CURRICULUM VITAE

KIRAN PALA,

Date of Birth: 7th April 1983, Nationality: Indian.
Mail me: kiran.pala@research.iiit.ac.in or kiranap@gmail.com

Stay At: #No 233, NEW BOYS HOSTEL,
International Institute of Information Technology-Hyderabad.

EDUCATION:

International Institute of Information Technology-Hyderabad, India.
Pursuing Ph.D in Cognitive Science (Virtual Environments and Learning)

Jawaharlal Nehru Technological University, Hyderabad, India.
B-Tech in Computer Science and Information Technology.

State Board of Technical Education and Training, Andhra Pradesh, India.
Diploma in Automobile Engineering.

State Board Secondary School Certificate, Andhra Pradesh, India.
Secondary School Certificate.

SPECIAL COURSES:

Course Name
Psycholinguistics
Computer Problem Solving
Introduction to Cognitive Science
Advance Cognitive Science
Natural Language Processing
Cognitive Neuroscience
Cognitive Neurophysiology
Artificial Neural Networks

RESEARCH INTERESTS:

My research interests can be summarized as:

➢ Natural Language Processing (Building and processing linguistic resources and Development of NL interfaces).
➢ Designing and developing virtual environments with cognitive aspects.

- ➢ Computing and Education
- ➢ Learning and cognition (After Puberty and Adulthood),

TOOLS and LANGUAGES:

Tools: Weka, GATE, CRF
Matlab for signal processing and analysis,
PRAAT for phonetic analysis,
SPSS for statistical analysis,
PsychoPy for psychological experiments,
Eye-tracker for trace the eye-ball movement,
Audacity for signal recording,
RAPTOR for programming design

Languages and Scripts: C, JAVA, PERL, PYTHON, VC++, Javascript, PHP, JSP
Servers: Tomcat, Apache, Http
Databases: MySQL, SQL,
Packages: VB, .NET
Operating Systems: Linux, Windows, Cent, Suse
IDEs: Netbeans, Eclipse, Visual Studio etc.
Others: Libraries and tools used for research, especially those meant for Natural Language Processing

EXPERIENCE:

Sep-2011 to Date: Working as researcher. Project: Computer Adaptive Testing: A Formalism, IIIT-Hyderabad.

June-2011 to Date: Participating as researcher. Project: WBS for Computational Thinking. IIIT-Hyderabad.

Dec-2007 to Date: Anusaaraka Lab, Language Technologies Research Center, IIIT-Hyderabad. Project: **Anusaaraka** (Machine Translation System).

Mar-07 to Dec-07:   Fuji Japanese academy & LTRC, IIIT-Hyderabad. **Language Engineer,** project: **Trilingual Lexicon** for **Japanese-English-Tamil**.

June-05 to Mar-07:   **Panacea Dream weavers Software (P) Ltd**, Chennai, India. **Software Developer,** projects are: Language oriented application development, keyboard drivers, Multi-lingual text processor for Indic languages, Semi-automated font converters, etc.

Nov-04 to June-05:   Natural Language Engineering Lab, **University of Hyderabad**, Hyderabad, India. **Project participant,** project: **AKSHARA** (ADVANCED MULTILINGUAL TEXT PROCESSOR), Enabling Indian language support to a text processing framework.

INVOLVED MAJOR PROJECTS:

ONLINE COMPUTER ADAPTIVE TESTING (CAT)

The main motive behind this project was to build a CAT system with a generic formulism. Assessment and testing is directly related to education and training. It is a way to measure the level of performance and comprehension of the students in a group. In this era, using computer technology in testing and assessment has become one of the most important and considerable factor. While constructing a mechanism of Computer Adaptive Testing (CAT), different parameters should be taken into account. The Computer Adaptive Testing refers to those tests that are conducted online considering various parameters of adaptability.

WORK BREAKDOWN STRUCTURE FOR COMPUTATIONAL THINKING

According to NASSCOM (National Association of Software and Services Companies), the percentage of engineering graduates who are employable in software industry is at a low rate of 25%. The majority of the Indian software companies are service based and they have high demand for the people who are good in programming languages. The students need better of learning a programming language where they can adapt to themselves in solving complex problems. The WBS (work breakdown structure) is constructed in a way to help the students (especially the non programmers) to have an effective and friendly program-design model. The WBS is a four step program design model that includes steps like identifying the problem, constructing a solution (that includes Management model), verifying the solution and documentation. The motive behind this program design model is also to reduce the errors/flaws occurring in a program. One of the main cruxes of WBS is management Model which is based on the belief that "every computing problem is a management problem".

ANUSAARAKA (MACHINE TRANSLATION SYSTEM)

Anusaaraka is an approach to Machine Translation based on Information Dynamics, inspired by Panini's grammar. It enables a person who knows any one Indian language, to understand texts in other languages like Sanskrit, Telugu, Kannada, or English. However the user needs some training to use the software effectively. The Anusaaraka package is under General Public License (GPL), and will be distributed free. At present it is available only for Hindi speakers.

Internal search engine for corpus

A search engine has been developed to extract word frequency and analysis purpose from corpus (Gyananidhi and British National Corpus), this was built using Lucene (functions based on a stemmer to find the root word or morphological analyzer) API, which is core tool for search engines.

Corpus alignment tools (behind philosophy "Information should not loss")

To develop any tool or a product related to languages it must need proper resources, which can be dictionaries or huge database with related data.

Tools for Corpus Development

Using GATE-4.0 package plug-in, we have developed a set of tools package for corpus improvement according to the requirements. *English Tokenizer*, English *Sentence splitter*

(using honorifics), Named Entity, Hindi Tokenizer, Hindi Sentence splitter, Hindi Named Entity.

Paragraph and Sentence alignment

From the cleaned corpus, the paragraph information has been identified and extracted, made numbering accordingly. Those numbered information (anchors) has been considered as seed with a proper distance between anchors to the paragraph aligner's algorithm. The respective outputs will supplies to the sentence aligner; this entire process will be monitored through web-based (not-internet) user-interface.

Sentence aligner uses the Anusaaraka and Heuristics as input information and it applies on the given aligned paragraphs to get sentences aligned.

At present developing an interactive interface to organize those sentence output to evaluate as well facilitate to annotation of the text inputs according to standard format.

Input Tools for Anusaaraka

Text data will be in various formats doc, txt, html, pdf etc. we have developed a mechanism by using GATE-4.0 plug-ins to extract the content from various formatted files without disturbing the existing content i.e. information should not loss, and directs to Anusaaraka for translation purpose. It is a very useful mechanism to convert a webpage from one language to other.

An Interface for Word Sense Disambiguation

In NLP and WSD is one of the major problem, most of the AI systems performance efficiency was questioned at this stage, to overcome this kind of problem, good word-sense database required, which can handle the grammatical relations also. Building and evaluating such process will be a hectic and tedious job, to reduce such kind of fatigue on users. We have provided an interface; it takes minimum input from the users and in back-end constructs information into machine required format.

TRILINGUAL LEXICON DEVLOPMENT

This work has been started from scratch in-part of Visiting Researcher Program at IIIT Hyderabad, from Fuji-Japanese Academy; we collected domain based wordlists from various open-sources, and in this approach considered root-words are base words to compare the meaning among the languages. Final output has evaluated manually, this is a one of big resource for **Japanese-English-Tamil** languages.

FONT & ENCLODING CONVERTORS

Online and offline encoding (ISCII to UTF-8 vice versa) and font convertors (one font to other font in same encoding) has been developed for Tamil and Hindi languages, it functions by reading and replacing with required characters in matrix form as well Akshar-based method, to reduce the redundancy of the rendering matters.

KEYBOARD DRIVERS

International Phonetic Association (IPA) keyboard driver has been designed and developed in part of dictionary development work at Panacea Dreamweavers, Chennai. I have handled issues during design

and development of project, like navigation, appearance, shape and functional issues are how to create a link between systems files and third-party software (page maker, MS office, and other editors),

User-Controls for Indian language Text Editors
I was a linguistic supporter and developer in Panacea Dreamweavers, Chennai. I have involved in development of offline Microsoft operating systems based text editors for Indian languages, especially for Tamil and Telugu. Functionally, it facilitates typing and displaying on the editors. We have developed user controls i.e. dynamic link libraries (DLL) for windows.

AKSHARA (ADVANCED MULTILINGUAL TEXT PROCESSOR)
I was involved in this project during my under graduation; this was funded by Dept. of Information Technology (DIT), developed at University of Hyderabad under the supervision of Prof. Kavi Narayana Murthi. Technically, it is a multilingual text processor; implemented for 13 major Indian Languages using Java as a programming language. It supports Microsoft Operating Systems, and the Linux. I was dealt with font encodings of ISCII and ASCII format. At back-end ISCII encoding behaves as a back-bone between fonts and encodings of the text. With this technique without UTF-8 support also the text can be manipulate accordingly. We have chosen this technique, because-off due to lack-off proper UTF support for Indian languages as well nature of rendering between characters. For this fonts and a text processor are developed separately and integrated, according to nature of Indian languages.

RECENT ACADEMIC ACTIVITES:

Involved in guiding of one UG student on their final year project

As a mentor guided 6 UG Interns in summer 2011

As a mentor guided 2 students (UG and PG) on their final year projects

Teaching Assistant (TA) for *"Computer Problem Solving"* Course

RECENT AWARDS and FELLOWSHIPS:

Selected for "Young Researchers Conclave" at IIT Gandhinagar, December-2011

Summer School on "Language and Communication" at NIAS, Bangalore-2011

Full fund to participate in "Yahoo Summer School on Information Retrieval" at IISc Bangalore-2011

Full fund to participate in WWW2011 Conference (Nixi-fellowship)

CIIL –NTS Module writing fellowship2011

PUBLICATIONS:

PUBLISHED:

1) *Experience of Speech Perception Mediates in Lexical Learning: An Experimental Study* **Kiran Pala** will appear in proceedings of ISPI 2011, November-2011.

2) *Games for Academic Vocabulary Learning Through a Virtual Environment.* **Kiran Pala** will appear in IALP 2011 IEEE Transactions, November-2011.

3) *An Experiment on Resolving Pronominal Anaphora in Hindi: Using Heuristics.* **Kiran Pala** and Rafiya Begum. Published in *Springer-CCIS (Lecture Notes in Computer Science)*, March-2011.

4) *The Relevance of Variations in Auditory Perception for Second Language Teaching and Learning.* **Kiran Pala**, et al. In Proceedings of the 4th Language Technology Conference (LTC)-Human Language Technology (HLT), **LTC09**. Poland. 2009.

5) *Enhancing effectiveness of sentence alignment in parallel corpora: Using MT & heuristics.* Sriram Chaudary, **Kiran Pala**, et al. In Proceedings of the Sixth International Conference on Natural Language Processing **(ICON)**. Pune, India. 2008.

6) *Estimating the Cost of Adapting the Resources of One Language for Another.* Anil Kumar Singh, **Kiran Pala** and Harshit Surana. In Proceedings of the Sixth International Conference on Language Resources and Evaluation **(LREC)**. Marrakech, Morocco. 2008.

IN PRESS:

1) *Challenges and Opportunities in Automatically Building Bilingual Lexicon for Telugu from Web Corpus,*

2) *Visual Experience in Cognitive Structures: An Experimental Study,*

3) *Decision making in Program Design for Problem Solving: Learning Perspective,*

4) *Three Kinds of Text Input Methods for Languages that Use Indic Scripts: A Comparison,*

5) *Computer Adaptive Testing: Program Designing,*

6) *Why Does "Lexical Access Preference" Exist Bilinguals*

IN PREPARATION:

PAPERS

1) *Computing as Management*

2) *LAP is a Cognitive Structure: Mediates in Vocabulary Learning.*

3) *Semi-supervised verb frame learning: Cluster-based Approach.*

4) *Knowledge Extraction from Web for Building Large Scale Lexicons: Template base Approach.*

BOOKS

1) *Computing as Management*: *WBS,* participating as a Co-author

2) *Virtual Environments Can Mediate Continuous Learning*: A Book Chapter.

UNPUBLISHED:

Grammar and Multilingualism: A Comparative Approach Tamil-Japanese. **Kiran Pala** Unpublished manuscript, IIIT Hyderabad.

PRESENTATIONS:

Statistics in Social Sciences Research. **Kiran Pala.** Presented at EnhanceEdu, IIIT Hyderabad, India, 2011.

Effect of Technology on Perception in-terms of Language Contact in Multilingual Society. **Kiran Pala.** Presented at the South Asian Languages Analysis Roundtable 29. Mysore, India, 2011.

A correlational Condition of Academic Background and Cognates. **Kiran Pala.** Presented at EnhanceEdu, IIIT Hyderabad, India, 2011.

Anusaaraka (Translation) engine for Webpages, Word document, pdf or text from English to Hindi based on. **Kiran Pala.** Presented In the Workshop on Free/Libre/Open Source Software (**FLOSS)**, Delhi. January 2010.

A Heuristic approach for Hindi Anaphora Resolution. **Kiran Pala,** Sai Gollapudi Rafiya Begum**.** Presented In the NLP Group at LTRC IIIT-H**,** Hyderabad, December 2009.

A Proposed Approach for Testing the Hypothesis that Cognates & Background Information Together Facilitate Vocabulary Learning. **Kiran Pala.** Presented In the International Conference of Language and Cognition Interface **(ICLCI),** Allahabad, December 2009.

Anusaaraka cum Machine Translation system. Sriram Chaudary, **Kiran Pala**, et al. At the R&D Showcase, International Institute of Information Technology-Hyderabad, India. 2009.

Modeling user categorization for language learning. **Kiran Pala**, Prakash Chandra Modal and Anil Kumar Singh. Presented At the 29th Annual Conference, Linguistic Society of Nepal. Khatmandu, Nepal. 2008.

Philosophy of Anusaaraka. Sriram Chaudary, **Kiran Pala**, et al. At the R&D Showcase, International Institute of Information Technology-Hyderabad, India. 2008.

Semi-supervised verb frame learning given seed frames: A Cluster based approach **Kiran Pala**, A. Ananth Ramakrishnan, et al. At the Winter School on Natural Language Processing, International Institute of Information Technology-Hyderabad, India. 2008.

Grammar and Multilingualism. **Kiran Pala** and Anil Kumar Singh. At the National Seminar on the Emerging Linguistic Scene in North East India. Shillong, India. 2007.

EVENTS ORGANIZED:

Organizing Member in Workshop on Excitement of Research ExOR (October 2011).

Organized a Workshop on Social Networking WebApps (April 2011).

Organized a National level Workshop on Applied Computer Science (WACS-2010 December).

Organized a Workshop on Python Programming, at IIIT-Hyderabad (November 2010).

RECENT ACADEMIC ACTIVITIES:

Participated at WORKSHOP ON PRONUNCIATION LEXICONS FOR INDIC LANGUAGES, JNU-Delhi, India (May-2011).

NIAS Summer School on Cognition and Mathematical Models (April 2010)

Active Participation in organizing CMU Winter School on Learning Sciences, December 2009, International Institute of Information Technology-Hyderabad, India.

Participated at Workshop on Ontology, NLP, IE and IR-2008, Indian Institute of Technology-Mumbai, India.

Participated in ICON-2007(International Conference on Natural Language Processing), International Institute of Information Technology-Hyderabad, India.

Participated as Resource Person from the Organization LRIL-2007(Lexical Resources for Indic Languages), CDAC- Mumbai, India.

EXTRA CURRICULAR:

- Branch Counselor, CSI (Computer Society of India) Student Branch IIIT-Hyderabad.
- Participating in various treks and travels.
- Participating in debates and science exhibitions.
- Active participation in social activities.
- Active participation in sports and cultural events in annual celebrations.

LANGUAGES KNOWN:

Native: [Telugu, Tamil], **fluent**: [English, Hindi, and Malayalam], **intermediate**: [Spanish, Kannada], **elementary**: [Greek, Japanese, and Bengali].

REFERENCES: available on request.