Fourth Meeting - Minutes

1. **Welcome**
   Prof. Rajesh Sachdeva, Director, Central Institute of Indian Languages and Chairperson, Linguistic Data Consortium for Indian Languages (LDC-IL) welcomed the Members of the Fourth Project Advisory Committee meeting.

2. **Release of LDC-IL Publications**
   Shri R P Sisodia, Director (Languages), Ministry of Human Resource Development released the following LDC-IL Publications;
   - Morphological Analysers and Generators
   - Part-of-Speech Tagging
   - Creation of Multilingual Speech Resources: Academic & Technical Issues
   - Making of Electronic Dictionary

3. **Agenda Items**
   After the Welcome, the agenda items were taken up in the order.

   I. **Confirmation of the Minutes of the Third PAC Meeting**
   The Minutes of the Third Project Advisory Committee Meeting of the Linguistic Data Consortium for Indian Languages (LDC-IL) held on December 8, 2008 were confirmed.

   II. **Action Taken Report**
   The actions taken on the recommendations of the Third Project Advisory Committee Meeting were explained item-wise.

   III. **Presentation on Review of Progress**
   Dr. L. Ramamoorthy, Reader cum Research Officer & Head, LDC-IL made a presentation on progress of work in LDC-IL from December 8, 2008 to July 31, 2010.

   IV. **Annual Plan : 2010-11**

   The proposed annual action plan for the year 2010-11 was considered and approved.
   During the review of progress, some of the issues related to text/speech corpus were discussed.

   A. **Text Corpus :**
   LDC-IL is downloading the data from leading newspapers for the project work. The PAC advised to verify whether such act attracts any copyright issues or not. If so, the collected data cannot be published or utilized without the consent.

   B. **Speech Corpus :**
   It was suggested that, collecting fine speech data by using advanced recording device is not necessary for automatic speech synthesis. It is better to have real life information by collecting telephonic data in 8HZ sampling rate. However, it was accepted that the data LDC-IL has, will be useful for ASR.
   The raw data of the speech corpus should be evaluated to check the quality of the data before taking up annotation work. The annotation at the level of utterance/sentence level is enough for ASR. However, LDC-IL will have annotation upto phone level for multipurpose analysis.
   The sample data has to be made available in the web.

C. **Standard Tag set:**

    It was felt to normalize the standard tag set across the research groups working on NLP. BIS has evolved a tag set after consulting concerned institutions and PAC advised LDC-IL to follow the BIS Tag set.

V.   **Consideration and Discussion of the Licensing issues:**

    Quality of the data: The data should be validated by the externals/experts to ensure quality data.

    Price: It has to be fixed by the costing committee by looking into the various kinds of data and different categories of users. Evaluation of the price has to be made every 2-3 years.

    Membership Fee: It was suggested to re-fix the membership fee by considering the different categories.

    Free Data: It was felt, providing huge quantity of data to the members at free of cost is not advisable. Only they can have membership but they should not take all data for full year. Every data must be priced.

    Pay per use: It was strongly recommended to provide specific data for specific price depending upon the requirement of various kinds of users.

Data Security: It was suggested to make proper security for the data available in the LDC-IL. It was also suggested to follow DRC norms to secure the safety of the data.

Deadline: It was told to fix deadline for finalizing the Licensing Policy. The draft policy may be circulated among the members.

Copyright issues: It was advised to include Shri Venkata Rao, Vice-Chancellor, National Law College, Bangalore as a member to the Licensing Policy Committee to obtain suggestions while resolving the copyright issues.

VI (a). Acquisition of Corpus from Publishing Houses:

It was proposed to acquire different kinds of data from various leading publishing houses. Committee felt that, they will charge more for the ready data and unless the price was fixed by the costing committee for the data distribution, purchase of the data also cannot be done. However, LDC-IL is advised to acquire data if it is freely available.

VI (b). Constitution of Costing Committee:

It was felt that, separate costing committee need not be constituted. Licensing committee can look after these matters. However, it was suggested to expand the Licensing Committee.

VI (c). Granting balance amount to IIT, Mumbai

Prof. Pushpak Bhattacharya, IIT Mumbai explained about the project titled "Sanskrit WordNet". It was decided to sanction remaining grant by considering the report submitted by them.

VI (d & e). Administrative Issues

It was told that, i) Extension of Foreign Service of Shri M. Venkatesan, Maintenance Engineer, LDC-IL and ii) Re-engagement of Shri R. Parthasarathy, Maintenance Officer (Administration & Accounts)I/c, LDC-IL has to be decided by the Ministry. The issue was brought to the notice of the PAC members.

VII. Other matters

    A. **Structural issues:**

        It was agreed to expand the members of each Working Group. The Working Groups have to make preliminary observations and guidelines and submit report for further process.

    B. Grant-in-Aid issues:

Expansion of the Committee: At present, the LDC-IL has 5 members in the GIA Committee. PAC advised to expand this committee and include 2-3 members from PAC.

Sanctioning of projects: PAC advised to invite grantee's also to the GIA Meeting to give presentation and brief explanation of their proposals. Preliminary evaluation has to be made by the GIA committee before approving the proposal and before inviting the grantee. After getting the consent of the members only, funding of the projects can be made. Further, it is advised to update the details in the LDC-IL website with respect to GIA.

C. Web-site update:

The committee advised LDC-IL to host the history of LDC-IL from its conception, and acknowledge the institutions involved in it etc., in the website.

The next meeting of the LDC-IL PAC will be held at Mysore within 3 months.

(RAJESH SACHDEVA)
Chairperson, Project Advisory Committee, LDC-IL &
Director, Central Institute of Indian Languages, Mysore

	MEMBERS PRESENT	
1	Prof. Rajesh Sachdeva Director, CIIL, Mysore	Chairperson
2	Shri. R. P. Sisodia Director (L), MHRD, New Delhi	Member Representing Language Bureau
3	Prof. Rajeev Sangal Director, IIIT, Hyderabad	Member
4	Director Indian Institute of Technology, Bombay	Member Represented by Dr. Pushpak Bhattacharya
5	Director Indian Institute of Technology, Madras	Member Represented by Prof. Hema Murthy
6	National Law School University Bangalore	Member Represented by Prof. R. Venkata Rao, Vice Chancellor
	SPECIAL INSTITUTIONAL INVITEES	
7	Director Indian Institute of Technology, Kharagpur	Member Represented by Prof. Anupam Basu
8	Director All India Institute of Speech & Hearing Manasagangotri, Mysore – 570 006	Member Represented by Dr. Shyamala
	SPECIAL INDUSTRY INVITEES	
12	MICROSOFT, Bangalore	Represented by Dr. Kalika Bali
13	MOTOROLA, Bangalore	Represented by Shri Shailesh Ramamurthy
14	Dr. L. Ramamoorthy RRO & Head, LDC-IL, CIIL, Mysore	Member-Convener

	MEMBERS ABSENT
1	Finance Division, MHRD, N. Delhi
2	MCIT, New Delhi
	SPECIAL INSTITUTIONAL INVITEES
3	C-DAC, Pune
4	One person from e-governance of DIT
	SPECIAL INDIVIDUAL INVITEES
5	Prof. C.N. Krishnan (AU-KBC, Chennai)
6	Prof. Peri Bhaskararao (ILCAA, Tokyo University of Foreign Studies, Japan)
	SPECIAL INDUSTRY INVITEES
7	HP Labs
8	IBM
9	MAIT
10	Google