Tel: +91-11-24363525, 91-11-24301272 Fax: +91-11-24363525