

PROFORMA FOR SUBMISSION OF PROPOSALS UNDER THE SCHEME OF RESEARCH & DEVELOPMENT

PART-I Information relating to department/Institute

1. Name of Institute with complete address:
Indian Institute of Technology, Bombay, Powai, Mumbai 400 076
Partner Institutes: i. Amrita University, Coimbatore
ii. Tamil University, Tanjore
iii. Dravidian University, Kuppam
iv. Central University, Hyderabad

2. Title of the Research Project:
Creation of Machine Translation tools and resources for English to Dravidian Languages:
Pilot Study

3. Department/Board Area:
Linguistics, Natural Language Processing

4. Major areas of research in the Department:

Artificial Intelligence, Algorithms, Databases, Networking, Natural Language Processing,
Theoretical Computer Science

5. Names & Designation of Principal Researches in the major areas and list of publication
during last 5 years based on work done in the Department :
 - i. Dr. Pushpak Bhattacharyya, Professor, Department of Computer Science and Engineering,
Indian Institute of Technology, Mumbai
 - ii. Dr. Soman KP, Professor & Head, CEN, Amrita University, Coimbatore
 - iii. Dr. S. Rajendaran, Professor and Head, Dept. of Linguistics, Tamil University, Thanjavur
 - iv. Dr. S. Arulmozi, Asst. Professor, Department of Comparative Dravidian and
Computational Linguistics, Dravidian University, Kuppam
 - v. Dr. K. Narayana Murthy, Professor and Head, Department of Computer Science, Central
University, Hyderabad

(Please refer resume's attached in the Appendix)

6. Is it Inter-disciplinary Project?: Yes.

7. Is it Inter-Institutional Project?: Yes.

8. Is any Industry/User agency participating?: No

9. Brief of completed and or ongoing research projects supported by MHRD/AICTE in the
Department during last 5 years:

Will be submitted in two days time.

PART-II Information relating to department/Institute

10. PRINCIPAL INVESTIGATOR

- a. Name : Dr. Pushpak Bhattacharyya
b. Designation : Professor
c. Age : 47
d. Educational Qualification : PhD
e. Areas of specialization : Natural Language Processing,
Machine Learning
f. Experience (Teaching & Research) : About 25 years
g. Selected list of papers published : Please see the CV in the appendix

11. In case it is a joint project with other Institution, research labs and industries, name(s) of participating investigators.

Sr. No.	Name	Institution	Email Id
1	Dr. Soman KP	Professor and Head, CEN, Amrita University, Coimbatore	kp_soman@amrita.edu
2	Dr. S. Rajendaran	Professor and Head, Dept. of Linguistics, Tamil University, Thanjavur-613 010.	raj_ushush@yahoo.com
3	Dr. S. Arulmozi	Asst. Professor, Department of Comparative Dravidian and Computational Linguistics, Dravidian University, Kummpam	arulmozi@gmail.com
4	Dr. K. Narayana Murthy	Professor and Head, Department of Computer Science, Central University, Hyderabad	kavinarayanmurthi@yahoo.com

12. In case industry/user agency is participating, whether a MOU has been signed or letter of intent given.
Not Applicable
13. Present commitments of the Principal Investigator
- (a) Teaching : 6 hours per week
 - (b) Ph.Ds registered : 5 students (2 co-guided)
under him/her
 - (c) Sponsored research : Ministry of IT funded projects on Machine Translation
and Cross Lingual Search; TCS funded project on NLP
 - (d) Consultancy : --
14. Other members of the Research Group to work on proposed Projects:
- i. Dr. Soman KP, Professor and Head, Amrita University, Coimbatore
 - ii. Dr. Rajendran, Professor, Tamil University
 - iii. Dr. S Arulmozhi, Professor, Dravidian University, Kuppam
 - iv. Dr Narayana Murthy, Professor, Central University, Hyderabad

(Please refer attached resume for details)

PART-III About Research Project

15. (a) Summary of the project (in brief)

Our main objective is to develop Machine Translation(MT) system and needed linguistic resources for English-Dravidian languages(Tamil, Malayalam, Telugu and Kannada), that would facilitate the creation of rich educational contents in Indian languages. Our research effort is to make all the tools and translation system to be based on Machine Learning methodologies so that computer graduates and other such non-linguists are able to immediately participate in the national mission on literacy by contributing additional tools for language translation.

In **Module II**, we aim at developing teaching material corresponding to the tools we are developing (most tools are based on machine learning) so that it can be delivered as part of undergraduate computer science and engineering curriculum on data mining/machine learning. This will ensure a critical amount of man power required for sustaining translation effort needed for national mission on education.

Module II also aims at training 500 faculties selected from across the country on machine translation methodologies using machine learning techniques.

Module III aims at developing a Dravidian Wordnet required for translation. This module requires collaboration between different universities. This will also link to the Hindi wordnet developed at IIT Bombay (www.cfilt.iitb.ac.in) which is being widely used for NLP involving Indian languages.

(b) Justification, importance of project, etc.

Module I: Machine Translation (MT) between natural languages is one of the biggest challenge in Natural Language Processing (NLP) or Computational Linguistics (CL) research. MT research started in early 1950s' and made lot of progress for English and European languages whereas MT research in India started much later and yet to produce tangible results. As a partner in the larger government initiative English to Indian Languages: Machine Translation system(EILMT), Amrita has been into MT research for past 3 years and has gained significant experience developing linguistic tools and annotated corpora for Tamil and Malayalam. Based on the experience we gained, we developed a working prototype MT system for English-Tamil and many linguistic tools and resources for Tamil.

The project aims at developing two sets of tools. One set is for assisting manual translation (Computer Assisted Translation tools or CAT tools) process and the other for semi-automated machine translation (MT tools).

Module II: First of all, computational linguistics has not become part of undergraduate curriculum of many Indian universities. Even in those universities where it is introduced, the course material is entirely biased towards linguistics of English language.

Secondly there is no undergraduate text book currently available that deals with computational linguistics of Indian Languages.

Thirdly we do not have open linguistic resources (for Indian Languages) and tools that can be used by students/teachers/researchers. Therefore this project aims at putting relevant theory, tools and data together so that the subject can be taught either as elective or as part of Data mining course to under graduate students.

A good mass of trained manpower in Computational linguistics will ensure continuous flow of free linguistic tools that facilitate creation or translation of scientific texts in all Indian Languages.

Societal need

Use of a language other than one's mother tongue in schools is commonplace. English is viewed as connecting children to the expanded opportunities of the global economy. But in villages, teaching in mother tongues makes education more accessible and gives native speakers an equal chance at succeeding in the final common examination.

The problems of students in villages are twofold.

a) No qualified teachers for teaching in English

According to a report in Hindu Daily, when a study was conducted among school teachers in Maharashtra, it was found that many of English teachers do not know the answers to the exercise-questions given in the text book they teach. Also most teachers do not have fluency in English.

b) Do not get help from parents as they do not know English

In this context availability of learning materials (including the latest developments in Science) in local languages will go a long way in achieving social justice.

Module III: Wordnet is an online lexical reference system. Its design is inspired by current psycholinguistic and computational theories of human lexical memory. Nouns, verbs, adjectives and adverbs are organized into synonymous sets, each representing one underlying lexicalized concept. Different semantic relations link the synonyms sets. Wordnet is neither a traditional dictionary nor a thesaurus but combines features of both types of lexical reference resources. The most ambitious feature of Wordnet is the organization of lexical information in terms of word meanings rather than word forms. The main objective of DWN is to develop an extensive and high quality multilingual database with wordnets for Dravidian languages in a cost-effective manner, and also to link it with the existing prominent Wordnet of languages like Hindi, English and other languages. The project will also develop a language independent set of semantic concepts linking the language networks

together. The resources will be field tested for adequacy in an information retrieval application. The ultimate objectives are to move toward standardization of semantic classification of information for all Dravidian languages and to provide resources for development of applications, which can operate in a selected language or over a range of languages.

(c) Details of the work already done by Principal Investigator in this area:

The PI has wide experience of teaching and R & D in various areas of NLP mentioned in the proposal. The work on machine translation done by him used the Universal Networking Language (UNL) which was funded in 1996 by the United Nations. Many tools and resources for machine translation were developed under this project. The Hindi wordnet was developed under his guidance which is serving as the pioneer for various Indian language wordnets.

16. Total amount required: Rs. 100 Lakhs

Breakup of this amount institute-wise is given below:

No:	Institute	Amount in Lakhs of Rupees
1	IIT Bombay	15
2	Amrita University, Coimbatore	40
3	Tamil University, Tanjore	15
4	Dravidian University, Kuppam	15
5	Central University, Hyderabad	15
	Grand Total	100

The breakup of these amounts for each university is given in the following tables:

17. (a) Recurring budget (not more than 30%) of the proposal along with item-wise breakup (Manpower, Contingency, Consumable, Travel, Miscellaneous year wise breakup.

Total Recurring Budget for all FIVE Universities

No:	Task	Amount in Lakhs
1	Manpower	50.7
2	Honorarium	6.4
3	Contingency	12.00
4	Travel	10.70

Total Amount: Rs. 79.8 Lakhs

(The project is highly manpower intensive and hence recurring budget is more than non-recurring budget)

- (b) Detailed breakup of non-recurring items (with the equipment to be procured along with cost)

Total Non-Recurring Budget for all FIVE Universities

No	Task	Amount
1	Equipment	20.2

Total Amount: Rs. 20.2 Lakhs

Detailed breakup for Each University as follows:

Amrita University - Recurring Budget

Project Duration: 12 months

No:	Task	Module I: Amount in Lakh	Module II: Amount in Lakh	Total: Amount in Lakh
1	Staff Salaries	20	3.30	23.30
2	Contingency	3.00	1.00	4.00
3	Travel	2.20	0.50	2.70

Grand Total Rs. 30 Lakhs

Amrita University - Non Recurring Budget*Project Duration: 12 months*

No:	Equipment	Module I	Module II	Total
1	Desktops	8.00	0.80	8.80
2	Laptops	1.20	-	1.20

Grand Total Rs. 10 Lakhs

Net Total Amount for Amrita University: Rs. 40.00 Lakhs**IIT Bombay - Recurring Budget***Project Duration: 12 months*

No:	Task	Amount in Lakh
1	Staff Salaries	6.85
2	Contingency	2.00
3	Travel	2.00
4	Honorarium (coordination and project management)	1.6

Grand Total Rs. 12.45 Lakhs

IIT Bombay - Non Recurring Budget*Project Duration: 12 months*

No:	Equipment	Amount in Lakhs
1	Desktops	2.55

Grand Total Rs. 2.55 Lakhs

Net Total for IIT Bombay: Rs. 15 Lakhs

Tamil University- Recurring Budget

Project Duration: 12 months

No:	Task	Amount in Lakh
1	Staff Salaries	6.85
2	Contingency	2.00
3	Travel	2.00
4	Honorarium	1.6

Grand Total Rs. 12.45 Lakhs

Tamil University - Non Recurring Budget

Project Duration: 12 months

No:	Equipment	Amount in Lakhs
1	Desktops	2.55

Grand Total Rs. 2.55 Lakhs

Net Total for Tamil University, Tanjore: Rs. 15 Lakhs

Central University- Recurring Budget*Project Duration: 12 months*

No:	Task	Amount in Lakh
1	Staff Salaries	6.85
2	Contingency	2.00
3	Travel	2.00
4	Honorarium	1.6

Grand Total Rs. 12.45 Lakhs

Central University - Non Recurring Budget*Project Duration: 12 months*

No:	Equipment	Amount in Lakhs
1	Desktops	2.55

Grand Total Rs. 2.55 Lakhs

Net Total for Central University, Hyderabad: Rs. 15 Lakhs

Dravidian University- Recurring Budget

Project Duration: 12 months

No:	Task	Amount in Lakh
1	Staff Salaries	6.85
2	Contingency	2.00
3	Travel	2.00
4	Honorarium	1.6

Grand Total Rs. 12.45 Lakhs

Dravidian University - Non Recurring Budget

Project Duration: 12 months

No:	Equipment	Amount in Lakhs
1	Desktops	2.55

Grand Total Rs. 2.55 Lakhs

Net Total for Dravidian University, Kuppam: Rs. 15 Lakhs

18 Summary sheet (common for all the 3 schemes)

SUMMARY SHEET

1. Name of the Institution: IIT-Bombay
2. Title of the Project: Creation of Machine Translation tools and resources for English to Dravidian Languages
3. Name of the Department: Computer Science & Engineering
4. Cost of the Project : Rs 100 Lakhs
5. Amount released earlier if any: None
6. Utilization position in respect of grants released earlier (upto-2001) for various projects (Details to be given project-wise)

(i) Fully spent : Not Applicable

(ii) Unspent, proposal to utilize it: Not Applicable

7. Reasons for unspent balance: Not Applicable

8. Name of the Principal Investigator responsible for implementation of the Project.

Dr. Pushpak Bhattacharyya
Professor, Department of Computer Science and Engineering, Indian Institute of Technology,
Mumbai

(NAME & SIGNATURE OF THE PRINCIPAL INVESTIGATOR)

For Office Use Only

1. Name of the Area :
2. Recommendations:
 - a. Approved
 - b. Not approved
 - c. Deferred Amount (Rs. _____ in lakhs)
 - d. Transferred to area
3. Remarks, if any
(common for all the 3 schemes)

19. Certificate

1. Certified
 - (i) that the Principal Investigator is due to retirement during the currency of this project.
 - (ii) that the Principal Investigator is a regular employee of this Institution. However in case he/she proceeds sabbatical/resigns/proceeds VRS etc., the Institute will ensure to replace PI by a compatible academician to ensure that without any brake whatsoever, project will be completed within the stipulated period of 2 years.
 - (iii) that it will be ensured that the implementation will be carried out on mission mode with no time or cost over run and we are aware that ministry will neither provide any extension of time nor additional funding.
 - (iv) that no over heads will be charges by the Institution for this project and all facilitation including other essential/infrastructure support like air-conditioning etc. will be provided by the Institution.

Seal of the Institution

Signature of Head of the Institution

(Name)

Detail Project Report

PREAMBLE

Our main objective is to develop Machine Translation (MT) system and needed linguistic resources for English-Dravidian languages (Tamil, Malayalam, Telugu and Kannada), that would facilitate the creation of rich educational contents in Indian languages. Our research effort is to make all the tools and translation system to be based on Machine Learning methodologies so that computer graduates and other such non-linguists are able to immediately participate in the national mission on literacy by contributing additional tools for language translation.

The project will be implemented by a consortium of the following universities:

1. IIT, Bombay
2. Amrita University, Coimbatore
3. Tamil University, Tanjore
4. Dravidian University, Kuppam
5. Central University, Hyderabad

We have divided the project into 3 modules:

We have divided the project into 3 modules:

Module I aims at developing Machine translation tools and linguistic resources for English to Dravidian languages. We will also work in creating Machine Translation tools for Malayalam to Tamil and vice versa. This later can be extended to other language pairs.

In **Module II**, we aim at developing teaching material corresponding to the tools we are developing (most tools are based on machine learning) so that it can be delivered as part of undergraduate computer science and engineering curriculum on data mining/machine learning. This will ensure a critical amount of man power required for sustaining translation effort needed for national mission on education.

Module II also aims at training 500 faculties selected from across the country on machine translation methodologies using machine learning techniques.

Module III aims at developing a Dravidian Wordnet required for translation. This module requires collaboration between different universities. This will also link to the Hindi wordnet developed at IIT Bombay (www.cfilt.iitb.ac.in) which is being widely used for NLP involving Indian languages.

1. OBJECTIVE

Module I Objective

Machine Translation(MT) between natural languages is one of the biggest challenge in Natural Language Processing(NLP) or Computational Linguistics(CL) research. MT research started in early 1950s' and made lot of progress for English and European languages whereas MT research in India started much later and yet to produce tangible results. As a partner in the larger government initiative English to Indian Languages: Machine Translation system(EILMT), Amrita has been into MT research for past 3 years and has gained significant experience developing linguistic tools and annotated corpora for Tamil and Malayalam. Based on the experience we gained, we developed a working prototype MT system for English-Tamil and many linguistic tools and resources for Tamil.

The project aims at developing two sets of tools. One set is for assisting manual translation (Computer Assisted Translation tools or CAT tools) process and the other for semi-automated machine translation (MT tools).

- The aim of Computer Assisted Translation (CAT) tools creation is to develop range of tools that would assist human translators to translate the contents fast.
- The aim of the semi-automated machine translation is to develop Machine Translation (MT) system for English-Dravidian languages (Tamil, Malayalam, Telugu and Kannada), and between Tamil and Malayalam and linguistic resources that would facilitate the creation of rich educational contents in Indian languages. The tools can be easily modified for other Indian languages. All the tools are based on machine learning approaches. The linguistic (annotated) resources developed out of this project for machine learning will make it easy for computer graduates and non-linguists to directly enter and contribute by developing additional tools for the national mission on education and literacy without being bogged down by the technicalities of the languages.

1.1 Scope of work

Two types of tools are proposed in this project.

1. Machine Translation tools
2. Computer Assisted Translation tools

These two technologies are the consequence of different approaches. They do not produce the same results, and are used in distinct contexts. MT aims at assembling all the information necessary for translation in one program so that a text can be translated without human intervention. It exploits the computer's capacity to calculate in order to analyze the structure of a statement or sentence in the source language, break it down into easily translatable elements and then create a statement with the same structure in the target language. It uses huge plurilingual dictionaries, as well as corpora of texts that have already been translated. As mentioned, in the 1980s MT held great promises, but it has been steadily losing ground to computer-assisted translation because the latter responds more realistically to actual needs.

CAT uses a number of tools to help the translator work accurately and quickly, the most important of which are terminology databases and translation memories. In effect, the computer

offers a new way of approaching text processing of both the source and target text. Working with a digital document gives us non-sequential access to information so that we can use it according to our needs. It becomes easy to analyze the sentences of the source text, to verify the context in which a word or a text is used, or to create an inventory of terms, for example. Likewise, any part of the target text can be modified at any moment and parallel versions can be produced for comparison and evaluation. All these aspects have profound implications for translation, especially in terms of assessing the results, since the translator can work in a more relaxed way because of the greater freedom to make changes at any time while the work is in progress.

It is important to stress that automatic translation systems are not yet capable of producing an immediately useable text, as languages are highly dependant on context and on the different denotations and connotations of words and word combinations. It is not always possible to provide full context within the text itself, so that machine translation is limited to concrete situations and is considered to be primarily a means of saving time, rather than a replacement for human activity. It requires post-editing in order to yield a quality target text. To understand why accurate machine translation is difficult, we must understand the process of the cognition itself.

1.2 Cognitive Processes

To understand the essential principles underlying machine translation it is necessary to understand the functioning of the human brain. The first stage in human translation is complete comprehension of the source language text. This comprehension operates on several levels:

- Semantic level: understanding words out of context, as in a dictionary.
- Syntactic level: understanding words in a sentence.
- Pragmatic level: understanding words in situations and context.

Furthermore, there are at least five types of knowledge used in the translation process:

- Knowledge of the source language, which allows us to understand the original text.
- Knowledge of the target language, which makes it possible to produce a coherent text in that language.
- Knowledge of equivalents between the source and target languages.
- Knowledge of the subject field as well as general knowledge, both of which aid comprehension of the source language text.
- Knowledge of socio-cultural aspects, which is, of the customs and conventions of the source and target cultures.

Given the complexity of the phenomena that underlie the work of a human translator, it would be absurd to claim that a machine could produce a target text of the same quality as that of a human being. However, it is clear that even a human translator is seldom capable of producing a polished translation at first attempt. In reality the translation process comprises two stages: first, the production of a rough text or preliminary version in the target language, in which most of the translation problems are solved but which is far from being perfect; and second, the revision stage, varying from merely re-reading the text while making minor adjustments to the implementation of radical changes. It could therefore be said that MT aims at performing the first stage of this process in an automatic way, so that the human translator can then proceed directly to the second, carrying out the meticulous and demanding task of revision. The problem

is that the translator now faces a text that has not been translated by a human brain but by a machine, which changes the required approach because the errors are different. It becomes necessary to harmonize the machine version with human thought processes, judgments and experiences. Machine translation is thus both an aid and a trap for translators: an aid because it completes the first stage of translation; a trap because it is not always easy for the translator to keep the necessary critical distance from a text that, at least in a rudimentary way, is already translated, so that mistakes may go undetected. In no sense should a translation produced automatically be considered final, even if it appears on the surface to be coherent and correct.

Semi Automated Machine Translation

Machine Translation technology has been improved over time to produce good-quality results, but by no means should be seen as the panacea for communicating with other cultures. In effect, it can only achieve a good output when working with very simple texts or when controlled language is used. For example, Hutchins and Somers (1992) show the effectiveness of machine translation for weather forecasts. But to exploit the real power of machine translation, some editing is required before and after processing a text in a machine translation system as Somers (1997) and Allen (2003) suggest.

Machine translation of Scientific texts (The Indian Context).

In the present context of scientific text translation by Education ministry of India, where most of the text is being produced by Indian experts with knowledge of English, we have an advantage. We have control over the structure of the sentences. The authors may be asked to produce text with simple sentence structures and make the sentences as unambiguous as possible. This is advantageous in three respects.

- 1) Machine Translation to multiple languages becomes easy
- 2) Requirement for post editing the sentences is less
- 3) The text is more comprehensible.

Still we must expect sentence structures that are not easily translatable without ambiguity. So we must have a tool that classifies sentences into 'good and bad' with respect to translation accuracy. This tool must also be able to break complex sentences into simple sentences. We call these tools as preprocessing tools. These tools also must identify all the key terms, Phrases, collocations etc that need predefined translation.

The second phase is Machine translation process. This phase may use multiple translating engines and the output may be chosen according to some quality criteria.

The third phase is post editing. This is initially done by a linguist/language expert and then by the subject expert.

1.3 Computer Assisted Translation Tools (CAT tools)

Translation Memory Systems

The translation process may be sped up depending on the text type by using translation memories. Basically, the purpose of a translation memory is to allow translators to leverage previous translations, that is, to re-use text parts that have been already translated and as such can be used in the process of translating new documents. One must be aware of the pros and cons of using a translation memory prior to starting a translation, as they can also slow down the

translation process in non-repetitive texts such as novels, in which case creativity is more important than terminological consistency.

Translation Memories

Translation memories represent one of the most important applications of on-line bilingual texts, going back to the beginning of the 1980s with the pioneering TSS system of ALPS, later Alpnet. This was succeeded at the beginning of the 90s by programs such as Translator Manager, Translator's Workbench, Optimizer, Déjà Vu, Trados and Eurolang, among others. In its simplest form, a translation memory is a database in which translator stores translations for future re-use, either in the same text or other texts. Basically the program records bilingual pairs: a source-language segment (usually a sentence) combined with a target-language segment. If an identical or similar source-language segment comes up later, the translation memory program will find the previously-translated segment and automatically suggest it for the new translation. The translator is free to accept it without change, or edit it to fit the current context, or reject it altogether. Most programs find not only perfect matches but also partially-matching segments. This computer-assisted translation tool is most useful with texts possessing the following characteristics:

- **Terminological homogeneity:** The meaning of terms does not vary.
- **Phraseological homogeneity:** Ideas or actions are expressed or described with the same words
- **Short, simple sentences:** These increase the probability of repetition and reduce ambiguity.

A translation memory can be used in two ways:

1. In **interactive** mode: The text to be translated is on the computer screen and the translator selects the segments one by one to translate them. After each selection the program searches its memory for identical or similar segments and produces possible translations in a separate window. The translator accepts, modifies or rejects the suggestions.
2. In **automatic** mode: The program automatically processes the whole source-language text and inserts into the target-language text the translations it finds in the memory. This is a more useful mode if there is a lot of repetition because it avoids treating each segment in a separate operation.

A translation memory program is normally made up of the following elements:

- A translation editor, which protects the target text format.
- A text segment localizer.
- A terminological tool for dictionary management.
- An automatic system of analysis for new texts.
- A statistical tool that indicates the number of words translated and to be translated, the language, etc.

Thus translation memory programs are based on the accumulation and storing of knowledge that is recycled according to need, automating the use of terminology and access to dictionaries. When translation tasks are repeated, memories save the translator valuable time and even physical effort: for example, keyboard use can be reduced by as much as 70% with some texts. Memories also simplify project management and team translation by ensuring consistency. However, translation memories can only deal with a text simplistically in terms of linguistic segments; they cannot, unlike the human translator, have a vision of the text as a whole with regard to ideas and concepts or overall message.

A human translator may choose to rearrange or redistribute the information in the source text because the target language and culture demand a different content relationship to create coherence or facilitate comprehension. Another disadvantage of memories is that training time is essential for efficient use and even then it takes time to build up an extensive database i.e. they are not immediate time-savers straight out of the box. Finally, it should be stressed that translation memory programs are designed to increase the quality and efficiency of the translation process, particularly with regard to specialized texts with non-figurative language and fixed grammatical constructions, but they are not designed to replace the human translator.

Terminology Management Tools

Terminological consistency should be of utmost importance when translating in group and/or dealing with a big translation project. For this reason, it would be ideal if terminologists could analyze the source text and build a glossary before translating to prevent last-minute terminology changes once the translation process has already begun. In order to do so, terminologists use term-extraction and concordance tools such as WordSmith Tools.

Proofreading Tools

Once the translation and its editing are finished, everything should be proofread before the final product is delivered. Proofreading is the process which focuses on the correction of errors such as misspellings or typos and mistakes in grammar and punctuation. Microsoft Word comes with a built-in proofreading function called "track changes" which lets the user not only make changes to the document that will require a later acceptance to become permanent, but also to add comments or suggestions. Thus, when the translator gets back the revised version of his/her work, the translator can still consider the modifications made, as well as the comments and suggestions, to decide whether to accept or reject the changes, so the translation remains faithful as well as polished.

Databases for CAT: Mono and bilingual corpora

Two main applications of the corpora in translation

Here, two applications of specialized corpora are introduced to describe their role in producing a high-quality translation.

1. Translating Collocations

We can search for different collocations which are frequently encountered in the text. Consider the noun phrase "pre-emptive war". Bilingual dictionaries won't usually help us to translate such collocations.

2. Verifying or rejecting decision taken based on other tools

While traditional translation tools (such as dictionaries) suggest more than one equivalent and sometimes improper ones, corpora become an effective solution to these problems. When you are in doubt about which one to choose among the equivalents suggested by dictionary, corpora are great tools for verifying or rejecting the suggested translation(s).

Background and Technology Status

Translation system architecture

Machine Translation methodologies can be broadly classified into two paradigms: Rule based and Corpus based. Traditional models of rule based Machine Translation system include direct, syntax transfer and Interlingua. In direct translation, source language words are directly translated to target language. Syntax transfer approach involves deeper understanding of syntactic structures on both sides and creation of transfer rules which takes care the transformation of source text into target text. In inter-lingual approach, source language is mapped to an abstract intermediate unambiguous representation which will then be used to generate morphologically, syntactically and semantically well formed target sentence.

Our rule based architecture belongs to syntax transfer approach as shown in Figure 1. Not all the modules in the architecture are rule based, tools such as Part of Speech (POS) tagger and parser are statistical ones. Our approach effectively focuses on developing target resources rather than focusing on English. i.e. we will be using state of the art resources available for English in our system, this will reduce development time. On the syntactic parser side, we have two options. One is, parsing the input sentences using Tree Adjoining Grammar (TAG) parser and another is parsing the sentences using probabilistic Context Free Grammar (CFG) parser. For English-Indian language MT, parsers available in the public domain will be used. For Indian-Indian language MT, we will be developing parser and grammar resources on our own which are based on machine learning.

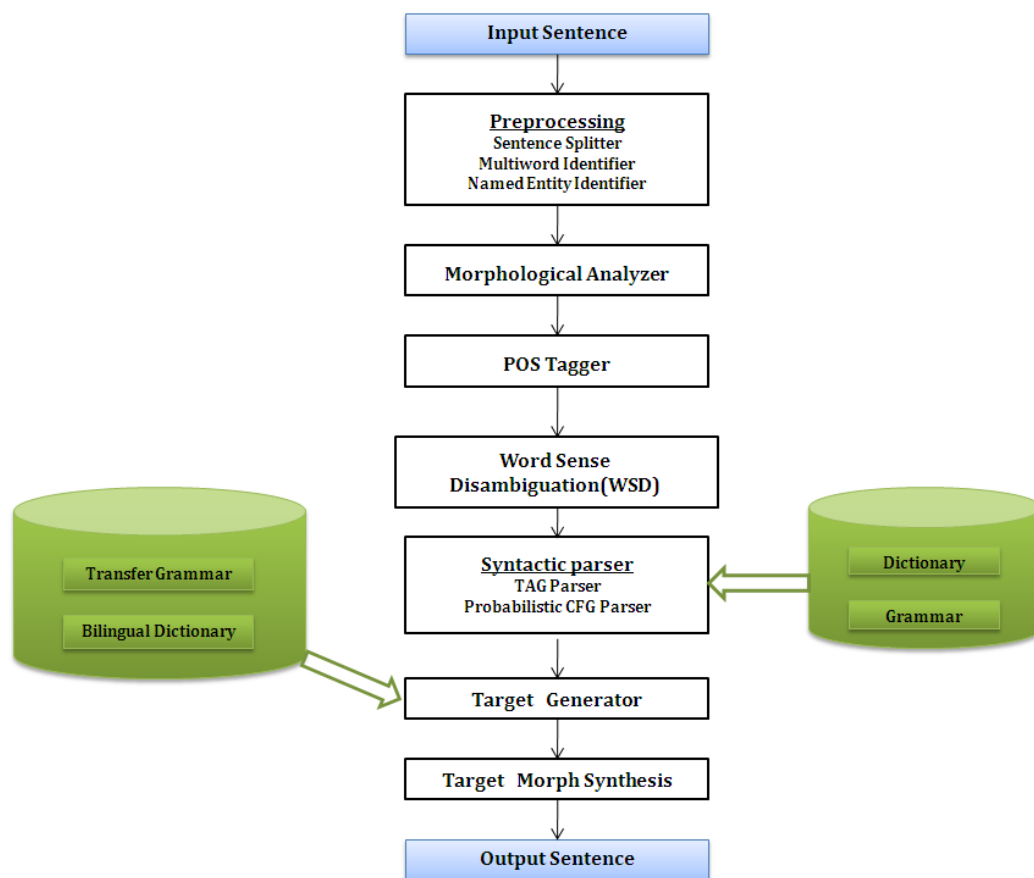


Figure 1: Architecture of the Rule based MT System

Comparison with existing MT systems in India

There are mainly three MT systems (TAG, Analgen and SMT) that are being developed by member institutions of the EILMT consortia. The TAG(CDAC-Pune) and Anlagen(IIT-Hyderabad) works adequately for English to Hindi translation. However the peculiar nature of Dravidian languages demands further fine tuning and research.

So we made an attempt to develop a new system. A prototype system is already developed which gives encouraging results. In parallel, we are also working on Machine learning based translation system, such as source tree to target string mapping and machine learning model that maps parse trees in the source language to parse trees in the target language (model is to be learned from parallel corpus).

We also propose to combine the merits of rule based and Machine learning based system.

Resources required for translation between Indian languages

1. Morph analyzer and tagger
2. Parser
3. Transliteration
4. Morph synthesizer
5. Wordnet

Module II: Objective

- 1) To develop a text book and teaching material (including video lectures, Lab assignments with linguistic data and resources) for '**Computational linguistics of Indian languages**' using machine learning. This is basically meant for undergraduate students in computer science and Engineering.

Text will be designed such that it covers the essentials of data mining/machine learning (Data Mining is currently taught as a compulsory subject for computer science in many universities) with applications in computational linguistics so that our country have enough manpower who can contribute towards developing more linguistic tools for translating educational documents from English to Indian Languages and thus strengthening the national mission on education.

- 2) Training of faculty in Engineering Colleges in '**Computational linguistics of Indian Languages**'.

Apart from basic data mining (machine learning) techniques the text book include the following

- 1) Transliteration from English to Indian Languages using data mining techniques
- 2) Morphological analysis and synthesis of Indian Languages using machine learning
- 3) Part of speech tagging of Indian Languages using Support Vector machines

- 4) Dependency parsing using support vector machines
- 5) Machine translation (rule based and Statistical) from English to Indian Languages and between Indian Languages
- 6) Text to speech synthesizer

Out of these six topics, machine learning tools for the first four topics achieves currently a minimum of 90% accuracy. Machine translation is yet to achieve good accuracy. However, for controlled text (text made with simple style and structures so that translation is easy) , fairly adequate accuracy can be achieved.

The course material developed will be taught to 500 faculties drawn from various engineering colleges in the country.

Module III: Objective

The main objective of DWN is to develop an extensive and high quality multilingual database with Wordnet for Dravidian languages in a cost-effective manner, and also to link it with the existing prominent Wordnet of languages like Hindi, English and other languages. The project will also develop a language independent set of semantic concepts linking the language networks together. The resources will be field tested for adequacy in an information retrieval application. The ultimate objectives are to move toward standardization of semantic classification of information for all Dravidian languages and to provide resources for development of applications, which can operate in a selected language or over a range of languages.

Major aims of the project are:

1. The major aim of the project is to prepare Wordnet for the following languages:
 - Tamil Wordnet
 - Malayalam Wordnet
 - Telugu Wordnet
 - Kannada Wordnet
 - Linking of Dravidian Wordnet with the Wordnet of Hindi and other languages in North India and also with English Wordnet
2. These individual Wordnet will be merged into Dravidian Wordnet and also linked with Wordnet of English and Hindi.
3. The notion of a synset and the main semantic relations will be taken over in Dravidian Wordnet. However, some specific changes will be made to the design of the database, which are mainly motivated by the following objectives:
 - (1) to create a multilingual database;
 - (2) to maintain language-specific relations in the Wordnet

(3) to achieve maximal compatibility across the different resources; to build the Wordnet relatively independently (re)-using existing resources

2. METHODOLOGY

Module I: Methodology

Machine learning based methodology will be used for creating tools like Transliteration, POS Tagger, Morphology Analyzer etc. 'Plagiarism detection Engine' developed by the coordinator for Ministry of Communication and Information Technology will be modified for making the tools like "Translation Memory System" and Terminology management system. Much of the time will be devoted to making linguistic resources and developing automated techniques for identifying complex sentences and finding a way of breaking into simpler sentences. Open-source parsers available like 'Stanford Parser' or 'LTAG Parser' developed will be tested for its adequacy for utilizing in Machine translation. Monolingual corpus will be collected from sources like Wikipedia in domains like Physics, Chemistry, and Mathematics etc for making terminology database.

Throughout, our research effort will be focused on developing machine learning resources and approaches for Machine Translation.

Module I: Tools developed so far

We have developed prototype versions of most of the tools mentioned above. These tools have to be scaled up for using in real life application systems. Screenshots of some of the tools are given below,

Application of MT for E-Learning content

The image displays two screenshots of educational software. The left screenshot shows a math problem in Tamil: "4 + 1 = 5" with a "சுவாஷ்" (Svash) smile icon. The right screenshot shows a "Reading" interface with student details, a score table, and a translation exercise.

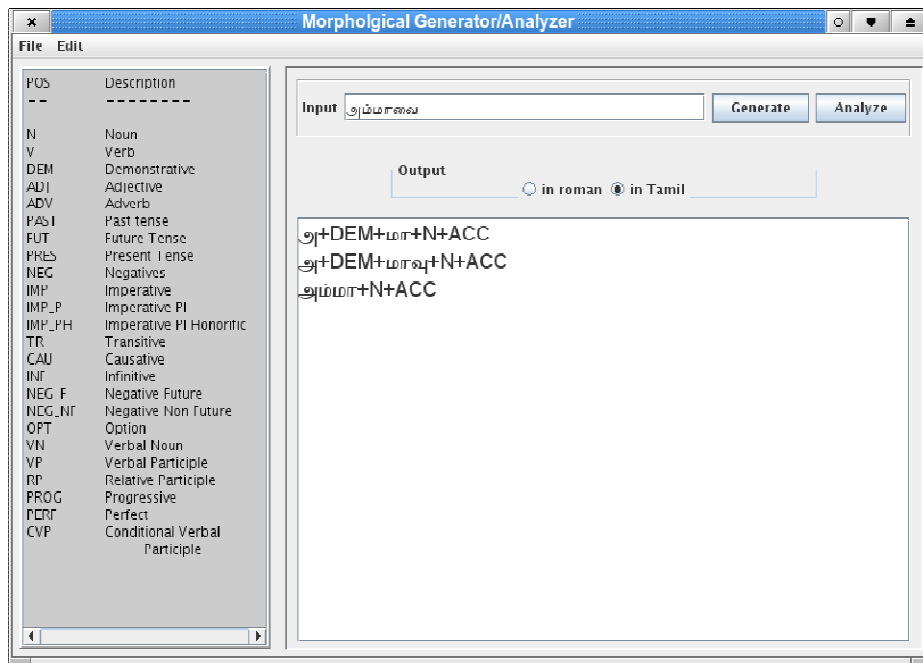
Left Screenshot:

- Navigation icons: HELP, AUDIO, TOOLS, SCORE, TUTOR.
- Text: 4 . ഒരണ്ണം കൂടി വാങ്ങി. ആകെ എത്ര ഉണ്ട്?
- Equation: $4 + 1 = 5$
- Smiley icon and text: സுவാഷ്
- Footer: Press F5 to play again Amrita Learning, 2006 - 2007

Right Screenshot:

- Student Detail: Student: Abhilash..., Level: III, Beginner, Lesson No: 3/35
- Score: 15 (marked with a red X), 15 (marked with a green checkmark)
- Reading section: "We live here." with a translation button: "நாங்கள் இங்கே வாழ்கிறோம்"
- Footer: Translate

Morphological Analyzer for Tamil



SVM Transliteration for Tamil

The screenshot shows a software window titled "SVM TRANSLITERATION". The window is divided into three main sections: "Input", "Conversion Process", and "Output".

Input Section: This section contains two radio buttons. The first is selected and labeled "Enter the lexicon name", with a text box containing "chandrababu". The second is labeled "Enter the filename" and is accompanied by a "Browse.." button. The word "OR" is centered between the two options.

Conversion Process Section: This section contains three buttons: "Convert to svm Format", "DO Transliteration...", and "Exit".

Output Section: This section is a text area displaying the transliterated output for the input "chandrababu". The output consists of five lines, each showing the input name followed by a different transliteration of the Tamil name "சந்திரபாபு":
chandrababu,சந்திரபாபு
chandrababu,சந்திரபாபு
chandrababu,சந்திரபாபு
chandrababu,சந்திரபாபு
chandrababu,சந்திரபாபு

Prototype English-Tamil MT System

English-Tamil Machine Translation(MT) System

Source Text

I was waiting at the restaurant to clean my car

Target Text

நான் என்னுடைய காலை சுத்தம் செய்ய உணவகத்தில் காத்துக்கொண்டிருந்தேன்
நான் என் காலை சுத்தம் செய்ய உணவகத்தில் காத்துக்கொண்டிருந்தேன்

Load MT System Translate Output in Regex format

General Syntax Tree Viewer

The image shows a software window titled "Syntax Tree Viewer" with a menu bar containing "File" and "Help". The window is divided into three main sections:

- Top Section:** A text area labeled "Tree String" containing the red text: `(S1 (S (NP (NNP India)) (VP (AUX is) (NP (DT a) (JJ beautiful) (NN country))))))`. To the right of this text area is a vertical scrollbar, a checkbox labeled "Penn String" which is checked, and a button labeled "View Tree".
- Middle Section:** A large white area displaying a syntax tree diagram. The root node is "S1", which branches into "S". "S" branches into "NP" and "VP". "NP" branches into "NNP", which leads to the word "India". "VP" branches into "AUX" (leading to "is") and "NP". This second "NP" branches into "DT" (leading to "a"), "JJ" (leading to "beautiful"), and "NN" (leading to "country").
- Bottom Section:** A text area labeled "Tree String" containing the green text: `(S1 (S (NP (NNP (India)) VP (AUX (is) NP (DT (a) JJ (beautiful) NN (country)))))))`. To the right of this text area is a vertical scrollbar, a button labeled "View String", and a button labeled "Clear".

LTAG Parser and TAG Tree Viewer

The screenshot displays the TAG Maestro v1.4 beta software interface. The main window is titled "TAG Maestro v1.4 beta" and contains two panels for viewing TAG trees, labeled "Tree102" and "Tree101".

Tree102: A syntactic tree for the sentence "I missed the flight". The root node is "S", which branches into "NP" (Pronoun) and "VP" (Verb and NP). The "NP" node contains "I". The "VP" node contains "missed" and another "NP" node. This second "NP" node branches into "Det" (the) and "Nominal" (flight).

Tree101: A syntactic tree for the sentence "you can book TWA flights". The root node is "S", which branches into "Aux" (can), "NP" (Pro), and "VP" (V, NP, NP). The "NP" node contains "you". The "VP" node contains "book", "PNoun" (TWA), and "Nominal" (Noun). The "Nominal" node contains "flights".

A central dialog box is overlaid on the trees, featuring the text:

ಅಥಿ
ಬ್ರಾಹ್ಮ
ಆ ಆ.

TAG Maestro v1.4 beta

Multilingual Editor for TAG Trees

© Amita Vishwa Vidyapeetham and CDAC Pune

At the bottom left, there is a "Transfer Grammar" field. On the right side, there is a control panel with various settings:

- Tree Name: Tree101
- Language: English
- Tree Type: Initial
- Category: CBrac, ConjAs
- Node Name: (empty)
- Constraints: NULL
- Properties: NULL
- Subscript: (empty)
- Person-Number-Gender: NULL
- Auxiliary Specifier: NULL
- Embedded Lexicon: (empty)
- Generate, Save buttons
- Source Tree: NULL
- Target Tree: NULL
- On Generation Mode checkbox
- Generate, Save buttons

Module II: Methodology

Our research has shown that a single data mining tool called “ Support vector machines” can be used for the first four tasks namely, transliteration , Morphological analysis, Part of speech tagging” and Dependency parsing . Highly accurate morphological analysis helps in machine translation between Indian languages. Statistical machine translation with huge parallel corpora will give ‘good enough’ translation of English text to Indian Languages. Rule based techniques if used for ‘controlled text’ give fairly good and grammatically right translation.

So the first part of the book will devote for explaining the basic theory behind the state of the art machine learning techniques. The second part will devote to computational linguistic problems and its modeling and solution using machine learning techniques. The text will be supplemented with

1. Video lectures for all the topics
2. lab assignments with real linguistic data sets

Module II: Tools and Expertise developed so far

Centre for Excellence in Computational Engineering and networking, Amrita Vishwa Vidyapeetham Coimbatore has developed several linguistic tools based on Machine learning methods. Some of tools available for free download are

- Morphological Analyzer/Generator for Tamil
- Tamil POS Tagger
- SVM based English to Tamil Transliteration Engine
- Rule based English to Tamil Transliteration Engine
- Linguistic Tree Viewer in Java
-

These tools are available for download from

<http://www.amrita.edu/cen/computational-linguistics.htm>.

Soon tools will be available for Hindi, Malayalam, Kannada and Telugu. The machine learning approach developed by the centre is so general that it is applicable for all Indian Languages.

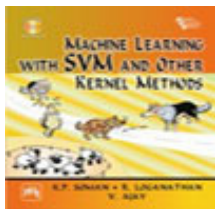
Other tools developed which are not yet released are

- 1) Parser
- 2) Excel based transliteration tool
- 3) Statistical Machine translation Engine.

Facilities Available

Amrita University is having e-learning studio equipped with video recording facilities and a multimedia group with expertise in developing 3D animation.

Relevant Books published by the Second Coordinator



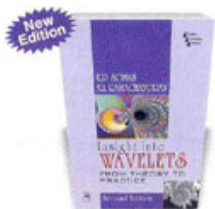
Machine Learning with SVM and Other Kernel Methods

Dr. Soman K.P., Loganathan .R, V. Ajay, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu. Published by: Prentice-Hall of India Pages : 496
ISBN : 978-81-203-3435-9 , Price : Rs. 425.00



Insight into Data Mining : From Theory and Practice

Dr. Soman K.P., Shyam Diwakar, V. Ajay, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu Published by: Prentice-Hall of India , Pages : 420 ,
ISBN : 81-203-2897-3
Price : Rs. 350.00



Insight into Wavelets: From Theory to Practice [Second Edition]

Dr. Soman K.P., Dr. Ramachandran K.I, Amrita Vishwa Vidyapeetham, Coimbatore, India. Published by: Prentice-Hall of India, Pages : 404 ,
ISBN : 81-203-2902-3
Price : Rs. 295.00

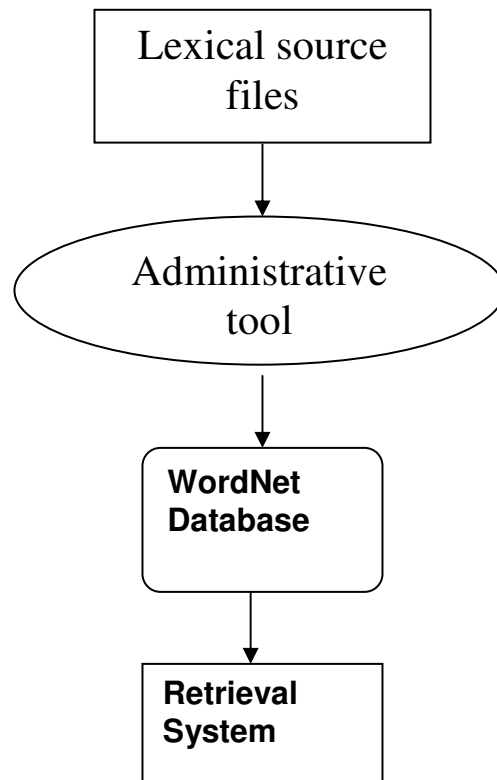
Module III: Methodology

Building Wordnet from scratches will take a lot of time and will consume a lot of money. In order to save time and money, it is proposed here to make use of the already available resources such as machine readable dictionaries, paper dictionaries, thesauri, etc. As some of the major tasks are shared between the consortium members, a lot of time and money will be saved. Top ontology and conceptual based thesaurus will be prepared by one of the members and will be shared by all in building the database. This makes it possible to build Wordnet for individual languages within a short stipulated period.

The word net system can be divided into four parts based on the specific tasks assigned to them:

- Lexical resource system
- Compiler system
- Storage system
- Retrieval system

But for our purpose we build lexical resource system, database and retrieval system only. We do away with the compiler system by building an administrative tool common for all the four language in which the lexical resource collocated by the lexicographer are entered.



The tasks will be distributed amongst various participating Institutes. The responsible institute for a given task will co-ordinate the development of various components and ensure delivery of the required component within the stipulated time frame.

Designing Data Base (T1)

Database is the core of WordNet. It will be created to pileup the information input into the Administrative tool so enable the retrieval system facilitate easy retrieval for the endusers.

Building Administrative tool (T2)

Administrative tool will be used to input the lexical information collected in the lexical files. The administrative tool facilitates inputting words and senses and linking them through lexical and semantic relations.

Building top ontology and concept based thesaurus (T3)

The ontology and concept based thesaurus helps in organizing the lexical information collected in lexical files to suit inputting in the administrative tool. The concept based thesaurus will help the individual language WordNet builders to build their lexical resource to suit WordNet.

Building Lexical Resource system (T4)

The lexical resource system consists of lexical files created by each language WordNet builders. The top ontology and concept based thesaurus helps the language groups to build their own lexical resource systems.

Entering the lexical information in the Administrative file (T5)

This is the major tasks to be performed by each language group. The lexical information collected using the concept based thesaurus will be input into the administrative tool. The lexical items are input with their senses along with sense relations.

Building Retrieval system (T6)

The retrieval system helps the end-users to retrieve the information they expect form the Wordnet. The lexical information stored in the database will be retrieved by user-interface management system.

Interlinking individual WordNets into Dravidian WordNet and also with WordNets of Hindi and English

This is a challenging task. The top ontology and the common concept based thesaurus on which the Wordnet are being built will make this possible.

The modular multilingual design of the DWN-database will have the following advantages:

- it will be possible to use the database for multilingual information retrieval, by expanding words in one language to related words in another language via the ILI;
- the different Wordnet can be compared and checked cross-linguistically which will make them more compatible;
- language-dependent differences can be maintained in the individual Wordnet;
- it will be possible to develop the Wordnet at different sites relatively independently;
- language-independent information such as the glosses, the domain-knowledge and the analytic Top Concepts can be stored only once and can be made available to all the language-specific modules via the inter-lingual relations;
- The database can be tailored to a user's needs by modifying the Top Concepts, the domain labels or instances, (e.g. by adding semantic features) without having to access the language-specific Wordnet.

The task of (semi) automatically linking the DWN with Wordnet of English, Hindi and other languages will be carried out with the long term expertise in this existing at IIT Bombay.

3. DELIVERABLES YEAR WISE AND ITS POSSIBLE CONTRIBUTION TO MAJOR OBJECTIVES OF MISSION.

Module I Phase 1 (12 months from the start), the following will be delivered

1. POS Taggers and tagged data for all Dravidian Languages
2. Morph Analyzer for all Dravidian Languages version 1
3. Transliteration Engine for all Dravidian Languages.
4. Translation Engine for English-Tamil – (Hybrid of Rule and Statistical)
5. Translation Memory system for manual translation
6. Parallel corpora (5000 sentence pairs) for all Dravidian languages

Module II: Phase 1 (12 months from the start), the following will be delivered

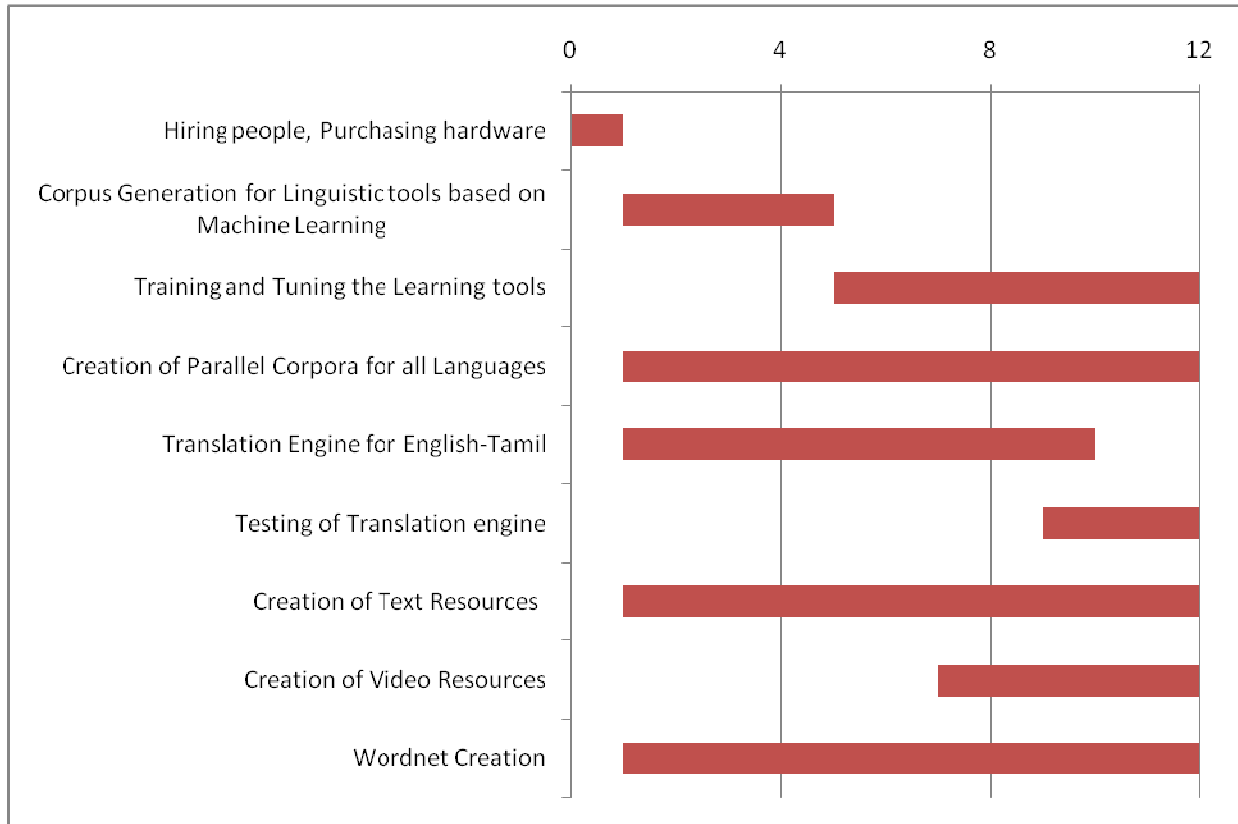
1. Text resources on Machine Learning based Computational Linguistics of Indian Languages.
2. Video Resources: Video Lectures of the key concepts in the textbook.
3. Lab Assignments with data sets.

Module III: Phase 1 (12 months from the start), the following will be delivered

1. First version of Wordnet for each language except Malayalam

4. TIME SCHEDULE (YEAR WISE)

Tasks	Start Month	Duration	End Month
Hiring people, Purchasing hardware	0	1	1
Corpus Generation for Linguistic tools based on Machine Learning	1	4	5
Training and Tuning the Learning tools	5	7	12
Creation of Parallel Corpora for all Languages	1	11	12
Translation Engine for English-Tamil	1	9	10
Testing of Translation engine	9	3	12
Creation of Text Resources	1	11	12
Creation of Video Resources	7	5	12
Wordnet Creation	1	11	12



5. DETAILS OF PERMANENT ASSETS TO BE PROCURED FROM THE PROJECT WITH ESTIMATED COST.

The list of equipments supplied in the next section 6 will constitute the permanent assets to be procured under this project.

6. DETAILS OF FINANCIAL OUTLAY IN YEAR WISE FOR RECURRING AND NON-RECURRING FUNDS.

Such a large-scale project needs inputs for the initial infrastructure and for building and maintaining a sizable team of technical and linguistic experts.

The total funding support being requested from the ministry for a one year period is Rs. 100 lakhs.

The major components for which the funding is required are outlined below:

6.1 The recurring outlay for the project is required for the following items

6.1.1 Manpower

The project is highly manpower intensive, requiring many research associates, linguists, programmers and data entry operators working together.

The total outlay for Manpower for all 5 universities amounts to Rs. 50.7 Lakhs for one year.

6.1.2 Travel

This is a project involving many languages and multiple universities are participating in this mega project. Frequent travel between universities is required to coordinate the work. The total amount for all 5 universities amounts to: Rs. 10.70 Lakhs for one year.

6.1.3 Contingency

The contingency for all the 5 universities is kept as Rs.12 Lakhs for one year.

6.2 The non recurring outlay for the project is required for the following items

6.2.1 Equipment

Since the project involves lot of manpower for programming, data entry and data storage, the project needs about 40 computers and a high capacity server. The total amount for equipment is kept at Rs 20.2 lakhs for one year.

7. MANAGEMENT OF DELIVERABLES & IPR ETC.

All the deliverables under this project will be released in the open source domain.

8. JUSTIFICATION OF THE PROJECTION WITH CLEAR CUT STATEMENT ABOUT OUTCOMES OF THE PROJECT CONTRIBUTING TO MISSION OBJECTIVE

One of the main objectives of the mission is to make the educational contents available in Indian Languages. English being Lingua-franca of science and technology, most of educational materials are being created in world over is in English. The aim of this project is to develop tools and resources that assist in translating such materials.

Translation on its own is a nascent industry in India. Tools and resources developed in this project will give a boost to this industry giving jobs to thousands in the country.

Curriculum Vitae of Principal Investigator

1. Dr Pushpak Bhattacharyya

- a. Education:
 - i. PhD, CSE, IIT Bombay, 1993
 - ii. M.Tech, CSE, IIT Kanpur, 1986
 - iii. B.Tech, EE, IIT Kharagpur, 1984
- b. Experience:
 - i. Professor, CSE Dept, IIT Bombay, 2002-present
 - ii. Visiting Professor, CSE Dept, Stanford University, 2004
 - iii. Visiting Professor, University Joseph Fouriere, Grenoble, France, 2005 and 2007
 - iv. Visiting Research Fellow, EECS, MIT, Cambridge, 1990
 - v. Research and Development Engineer, Wipro InfoTech, 1986-88
- c. Areas of Interest:
 - i. Natural Language Processing
 - ii. Machine Learning
 - iii. Information Extraction
 - iv. Machine Translation
- d. Projects:
 - i. Universal Networking Language, United Nations funded, 1996-99
 - ii. Laboratory for Intelligent Internets, TCS funded, 2001-04
 - iii. Technology Development in Indian Languages, DIT funded, 2000-03
 - iv. Meaning Based, Multilingual Search, under DIT funded MLAsia project, 2002-05
 - v. Tools and Resources for Indian Languages, under WB funded DGF project, 2003-2006
 - vi. Cross Lingual IR (DIT funded consortium project; IITB consortium leader): 2006 till date
 - vii. English to Indian Language Machine Translation (DIT funded consortium project; leader CDAC Pune): 2006-till date
 - viii. Indian to Indian Language Machine Translation (DIT funded consortium project; leader IIIT Hyderabad): 2006-till date
 - ix. Laboratory for Intelligent Systems, TCS funded: 2005 till date
- e. Select publications:
 - i. Rajat Mohanty and Pushpak Bhattacharyya, *Lexical Resources for Semantic Extraction*, Lexical Resources Engineering Conference (LREC08), Marrakech, Morocco, May 26-June 1, 2008.
 - ii. Rajat Mohanty, Pushpak Bhattacharyya, Prabhakar Pande, Shraddha Kalele, Mitesh Khapra and Aditya Sharma, *Synset Based Multilingual Dictionary: Insights, Applications and Challenges*, Global Wordnet Conference (GWC08), Szeged, Hungary, January 22-25, 2008.

- iii. J.Ramanand, Akshay Ukey, Brahm Kiran Singh and Pushpak Bhattacharyya, *Mapping and Structural Analysis of Multilingual Wordnets*, IEEE Data Engineering Bulletin, 30(1), March 2007
 - iv. Manish Sinha, Mahesh Reddy and Pushpak Bhattacharyya, *An Approach towards Construction and Application of Multilingual Indo-Wordnet*, 3rd Global Wordnet Conference (GWC 06), Jeju Island, Korea, January, 2006.
 - v. Rajat Mohanty, Anupama Dutta and Pushpak Bhattacharyya, *Semantically Relatable Sets: Building Blocks for Representing Semantics*, 10th Machine Translation Summit (MT Summit 05), Phuket, September, 2005.
 - vi. G. Ramakrishnan, Krishna Prasad Chitrapura, Raghu Krishnapuram and Pushpak Bhattacharyya, *A Model for Handling Approximate, Noisy or Incomplete Labeling in Text Classification* , International Conference on Machine Learning (ICML 05), Bonn, August, 2005.
- f. Awards:
- i. Microsoft Distinguished Researcher in a focused area (NLP): 2007
 - ii. IBM Innovation Award for Faculty (UIMA platform for Indian Languages): 2007
 - iii. Research Grant Award from United Nations: 1996
- g. Honorary service:
- i. Area Chair: ACL 2007
 - ii. Program Committee member: COLING 2008; Global wordnet conferences India, Hungary, Czechoslovakia; LREC conference, 2008

Curriculum Vitae of Second Coordinator

Name : **Dr. Soman K. P.**
Designation : Professor & Head, CEN,
 Centre for Excellence in Computational Engg. and Networking,
 Amrita Vishwa Vidyapeetham.

Degrees conferred

Degree	Institution Conferring the Degree	Field(s)	Year
B.Sc. (Engg)	Regional Engineering College, Calicut	Electrical Engineering	1984
P.M. Diploma	Indian Statistical Institute, Calcutta.	Quality Control and Operations Research	1986
M.Tech.		Reliability Engineering	
Ph.D.	Indian Institute of Technology, Kharagpur.	Systems Engineering	1989
	Indian Institute of Technology, Kharagpur.		1994

Research Work / Training Experience :

Institution	Nature of Work Done	Duration
Indian Institute of Technology, Kharagpur	Estimation of System Reliability	4 years (1990 – 94)
Keltron Counters, Trivandrum.	Quality Control	3 months (1995)
GKW, Howrah.	Quality Control	6 months (1996)

Current fields of research : Interior Point Methods for Optimisation Problems, High Performance Computing, Machine learning using Support Vector Machines, Signal Processing, Wavelets and Fractals.

Sponsored Research

- “Analysis of PIV and PLIF Images for Fluid Flow Characterization and Visualization” for Indian Space Research Organization (ISRO), Government of India
- “Study for Development of Methodology Detection Tools for Digital Contents Plagiarism and other Piracy Software.” for the DIT, Ministry of Communications and Information Technology, Government of India.
- “Under Water Sonic Data Classification using Massively Parallel Support Vector Machines”, for National Physical Oceanographic Laboratory, Government of India.
- “Study of Image fusion techniques and Quality Metrics ” for Defense Research Development Organization, Government of India.
- Machine Translation System – From English to Tamil for the DIT, Ministry of Communications and Information Technology, Government of India.
- Video Summarization for content based Video Retrieval, Indian Space Research Organization (ISRO), Government of India

List of publications : (partial list of publications)

Books

- “**Insight into Wavelets**”, Soman K.P., K.I. Ramahandran, Prentice Hall India, 2004.
- “**Insight into Data Mining**”, Soman K.P., Shyam Diwakar, Ajay V, Prentice Hall India, 2006.
- “**Support Vector machines and Kernel Methods**”, Soman K.P., Ajay V., Loganathan R., Prentice Hall India, 2009.

Partial List of Recent Papers:

1. Sindhiya Binulal Golla, Anand P, Soman KP, “SVM based approach to Telugu Parts Of Speech Tagging using SVMTool”, *International Journal of Recent trends in Engineering, Academy Publishers, Finland, 2009*
2. Sumaja S, Loganathan R, Soman KP, “English to Malayalam Transliteration Using Sequence Labeling Approach”, *International Journal of Recent trends in Engineering, Academy Publishers, Finland, 2009*
3. Dhanlakshmi V, Anand Kumar M Shivapratap G, Soman KP, Rajendran S, “Tamil POS Tagging using Linear Programming”, *International Journal of Recent trends in Engineering, Academy Publishers, Finland, 2009*
4. Vijaya MS, Shivapratap G, Ajith VP, Soman KP, “ English to Tamil Transliteration Using WEKA A Data Mining Tool”, *International Journal of Recent trends in Engineering, Academy Publishers, Finland, 2009*

5. Dhanalakshmi V, Anandkumar M, Vijaya MS, Loganathan R, Soman KP, Rajendran S, “Tamil Part of Speech Tagger based on SVMTool”, *Proceedings of The International Conference on Asian Language Processing*, Chiang Mai, Thailand, 2008
6. Vijaya MS, Loganathan R, Shivapratap G, Ajith VP, Soman KP, “English to Tamil Transliteration using Sequence Labeling Approach”, *Proceedings of The International Conference on Asian Language Processing*, Chiang Mai, Thailand, 2008
7. Dhanalakshmi V, Anandkumar M, Vijaya MS, Soman KP, “Comparison of SVMTool based Tagger Generator with Other Methods”, *Proceedings of 6th International Conference On Natural Language Processing (ICON)*, CDAC, Pune, 2008
8. Vijaya MS, Dhanalakshmi V, Shivapratap G, Ajith VP, Soman KP, “Sequence Labelling Approach for English to Tamil Translation using Memory based Learning”, *Proceedings of 6th International Conference On Natural Language Processing (ICON)*, CDAC, Pune, 2008

BRIEF RESUME OF CURRICULUM VITAE OF S.RAJENDRAN

Name: S.Rajendran
Age and date of birth: 58 years; 3-12-1950
Place of birth: Nagercoil, Tamilnadu, India
Nationality: Indian
Marital Status: Married
Country, State and District: India, Tamilnadu, Kanyakumari District
Mother tongue: Tamil
Residential Address: 55/122, Parisutham Nagar, Thanjavur 613007, Tamilnadu, India.
Office Address: Professor, Department of Linguistics, Tamil University, Thanjavur 613010, Tamilnadu, India.

Educational Qualification

Sr. no.	Name of the degree	Name of the University	Subjects	Year of Pass	Class
1	BSc	Madurai University, Madurai, India	Chemistry, Physics & Zoology	1971	Second
2	MA	University of Kerala, Thiruvananthapuram, India	Linguistics	1973	First
3	PhD	University of Poona, Pune, India	Linguistics	1978	

Professional Experience

Sr.No.	Name of the Post held	Name of the Institute where employed	Period of service
1	Project Assistant	Central Institute of Indian Languages, Mysore, India	30-1-1979 to 6-12-1980
2	Research Associate (Sponsored by University Grants Commission)	Deccan College Post-graduate and Research Institute, Pune, India	12-12-1980 to 9-9-1983
3	Research Officer	Language Division, Ministry of Home Affairs, Calcutta, India	12-9-1983 to 19-1-1989
4	Lecturer	Department of Linguistics, Tamil University, Thanjavur, India	23-1-1989 to 22-1-1997
5	Associate Professor	Department of	23-1-1997 till date

		Linguistics, Tamil University, Thanjavur, India	
--	--	---	--

Teaching Experience: 25 years

Research Experience: 35 years

Area of Specialization: Syntax, Semantics, Morphology and Phonology in the light of Natural Language Processing, Machine Translation

Number of Training Programmes attended: 4

Proficiency in Languages: Can speak, read and write: Tamil, Malayalam, English, Hindi

Number of Research projects at hand: 2

Number of Research Projects completed: 15

Details of the Workshops, Seminar, Training Courses, etc. conducted:

Number of Workshops Conducted = 5

Number or Refresher Course Conducted = 5

Number of Papers Published: Total 55, International 13, National 42

Number of books published = 4

Research Guidance:

Number of M.Phil Students guided = 25

Number of M.Phil Students under supervision = 5

Number of Ph.D. Students guided = 6

Number of Ph.D. Students under supervision = 10

S.ARULMOZI

Assistant Professor
Dept. of Dravidian & Computational Linguistics
Dravidian University, Kuppam 517425, India

Mobile: +91-9441330510
Telefax: +91-8570-278230
Email: arulmozi@gmail.com

EDUCATION

- Ph.D., Applied Linguistics, University of Hyderabad, 1999.
- PGDTS, Translation Studies, University of Hyderabad, 1995 (66%)
- M.Phil., Applied Linguistics, University of Hyderabad, 1992 (70.8%)
- M.A., Linguistics, Bharathiar University, 1990 (64.4%)
- B.Sc., Chemistry, Bharathiar University, 1988. (54.6%)
- H.Sc., Board of Higher Secondary Education, Tamil Nadu, 1985 (61.75%)
- S.S.L.C., Board of Secondary Education, Tamil Nadu, 1983 (64.5%)

EMPLOYMENT

- *Assistant Professor* (29 December 2005 onwards)
Dept. of Dravidian & Computational Linguistics
Dravidian University, Kuppam
- *Guest Faculty* (19 January 2005 - 28 December 2005)
Centre for Applied Linguistics & Translation Studies
University of Hyderabad, Hyderabad
- *Member Research Staff* (29 December 2000 - 15 January 2005)
AU-KBC Research Centre
MIT Campus of Anna University, Chennai
- *Project Fellow* (21 January 1999 to 2 July 2000)
UGC-Major Research Project, Department of Linguistics
Tamil University, Thanjavur
- *Language Assistant-Tamil* (8 June 1998 to 31 October 1998)
DoE Project, Central Institute of Indian Languages, Mysore.

RESEARCH EXPERIENCE

- Ph.D Thesis (Research Guide: Prof. Probal Dasgupta & Dr. N. Krupanandam)
Title: *Aspects of Inflectional Morphophonology: A Computational Approach*
- M.Phil. Dissertation (Research Guide: Prof. Probal Dasgupta)
Title: *Dynamics of Translation in Reconstructing Sci-Tech Terminologies*
- M.A. Project (Research Guide: Dr. C. Shanmugom)
Title: *Advertisement Headlines*

PROFESSIONAL EXPERIENCE

1. Invited Member, *Project Review and Steering Group (PRSG) Meeting of Indian Language-Indian*

- Language MT Consortium, IIIT-Hyderabad, 23-25 July 2008.*
2. Participant, *Workshop on Ontology, NLP, IR and IE* organized by IIT-Bombay and HP Labs, 16-18 July 2008.
 3. Resource Person, *Advanced Course in Computational Linguistics* organized by the Dravidian University, 16-25 March 2008.
 4. Resource Person, Summer Institute, *New Linguistic Survey of India* organized by CIIL, Mysore 24-25 June 2007.
 5. Participant, *Summer Institute, New Linguistic Survey of India* organized by CIIL, Mysore from 1-10 June 2007.
 6. Participant, *Summer Institute in Natural Language Processing* organized by Microsoft Research Lab India and IISc, Bangalore from 3-16 May 2007.
 7. Participant, Workshop on *Linguistic Annotation of Speech Corpora for Text to Speech Systems* organized by IISc, Bangalore and HP Labs, Bangalore from 25th August to 5th September, 2003.
 8. Resource Person, National Seminar on *Cross Language Communication Tools (Machine Translation) for Indian Languages* organized by University of Madras and CIIL, Mysore from 25-27 February 2004.
 9. Resource Person, All India Symposium on *Computational Linguistics with Special Reference to Tamil and other Indian Languages* organized by University of Madras and CIIL, Mysore from 1-3 November 2000.
 10. Participant: NLP Advanced course in *Computational Linguistics-Teacher's Training Programme* organized by Centre for ALTS, University of Hyderabad from 1 February to 5 March 1994.

REFRESHER/ORIENTATION COURSES

1. Participant, *Refresher Course in Language Studies* organized by the UGC-Academic Staff College, Bharathiar University from 30 October 2008 to 19 November 2008.

WORKSHOPS/SEMINARS COORDINATED

1. *National Seminar on Classicalness of Languages: The Dravidian Context, 25-26 January 2008*, organized by the Department of Dravidian and Computational Linguistics, Dravidian University – As Coordinator.
2. *National Workshop on Introduction to Computational Linguistics, 8-16 December 2006*, organized by the Department of Dravidian and Computational Linguistics, Dravidian University – As Coordinator.
3. *National Workshop on WordNet for Dravidian Languages, 2-3 June 2003*, Organized by AU-KBC Research Centre, Chennai, Central Institute of Indian Languages, Mysore and Tamil University – As Coordinator.
4. *International Workshop Corpus-based Natural Language Processing 17-31 December 2001*, organized by AU-KBC Research Centre, Chennai, RCILTS-Tamil, Anna University, National Centre for Software Technology, Mumbai, Language Technology Resource Centre, IIIT, Hyderabad and Tamil University, Thanjavur – As Coordinator.

PROJECTS

- Title: *Tamil WordNet to Enhance Machine Translation* (funded by Tamil Nadu Software Development Fund of Tamil Virtual University)
Duration: 18 months (April 2002 – October 2003)
Project executed at: AU-KBC Research Centre, MIT Campus of Anna University
Role: Coordinator
- Title: *Development of Indian Language to Indian Language Machine Translation System*

[*Telugu-Tamil*] (sponsored by the Department of Information Technology, Ministry of Communications and Information Technology, GoI).
Duration: 2 years (September 2007 – August 2009)
Role: Coordinator

- Title: *Development of VerbNet for Tamil*, Individual Project carried out at Dravidian University (completed in May 2007).
- Title: *Indian Language Corpora Initiative* [Telugu] (sponsored by the Department of Information Technology, Ministry of Communications and Information Technology, GoI.)
Duration: 2 years (February 2009-January 2009)
Role: Chief Investigator

RESEARCH GUIDANCE

No. of Ph.D Students under supervision: 2

RESEARCH PUBLICATIONS BOOKS

- '*Papers on Language Technology*', Dravidian University Publication, May 2008, (*Co-Edited with Panchanan Mohanty*).

DICTIONARIES

- '*Common Man's Multilingual Dictionary (6 volumes)*', Dravidian University Publication, May 2008, (*Technical Coordinator*).

IN JOURNALS/PROCEEDINGS

1. '*Towards Building a Computational Model for Tamil Verbal Compounds*', South Asian Language Review, Vol.17:1, Jan 2007, 28-68.
2. '*tamizh coRpinnal: oor aRimukam (in Tamil)*', Dravida Malar, Nov-06-Jan-07, 32-35.
3. '*Verb Sense Disambiguation of Simple Sentences in Telugu-Tamil MT*' (with Dr.M.C.Kesava Murty), Proceedings of the Symposium on Modeling and Shallow Parsing in Indian Languages, IIT-Bombay, 2-4 April 2006.
4. '*A Tamil Morphological Analyser*', Recent Advances in National Language Processing, 2003, 31-39, Hyderabad.
5. '*Computational Morphology of Verbal Complex*', Word Structure in Dravidian, Kuppam: Dravidian University Publication, 2003, 376-396.Mysore.
6. '*Tamil WordNet*' (co-authored with Dr. S.Rajendran, et.al), Proceedings of the First Global WordNet Conference, Mysore: Central Institute of Indian Languages, 21-25 January 2002.
7. '*urpaniyalum te.po.mii.yum*' (coauthored with Dr. S. Rajendran), tamiz pozil, tuNar 74, malar 3, 2000, 89-97.
8. '*Diglossia and Tamil: Some Sociolinguistic Observations*' (co-authored with Prof. Probal Dasgupta), Indian Journal of Linguistics, Vol-19, Jan-Dec 2000.

PAPERS PRESENTED

1. '*Enriching Verbs in Tamil WordNet*', National Seminar on Applied Linguistics, Department of

- Linguistics, Bharathiar University, Coimbatore, 22-24 October 2006.
2. '*WordNet: State of the Art*', Talk and Demonstration in VIBUDHA, Dravidian University, 26 April 2006
 3. '*Verb Sense Disambiguation of Simple Sentences in Telugu-Tamil MT* (with Dr. M.C.Kesava Murty), Proceedings of the Symposium on Modeling and Shallow Parsing in Indian Languages, IIT-Bombay, 2-4 April 2006.
 4. '*Machine Translation: An Introduction*', National Workshop on Kannada-Telugu Translation, Dravidian University, 22-26 March 2006.
 5. '*Tamil WordNet*', Workshop on WordNet: Current Perspectives and Future Prospects, Centre for ALTS, University of Hyderabad, Hyderabad, 24-25 February 2006.
 6. '*A Tamil Morphological Analyser*' (co-authored with S.Viswanathan, et.al), International Conference on Natural Language Processing, IIIT, Hyderabad, 18-20 December 2003.
 7. '*Building Telugu WordNet as a Component of Dravidian WordNet*' (co-authored with G. Uma Maheshwar Rao & S. Rajendran), National Seminar on Language Technology Tools: Implementation of Telugu, University of Hyderabad, Hyderabad, 8-10th October 2003.
 8. '*Tamil WordNet*' (co-authored with S.Rajendran, et.al), First Global WordNet Conference, Central Institute of Indian Languages, Mysore, 21-25 January 2002.
 9. '*Computational Morphology of Verbal Complex*' (co-authored with S. Rajendran, et.al), National Seminar on Word Structure of Dravidian Languages, Dravidian University, Kuppam, 26-28 November 2001.
 10. '*urpaniyalum te.po.mii.yum*' (co-authored with S. Rajendran), T.P.M. Centenary Celebration Conference, Department of Linguistics, Tamil University, Thanjavur, 23-24 March 2000.
 11. '*Spell and Grammar Checker for Tamil*' (co-authored with S. Rajendran and S. Boopathy), 27th All India Conference of Dravidian Linguists', Kerala University, Thiruvananthapuram, 17-19 June 1999.
 12. '*Computational Model for Tamil Verbal Compounds*', 26th All India Conference of Dravidian Linguists, PS Telugu University, Hyderabad, 18-20 June 1998.

MEMBERSHIP

- Life Member, International Journal of Dravidian Linguistics.
- Member, Project Advisory Committee, Project on Corpus Development of Ancient Tamil Works, Scheme for Classical Tamil.
- Board of Studies Member, Department of Linguistics, Bharathiar University.

AWARDS/SCHOLARSHIP

- Award for the best research paper presented at the 26th All India Conference of Dravidian Linguists held at the PS Telugu University, Hyderabad during 18-20 June 1998.
- Merit Scholarship for Ph.D. research work from the Central Institute of Indian Languages, Mysore from September 1995 to September 1997.
- Award for the President Scout in the year 1984.

AREAS OF INTEREST

- Computational Linguistics, Corpus Studies, WordNet, Machine Translation

CURRICULUM VITAE

Kavi Narayana Murthy

Professor, Department of Computer and Information Sciences
University of Hyderabad

Education:

Ph.D. (Computer Science) - University of Hyderabad - 1996

Area of specialization: Natural Language Processing

Thesis Title: Universal Clause Structure Grammar

Master of Technology (MTech) (Artificial Intelligence and Robotics)

University of Hyderabad - 1989 - First Class with Distinction - First Rank

Bachelor of Engineering (BE) (Mechanical Engineering)

Bangalore University - 1983 - First Class with Distinction

Professional Experience: Total: 23 Years

PG Teaching Experience: Total: 18 Years

Areas of Research Interest:

- Natural Language Engineering: - Syntactic Parsing, Word Sense Disambiguation, Text Categorization
- Speech technologies: - Automatic Speech Recognition, Speaker Recognition

Publications:

Books: 3, Journal Articles: 9, Conf. Proceedings: 44, Papers Presented in Conf.:12, Tech Reports: 12,

Relevant Books:

1. K. Narayana Murthy, "Natural Language Processing - An Information Access Perspective", Sarada Ranganathan Endowment in Library Science, Bangalore, 2006
2. K. Narayana Murthy, B. B. Chaudhuri (Eds), "Language Engineering Conference", Proceedings of the International Conference on Language Engineering LEC-2002, 13-15 December 2002, Hyderabad, India, IEEE Computer Society Press

Recent Journal Publications:

1. Kavi Narayana Murthy, "Technology for Telugu" - Bhasha, Vol 1, No 3, Feb 2008, pp 70-95
2. G Bharadwaja Kumar, Kavi Narayana Murthy, B B Chaudhuri, "Statistical Analysis of

Telugu Text Corpora" - IJDL, Vol 36, No 2, June 2007, pp - 71-99

3. Surabhi Sreekanth, Kavi Narayana Murthy, "Text-Dependent Speaker Recognition System for Telugu", Osmania Papers in Linguistics, Vol. 31, 2005, pp 84-99

4. Kavi Narayana Murthy, G Bharadwaja Kumar, "Language Identification from Small Text Samples", Journal of Quantitative Linguistics, Vol 13, No. 1, 2006 pp. 57-80

5. Atul Negi, Kavi Narayana Murthy, Chakravarthy Bhagvati, "Foundational Issues of Document Engineering in Indian Scripts and a Case Study in Telugu", Vivek, Vol 16, No. 2, 2006, pp 2-7

Research Guidance:

1. Ph.D: Awarded: 2 - On-Going: 4
2. M.Tech: 38 (Awarded) - 7 (On-Going):
3. MCA: 12

Theses Review:

Reviewed several PhD and MS theses from reputed institutions from India and abroad.

Foreign Collaboration:

1. Member, Indo-French Research Network in Computational Linguistics - Indian coordinator for 2001-2004
2. Member, Indo-German Team on Language Technologies on Cross-Lingual Information Access - V-MIA (Voice Based Multi-lingual Information Access)

Visiting Professor/Scientist:

1. Center for Advanced Study in Linguistics, Osmania University 1st December 2005 to 30th December 2005
2. AU-KBC Research Centre, Chennai 10th January 2005 to 30th November 2005
3. Documentation Research and Training Centre (DRTC), Indian Statistical Institute (ISI), Bangalore 12th January 2004 to 23rd January 2004
4. Center for Advanced Study in Linguistics, Annamalai University 22nd February 2003 to 7th March 2003
5. Also visited INRIA and University of Paris (France), Addis Ababa University (Ethiopia) and delivered lectures

Other Honours and Distinctions:

1. Delivered Sarada Ranganathan Endowment Lectures 2004 on Natural Language Processing and Information Retrieval 9-11 August 2004
2. Achievement Award - In Recognition of Research to the Field and in appreciation of valuable services to the IICAI 2005 conference (Second Indian International Conference on Artificial Intelligence, 20 - 22 December 2005, Pune, India)

Funded Projects and Consultancy:

1. Development of a Robust Document Analysis and Recognition System for Printed Indian Scripts, DIT, Government of India, Co-Investigator

2. Language Engineering Research - UGC (under University with Potential for Excellence scheme) - April 2002 to March 2007 - Coordinator
3. Resource Centre for Indian Language Technology Solutions (Telugu) - Ministry of Communications and Information Technology - April 2000 to December 2004 - Chief Investigator - Several products and services developed and deployed. Released to public in CD form by DIT, Govt. of India on 12 September 2005
4. Hexapod Control System - Euroflex Transmissions (India) Pvt. Ltd. - February 1999 to January 2000 - Investigator
5. Machine Aided Translation from English to Kannada - Government of Karnataka - Feb.1998 to Feb. 1999 - Chief Investigator
6. Spelling and Grammatical Error Detection and Correction - Department of Science and Technology (DST), Govt. of India (Under DST's Opportunities for Young Scientists Scheme) - April 1995 to September 1997 - Chief Investigator
7. POWERSHIFT: A conversion tool for converting PowerBuilder applications to Oracle Forms 4.5 Applications - Optima Software Services Ltd. - May 1995 to April 1996, - Investigator

Products and Services Developed:

Several products and Services in Language and Speech Technologies have been developed and deployed through transfer of technology and other means. Some of these have been released by DIT, Govt. of India to public on 12 Sept 2005. Here is a brief listing.

See www.LanguageTechnologies.ac.in for more details.

- UCSG Syntax, UCSG Shallow Parsing Architecture (FSM+HMM+A*), Wide Coverage Shallow Parser for English, UCSG Demo parsers for Telugu and Kannada
- Language Identification from small Text Samples - Multiple Linear Regression based Classifier - over 99% accuracy within Indian languages, pairwise
- Automatic Text Categorization System - SVM - MI for Dimensionality Reduction - Language Independent
- Speaker Independent Continuous Speech Recognition System for Telugu - htk
- Speaker Recognition/Verification System - Password Authentication
- Kannada dictionary, English-Kannada Dictionary, Kannada Thesaurus
- MORPH: Network and Process Model for Morphological Analysis and Generation, Kannada Morphological Analyzer and Generator
- MAT Architecture for Machine Aided Translation, English-Kannada Machine Aided Translation System for Government of Karnataka
- A 40 Million Word Text Corpus for Telugu
- AKSHARA advanced multi-lingual text processor - Akshara 1.1 released by DIT, Govt. of India to public on 12 Sept 2005
- DRISHTI OCR system for Telugu and other Indian languages
- Tel-Spell: Spelling Error Detection and Correction system for Telugu - ToT to Modular Infotech, Pune
- VIDYA: Comprehensive Web Based Education System
- VAANI: TTS system for Telugu
- Lexical Resources: Several Monolingual and Bilingual Dictionaries

- PSP2: A Natural Language Question Answering system for Harita Finance - Featuring Spelling Correction, Handling Ungrammatical/Incomplete Sentences and Context Based Interpretation
- CHART Parser for several Indian Languages, Language Games, Text Processing Tools in Prolog and Perl

Industry Links:

1. Media-Cognition Inc. - Technical Advisor, Media-Cognition India Pvt. Ltd.
2. Kalki Communication Technologies Pvt. Ltd, Bangalore - Member, Advisory Board on AI and NLP
3. New Horizons Cybersoft Ltd, Hyderabad - Member, Advisory Group
4. Consultant, Optima Software Services Ltd. Hyderabad
5. Consultant, Euroflex Transmissions (India) Pvt. Ltd. Hyderabad

Membership in Professional Bodies:

1. (IEEE)
2. SMATAC (Society for Machine Aids for Translation and Comunication)- Life Member
3. DLA - Life Member

Address for Correspondence:

Office: Home:

Dr. K. Narayana Murthy

Reader, Dept. of CIS 337, Doyens Township

University of Hyderabad Seri Lingampally,

Hyderabad 500 046 Hyderabad 500 019

INDIA INDIA

Ph: +91 040 2313 4017

FAX: +91 040 23010 780/145/120

email: knmuh@yahoo.com

web: www.LanguageTechnologies.ac.in