*Under 2) The EOI should carry the following information (details in the attached doc):*

Title: Application Form for Expression of Interest (EOI) for Technology Transfer/Commercialization of “Indian Language Machine Translation System (ILMT) Phase-II”.

"The institute invites an Expression of Interest (EOI) for the below listed Language Technology resources developed at the Language Technologies Research Center (LTRC), IIIT Hyderabad as a part of the Indian Languages Machine Translation (ILMT) consortium sponsored under the Technology Development for Indian Languages (TDIL) Programme initiated by the Department of Electronics & Information Technology (DeitY), Ministry of Communication & Information Technology (MC&IT), Govt. of India (GoI).

The EOI should be in the format given in ANNEX A. The companies/startups to be granted license to utilize the transferred technology on an as-is basis will be shortlisted based on the information furnished per the requirements as stated in the Annexures/Form, the supporting documents with the application and assessment by the institute committee.

The submission of the EOI shall include all such documents that are specified herein and applicable to prove the authenticity of the entity and any claim made therein. The burden of proving such claims shall lie with the bidder.

Any cost and expenses associated with submission of EOI, as maybe applicable shall be borne by the company while submitting the EOI and the Institute shall have no liability, in any manner in this regard, or if it decides to terminate the process of short listing for any reason whatsoever.

Signed on behalf of Grantee Institute Authorized Signatory:

Date:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Under 3) Scope / Resources Available <To be specified>*

a) Introduction

(i) Brief about the Institute/Research Center:

* + - **Institute**: The International Institute of Information Technology, Hyderabad (IIIT-H) is an autonomous university founded in 1998. It was set up as a not-for-profit public private partnership (N-PPP) and is the first IIIT to be set up under this model in India. The Government of Andhra Pradesh lent support to the institute by grant of land and buildings. A Governing Council consisting of eminent people from academia, industry and government presides over the governance of the institution.

IIIT-H was set up as a research university focused on the core areas of Information Technology, such as Computer Science, Electronics and Communications, and their applications in other domains. The institute evolved strong research programmes in a host of areas, with computation or IT providing the connecting thread, and with an emphasis on the development of technology and applications, which can be transferred for use to industry and society. This required carrying out basic research that can be used to solve real life problems. As a result, a synergistic relationship has come to exist at the Institute between basic and applied research. Faculty carries out a number of academic industrial projects, and a few companies have been incubated based on the research done at the Institute.

Our decade long work at IIIT-H has shown how an academic institute can build large research groups to carry out cutting edge research, and use it to solve real-life industrial and societal problems.

IIIT-H is organized as research centres and labs, instead of the conventional departments, to facilitate inter-disciplinary research and a seamless flow of knowledge within the Institute. Faculty assigned to the centers and labs conduct research, as well as academic programs, which are owned by the Institute, and not by individual research centers.

* **Research centre**: The Language Technologies Research Centre (LTRC) addresses the complex problem of understanding and processing natural languages in both speech and text modes.
* LTRC conducts research on both basic and applied aspects of language technology.
* It is the largest academic centre of speech and language technology in India.
* LTRC carries out its work through four labs, which work in synergy with each other.

LTRC is also a lead participant in nation-wide mission-mode consortia projects to develop deployable technology in the areas of Indian Language Machine Translation, English to Indian Language Machine Translation, and Cross Language Information Access (search engines).

**NLP-MT Lab**

The NLP-MT lab does fundamental work on developing grammatical as well statistical modelling of languages. Linguistic approaches are combined with machine learning techniques leading to new theories and technology development. This has resulted in higher accuracy parts-of-speech taggers, chunkers, constraint-based parsers as well as broad coverage statistical parsers, and semantic analyzers for Indian languages on the one hand, and annotated data including dependency tree banks, discourse banks, parallel corpora, etc. on the other.

**Anusaaraka Lab**

Anusaaraka lab is concerned with the development of machine translation systems which in addition to the usual machine translation output also allows a user to understand the source language text in a pseudo target language. For example, a reader who knows Hindi (target language) would be able to read the English source text, in a pseudo Hindi output after a small amount of training.

[**Search and Information Extraction Lab (SIEL)**](http://search.iiit.ac.in/)

The Lab focuses on solving research problems in the areas of Information Retrieval and Extraction using NLP techniques. SIEL is engaged in building technologies for personalized, customizable and highly relevant information retrieval (IR), Information Extraction (IE) and Information Access (IA) systems. Current research includes summarization, cross language information access systems and building semantic search engines by identifying entities and relations between these entities.

**Speech lab**

The research focus of the speech group is to address issues in processing, analyzing, understanding and manipulation of “real” speech i.e., read, conversational, emotional as found in real-environments for the development of robust speech systems in multiple languages specifically in the context of India. The speech group is also focused on development of speech recognition, speech synthesis, prosody models, spoken audio search, phonetic engine for Indian languages, language identification, speaker recognition for biometrics, voice conversion, speech summarization and spoken dialog systems.

(ii) Brief description about the Product/Technology/Prototype to be transferred:

The institute wishes to seek technology transfer for the following products and technologies developed under ILMT project:

1. Indian Language to Indian Language Machine translation systems: In the project, 18 MT systems between nine Indian languages were developed. All the 18 systems are deployed at IIIT-H’s server for internal access. Eight of the systems - Punjabi-Hindi, Hindi-Punjabi, Urdu-Hindi, Marathi-Hindi, Telugu-Hindi, Tamil-Hindi, Tamil-Telugu, and Telugu-Tamil - are now available and may be accessed at [http://sampark.org.in](http://sampark.org.in/).

The Indian language to Indian language machine translation systems are based on hybrid analyse- transfer-generate paradigm. First, analysis of the source language is done, then a transfer of lexical and grammatical structure from the source language to the target language is carried out, and finally the target language is generated. The larger task of system building was subdivided into smaller tasks. Language independent software engines were built commonly (as horizontals) and were used by the different language groups for their language by supplying data, resulting in major savings in effort and higher software quality. Each participating institute took up one or more tasks. Thus, it was possible to develop MT systems for several language pairs in parallel by developing common technologies and sharing tasks.

The IL-IL Machine translation systems available currently are :

* 1. Punjabi-Hindi,
  2. Hindi-Punjabi,
  3. Urdu-Hindi,
  4. Hindi-Urdu,
  5. Marathi-Hindi,
  6. Hindi-Marathi,
  7. Telugu-Hindi,
  8. Hindi-Telugu,
  9. Tamil-Hindi,
  10. Hindi-Tamil,
  11. Tamil-Telugu,
  12. Telugu-Tamil,
  13. Hindi-Kannada,
  14. Kannada-Hindi,
  15. Tamil-Malayalam,
  16. Malayalam-Tamil,
  17. Hindi-Bangla,
  18. Bangla-Hindi

1. Indian Language Analysis and generation tools: Apart from the complete translation pipelines for IL-IL Machine Translation, specific tools developed for analysis of the languages are also available. These tools include
   1. Indic Tokeniser
   2. Morph Analyser, POS tagger, Chunker and simple parser (for all mentioned languages)
   3. Full parser for Hindi and Urdu)
   4. NER and MWE identifier for Hindi
   5. Generation pipeline for all mentioned pipelines

(iii) Current status of Product/ Technology/Prototype:

All the available products are in deployable stage.

\*Points of Note:

1. The resources will be shared on an as-is-where-is condition to interested parties who are currently pursuing research and/or planning to utilize the available resources for further advancement of technology in Indian languages. The resources may be licensed for a fee or on a free-of-cost basis depending on the purpose of usage type or resource, validity of application time period and any other conditions as applicable. The grantee institute will have the final authority to decide on the application on a case-to-case basis.
2. The resources shared cannot be utilized directly or indirectly for any commercial purpose, unless explicitly stated to and having the written approval by the granting institute.
3. No additional support will be provided on any of the resources.
4. At any time before the submission of EOI, the grantee institution may carry out amendment(s) to this EOI document. The amendment will be made available on the website (www.iiit.ac.in) and will be binding on them.
5. The grantee institute reserves the right to accept or reject any application without assigning any reason thereof.
6. EOIs that are incomplete in any respect or those that are not consistent with the requirements as specified here or those that do not adhere to formats, wherever specified may be considered non-responsive and may be liable for rejection and no further correspondences will be entertained with such organizations/entities.

For any clarifications on the Expression of interest document, the following may be contacted through e-mail/FAX/Letter:

Details of the Contact Person(s) -

Competent Authority: Prof. Dipti Misra Sharma

Head of Language Technologies Research Center (LTRC)

Email:

Fax: +91-40-6653 1413

Address: Prof. Dipti Misra Sharma, LTRC, Vindhya Research Building,

International Institute of Information Technology (IIIT-H), Gachibowli,

Hyderabad 500 032 Telangana, INDIA.

Grantee Institute - IIIT Hyderabad

*\****EOI Form (Annexure A)** *- <with the following details>*

EOI Notice: <number, date, website, location>

EOI Title: Application Form for Expression of Interest (EOI) for Technology Transfer/Commercialization

Project Title: Indian Language to Indian Language Machine Translation System (ILMT) Phase-II

Affiliation: <to be mentioned, as stated in our Contract Notice>

EOI File Number: <as stated in the Contract Notice> (optional)

DeitY Sanction No. & Date: 11(1)/2010, HCC(TDIL), dated 27.12.2010

Name and complete Legal Address of the Grantee Institute: <as specified in our registration document>

The information to be furnished by the interested party for Expression of Interest (EOI) along with the supporting documents:

**PART-A Company Profile**

1. Name of Legal Entity or Entities:

2. Name of the Contact Person:

Address:

Telephone:

Fax:

E-Mail:

Website:

3. Year of Incorporation:

4. Type of Business Entity:

a. Educational Institute / Sole Proprietorship / Partnership / Limited Liability Partnership / Private Limited Company / Public Limited Company / Any Other

b. Whether Foreign Equity Participation (please give name of foreign equity participant and percentage thereof) Y/N

c. Names of Directors of the Board/ Proprietors (supporting document(s) to be attached, as listed in PART-B):

5. Category of the firm: Large / Medium / Small Scale Unit / Startup

6. Address of the Registered Office (supporting document(s) to be attached, as listed in PART-B)

7. Number of Employees (full-time):

8. Certificate of Registration (supporting document(s) to be attached, as listed in PART-B)

9. Permanent Account Number (supporting document(s) to be attached, as listed in PART-B)

10. Sales Tax Number/ VAT Number (supporting document(s) to be attached, as listed in PART-B)

11. Status of IS09001/ISO 13485 Certification (supporting document(s) to be attached, as listed in PART-B)

12. Nature of Business: Academic Institute / Research Organization / Manufacturing Unit / Services / Any Other (please specify) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

13. Turnover (supporting documents to be attached, as listed in PART-B)

14. Purpose for the EOI (how and where will the resources be utilized)

The Expression of Interest (EOI) is for participation of defined with the purpose for the EOI as listed:

a.

b.

c.

d.

**PART-B Essential Requirements**

1. The organization must be a reputed firm/company/SME/startup/R&D company incorporated in India with A standing of at least 2 years.

2. The turnover is to be supported by financial statements of accounts/ Annual reports duly certified by a Chartered accountant/ Balance sheets of last 3 years/ Income tax returns for the last 3 years period / Company PAN number.

3. Company profile, giving details of current activities and management personnel structure including evidence of incorporation. The company should be registered and ISO 9001/ ISO13485 or equivalent certified.

4. Details of absorption of technology for a product/ know-how that has been taken up on production scale in the past may also be given.

I hereby declare that the above information is true to the best of my knowledge.

Authorized Signatory on behalf of the Applicant:

Name & Designation:

Contact Details (Email, Phone, Postal Address for further communication):

Place: Date:

Company Stamp: