Minutes of the 4th Dravidian Languages Workshop held on

20th and 21st January, 2015

- Prof. Pushpak inaugurated the workshop by his Welcome speech.

- Prof. Rajendran spoke about the current situation of Dravidian Language WordNet groups.
  - He gave a clear picture of how staff recruitment is stagnated due to lack of funds and how the old staff has been working without their salaries.
  - He requested a combined effort to be made to convince funding agencies to release the previous funds.
  - He also mentioned that despite all the difficulties, the work should continue for the sake of our culture and language, as it should be considered a duty towards the nation.
  - Prof. Pushpak was praised for his worldwide efforts for Indian NLP community.

Presentations:-

- Presentation on Kannada WordNet was given by Prof. Bisembli Hemananda.
  - He also suggested that the names of all the districts and villages of the respective states with details about latitude and longitude to be added as Synsets, which can be linked to Google maps.
  - An interface was suggested for WordNets with proper ‘native’ pronunciation for WordNet concepts along with images.

- Presentation on Malayalam WordNet was given by Prof. S. Rajendran.

- Presentation on Telugu WordNet was given by Dr. Keshav Murthy.

- Ongoing WSD work in IIT Bombay was presented by Sudha.

- 'Hindi WordNet Android Application' along with Input Mechanisms was presented and demonstrated by Diptesh.
  - He suggested the use of open source responsive web layouts for WordNet websites, thus leading to creation of a mobile application.
  - A possibility of an iOS based mobile application was also discussed, using similar methods.

- PaCMan Workbench was presented and demonstrated by Diptesh.

- IL-MultiDict Tool for WordNet Synset Creation was demonstrated by Diptesh.

- Sense Marker Tool was demonstrated by Diptesh.
- Malayalam language group requested a separate copy of Sense Marker Tool, along with the embedded latest Malayalam Dictionary in order to Sense Mark.

- Morph Analyzer and Lexicon Entry tool were presented and demonstrated by Diptesh.

- 'Human Mediated Generic Stemmer' based on TRIE data structure was demonstrated by Diptesh.
  - After trying a few words from all the Dravidian languages, Prof. Pushpak suggested Dravidian Language TRIE should be re-constructed.

- Hanumant demonstrated the Indradhanush website with all the downloadables for creating resources among the Dravidian group.
  - He also demonstrated the websites of languages under Indradhanush project.

- Dhirendra presented his work on MultiWords, and detailed the challenges in identifying MultiWords, along with their classification and applications.
  - He also professed the need to add MultiWords in WordNets.
  - Prof. Rajendran inquired whether MultiWords are synonymous to Samaasas (Compound Words), which resulted in a discussion between Prof. Pushpak and Prof. Rajendran.

- Irawati presented their work on Semi Automatic approach to Populate Sanskrit WordNet using Monier-Williams Dictionary.

- Sudha presented her work on Word Embeddings and how they could be useful in creating WSD systems for Dravidian languages.

- Hanumant presented their work on World WordNet Database Structure.
  - He was suggested to add screenshots and comparison between Indian Language WordNets, and other foreign language WordNets.
  - He was also suggested to add Hindi WordNet data to the database structure.

- Neha presented her study on Crowdsourcing and Gamification of WordNets and our current resources.
  - Prof. Pushpak suggested her to deepen the reading with more papers.
  - She demonstrated the Hangman game created using WordNet as a resource.

- Apurv presented his work on Linked Open Data and demonstrated RDF structure for Konkani WordNet.

- Arjun presented his work in CLIA, and demonstrated 'Sandhaan', an Indian Language search engine.
  - He explained how IndoWordNet as a resource is helping Sandhaan.

- Rohit presented his study on how MultiDicts created using IndoWordNet is being used in Indian Language to Indian Language Machine Translation scenario.
Gajanan presented the applications of multi-dictionaries in the English to Indian Language Machine Translation System and showed the status of sense-tagged corpus using multi-word dictionary and what are the current steps being taken to improve.

Decisions:-

• All Dravidian language WordNet groups also need to approach their respective state governments with appropriate proposals for funds.

• Schools need to be informed about using WordNets as a teaching aid, which in turn would also get us much required feedback.

• Telugu and Malayalam WordNet sites were set up and hosted at IIT Bombay, using the data they had received earlier.

• Kannada and Tamil teams need to send their webpages along with databases, so that beta versions of their websites can also be hosted.

• Prof. Pushpak suggested the propagation of WordNets to public, via Social Networking.
  o It was suggested that WordNet groups be active over Twitter, LinkedIn, Facebook, Academia etc.
  o Facebook page for IndoWordNet was created in order to publicize it, and let people know about such a freely available resource.

• Prof. Pushpak advised Hanumant to send the previous proposal to Sandhya, Rajita, Neha, and Apurv.
  o Prof. Pushpak advised Rajita, Sandhya and Lata to prepare an Educational module along with facts and anecdotes for the proposal.
  o Prof. Pushpak requested Prof. Bisembli Hemananda to revise and re-plan the budget of the proposal, along with the facts pertaining to the manufacturing costs of lexical resources like dictionaries etc.
  o Neha and Apurv were advised by Prof. Pushpak to craft a Gamification module for the project proposal with research about the cost of rendering such an application.
  o Hanumant was asked to mention the 'Hindi WordNet Android Application (Beta)' in the proposal, and cite further such steps in accordance with Diptesh.
  o Hanumant will co-ordinate to create a final project proposal along with the above suggested modules.

• Prof. Pushpak suggested Medari Tham to approach the North Eastern council for funds regarding the project.

• Prof. Hemananda proposed a three step validation plan for synsets.
  o Prof. Pushpak suggested the use of an interface for doing such a task, where a check list can be formed for validating a particular synset on each step.
Diptesh and Prof. S. Rajendran discussed the possibility of one such tool among the group, and also a tool which can suggest examples based on a concordancer like search module.

**Discussions:**

- TRIE algorithm for generic stemmer was discussed by Prof. Pushpak and Prof. Hemananda.
  - It was observed that the 'schwa deletion' phenomena is absent for Kannada language.

- Budget proposals and financial aspects of WordNets were discussed, and a revised plan was distributed among the group.

- Gamification was discussed and the need for a deeper insight was professed by Prof. Pushpak.

- All the Dravidian WordNets should set up webpages using the CMS templates and send them to IIT Bombay for hosting.

- All possible sources for generating funds should be explored by the individual groups.

- It was discussed that WordNets should contain images for a better understanding of the concept.