Winter School for Women

The Winter School for Women 2019, on *"Natural Language Processing (NLP) & Machine Learning (ML)"*, organized by International Centre for Free and Open Source Software (ICFOSS) in collaboration with central institute of Indian Languages (CIIL) and co-sponsored by Kerala IT Mission, was held from Jan 15th to 25th 2019 at Club House, Technopark. The school convened women who were Post Graduate students, Research scholars, Faculty and Professionals from industry. The aim of the school was to improve women participation in FOSS communities as well as mould the individuals in the field of Natural Language Processing. The invited talks and hands-on sessions provided an opportunity for the participants to acquire individual attention of mentors and to develop network between themselves. The winter school encouraged the participants to come up with the best and brightest ideas in various areas of natural language processing and machine learning.

Participants

A total of 42 women participants attended the school, with 30 registered participants and 12 women researchers from ICFOSS as internal participants. The details are provided.

Registered participants:

PG Students	-	4
Research Scholars	-	10
Faculty	-	13
Professionals	-	3
Internal participants	-	12

Highlights of the Event

Day 1 : 15-01-2019

The formal session of Winter School for Women on *"Natural Language Processing (NLP) & Machine Learning (ML)"* commenced with the inaugural function at Orchid Hall, Club House, Technopark, on 15th January 2019 at 10:45 AM. Shri. M Sivasankar IAS, IT Secretary, graced the occasion as the Chief Guest and Dr. Rajasree M, Principal of GEC Barton Hill, as the Guest of Honor, while Dr. Jayasankar Prasad C, Director of ICFOSS and Dr. Rajeev R R, Program Head (e-Governance and Development), ICFOSS, presided over the

function. Dr. Rajeev R R formally welcomed the dignitaries on the dais and participants of the school. He made the audience aware about the origin and significance of ICFOSS and mentioned its achievements in the free and open source community. Shri. M Sivasankar IAS, in his inaugural address, praised ICFOSS for taking the initiative to put forth an effort in devising a programme involving only women participants and mentioned its significance in the current society. Dr. Jayasankar Prasad addressed the audience and expressed the contentment to instigate a programme, which is novel to the state and ensured that similar activities would continue to be part of ICFOSS in the future. The Guest of Honour for the day Dr. Rajasree M enlightened the participants with her thought provoking speech. *The formal inaugural session ended at 10:55 AM.*

Session 1: Day 1

The first session of the day by Dr. Rajasree M on "Research, Technology and Realization" started at 11:05 AM. She discussed about the evolution of NLP over years and the biggest open problems in NLP. The participants was introduced to the research opportunities in Natural Language Processing (NLP) and Machine Learning (ML) and she directed their attention to the major research projects in the field and the technology trends of 2018. Technologies such as Artificial embryos, 3-D metal printing and Blockchain, were pointed out during the session and the concepts of Dueling Neural Networks, Genetic fortune telling were also discussed. The session ended at 12.00 pm.

After the session, participants introduced themselves and detailed on their current NLP/ML experience status. Dr. Rajeev R R briefed on the issues faced by NLP researchers on each stages of their projects and on different NLP tasks.

The session gave a brief introduction about the Artificial Intelligence area. It helped the participants to familiarize with the winter school theme – Machine learning and Natural Language Processing. The participants got insights about current trends and technologies in the research field.

Afternoon: Session 2- Day 1

The next session was by Dr. Richard M Stallman. The topic was What Free software means. The session started at 3.00 pm. Dr. Stallman talked about importance of software freedom. The session discussed free software movement in detail. The session ended at 6.00pm.

Session by Dr. Richard M Stallman emphasized on the the idea and philosophy behind Free and Open Source Software (FOSS). The session was an opportunity for the participants to meet the person behind the commencement of Free software movement and to realize the importance and motives behind building an organization like ICFOSS.

Day 2 : 16-01-2019

Session 1 - Day 2 Forenoon:

The session, by Dr. Girish Nath Jha on "Computational Linguistics in multilingual societies", started at 10.25am. He gave introduction about the concepts in NLP. and discussed about Speaker recognition. Speaker recognition is the process of identifying the speaker from a speech input. Feature extraction is performed on the input. These are compared with reference templates and similarity is measured to find the speaker. Some interesting topics like Automatic speech recognition, which is very important in the case of Indian languages as most of people don't know how to read and write.

The other important topic discussed during the session was, Parsing. Dr. Jha shared his experiences in making parsers for Sanskrit. He explained the challenges in the research area of NLP/COLING(Computational Linguistics). He discussed common NLP problems like ambiguity, machine translation and also discussed about different tools used in the research.

Dr. Jha shared his practical experience in developing different tools for sanskrit language. The methods used for defining grammars, parsing mechanisms, solving ambiguity were discussed during the session.

The participants dispersed for lunch at 1.20 pm.

In the session participants were famialiarised with computational linguistics as a sub branch of AI. The evolution of this research area and its importance in the current society was discussed in detail.

Session 2 - Day 2 Afternoon:

The session by Dr. Dipti Misra Sharma about "Computational Morphology" started at 2.00 pm. Dr. Dipti talked about basic concepts of morphology. She started the lecture by talking about words. Words are of two categories: Open as well as closed classes. She took examples from English language to explain the basic concepts.

Dr. Dipti explained about morphemes, the basic meaningful units of a language. 'Ladak' in Hindi language is an example for the morpheme. The words 'Ladaka', 'Ladaki', 'Ladake' are formed from this morpheme. Morphemes are of two categories, Bound and Free morphemes.

Then the discussion moved on to another basic but important concept, Stems and Roots. With an example of word 'boy', the concept is explained. A root carries the primary content or meaning part of a word. Both words boy and boys 'formed' from the same root - boy. A stem is a content part of a word which is derived from another root. Example, arrangement (arrange+ment). Affixes are another concept discussed during the talk. Affixes are units that are attached to a root or stem to form a word. Affixes can be inflectional or derivational. Dr. Dipti gave examples from English as well as Hindi for affixes. Then gave an exercise to the participants to find different inflections of a word in Malayalam.

The word formation process is discussed. And Morphology is discussed in detail. There are two categories,

1. Derivational Morphology

2. Inflectional Morphology

Inflectional morphology is process of inflecting words with grammatical information such as number, gender, person etc and Derivational morphology is the process of creating new words from the existing words. The hierarchy in word structure was the next topic. Paradigm, Set of all word-forms of a lexeme/root is explained with examples from Hindi language.

After tea break, speaker discussed about types of Languages based on morphology. Isolating/Analytic languages are the languages that has very little morphology. Examples are Chinese and Vietnamese. Agglutinating languages are morphologically rich languages. Malayalam, Turkish, Finnish are the examples. Inflectional/fusion language is a type of synthetic language which has a tendency to overlay many morphemes in a way that can be difficult to segment. Sanskrit , Greek, Russian are some examples. PolySynthetic is the last type. Its a traditional term for languages or morphological systems which permit processes such as noun incorporation

The session discussed about Computational Morphology, the way automatically analyse/generate words. There are rule based and statistical approaches. Rule based method is easy to implement. There should be some rules written for the morphological analysis. Finite

state transducers (FST) is one of the major method used in rule based systems. Rule based solution gives multiple solutions for same case. Whereas the statistical one gives the best in the context as output to the morphological analyser. The advantages and disadvantages of rule based method are discussed in detail. She gave case studies of building morphological analysers in Indian Languages. She shared her practical experience regarding the same. Dr.Dipti concluded her lecture with a question that " Which method is better for your Language - Rule based or the statistical one?". There was a five minute Q&A session followed by the lecture.

After session by Dr. Dipti, a session about basics of NLP was given by Dr. Rajeev R R of ICFOSS. Basic concepts in NLP and CL are discussed during the session for the participants who are from non- NLP background. Difference between Natural language processing, computational linguistics and language technology is discussed. Some concepts like POS tagging, morphology, phonology were discussed. Technological aspects of a language is representation of the language. Unicode, ASCII representations, Speech and text processing were discussed during the session. Dr. Rajeev gave some examples of NLP and AI in daily life. Different applications in the area were mentioned.The lecture ended at 6.30.

This session discussed basics of morphology. Fundamental concepts of natural language processing were discussed in detail. It helped the participants to understand how a language could be viewed in a computational perspective.

Day 3 : 17-01-2019

Session 1 - Day 3 Forenoon:

The session by Dr. Umesh P from College of Engineering, Aranmula started at 10.05 am. Dr. Umesh handled an introduction to Python. Basics of Python is discussed in the class. He introduced command line python coding. The lecture started with basic arithmetic operations like addition, subtraction etc. And data types and fundamental programming elements were introduced. Python packages like numpy, Scipy, pandas,matplotlib were installed during the session.

This session covered practical applications of language processing using Python. Basics of python, implementing concepts like neural networks, machine learning algorithms were discussed . It was a an interactive session in which participants got a chance to practically experiment the concepts discussed during the previous sessions.

The class covered basic python lessons. Example programs to familiarize with concepts like list, tuples, dictionaries , string operations were done as part of the session. Regular expressions, Conditional operations, Loops in python and file operations were introduced. Pandas library familiarization is given. Arrays in python were handled using pandas library. The class dispersed for lunch at 1.00 pm.

Session 2 - Day 3 Afternoon:

After lunch the session by Dr. Umesh continued. The session topic was "Introduction to ML, Linear Algebra Review- Probability and statistics". Different ML algorithms and their implementation were discussed in detail. The speaker showed solving Linear equations using python libraries.

Regression models were built using scipy and pandas. Data for the ML algorithm training were given to the participants and the algorithms were implemented. Artificial Neural networks and deep learning basics were introduced. Examples using Tensorflow and Keras were demonstrated to familiarize the participants with the concepts. The class ended at 5.00 pm followed by a 15 minutes Q&A session.

The participants were grouped into 6 groups for the project idea presentation scheduled on the final day. 6 different real life situations were assigned to them to find a problem in that area. The group members have to find a solution to the problem.

Session 3 - Day 3 Afternoon:

An interactive session by Ms. Gopika Kumar on Dance & Theatre expressions followed. The participants interacted to each other in the session. The program ended at 7.00 pm.

Day 4 : 18-01-2019

4th day of the school opened with a talk by Dr. Asharaf S, Professor, IIITMK. The topic of the session was "A Gentle Introduction to Machine Learning Algorithm". Dr. Asharaf started the lecture with some real life examples. He talked about learning and memorization. Examples like "How a kid learns his mother's face", " How pigeons remember" were discussed. He talked about various topics like Blockchain technology, Autonomous Cars, Machine Learning Translations.

The basic concepts of Machine learning and ML algorithms were discussed. The three categories of learning in Machine learning are Supervised learning, unsupervised learning, reinforcement learning. A training phase is required for the implementation of ML algorithm . Thereafter a trained model is built on the data given for learning process. And then the model is tested on some test data after the model construction. Spam filtering application was discussed as a practical application of ML.

After the tea break, the session continued with ML versus deep learning .

The session ended at 12.25 pm. A Q&A session of 10 minutes followed the session.

After the session, Mr. Aron G, Intern of ICFOSS gave a practical session on Application of Machine learning using javascript. The participants dispersed for lunch at 1.00 pm.

The session discussed about various machine learning algorithms with some real life examples. The main outcome of the discussion was, participants got a revelation of " How Machine learning differs from deep learning ".

Session 2 - Day 4 Afternoon:

The session by Ms. Jannath H S, Research Scholar from IIITMK started at 2.00 pm. The discussion topic was " Supervised Learning --- Classification , Regression". Ms. Jannath gave an introduction to Machine learning. Types of machine learning are Supervised learning, unsupervised learning, reinforcement learning. The difference between supervised and unsupervised learning was discussed. Classification and regression are two kinds of supervised machine learning algorithms. She talked about perceptron learning.

The historical background of neural networks discussed. Back Propagation, an abbreviation for "backward propagation of errors" , is a common method of training artificial neural networks used in conjunction with an optimization.

Classification algorithms like decision trees and support vector machines were also discussed.

After the tea break, Regression models were discussed. Regression modeling was explained in detail. Two regression models, linear regression and logistic regression models were discussed in detail. The session ended at 5.00 pm.

A interactive session by Miss. Aardra Chandramouli started at 5.30 pm. Ms. Aardra, a successful woman entrepreneur shared her experience with the participants. The session ended at 7.00 pm.

The discussion gave detailed insight about machine learning algorithms to the audience. The different approaches and methodologies used in machine learning research area were mentioned. This session was a continuation from forenoon discussion.

Day 5 : 19-01-2019

Session 1 - Day 5 Forenoon:

The session by Dr. Sobha Lalitha Devi from AUK- BC, Anna University Chennai started at 10.00am. The topic for the discussion was "POS tagging". Dr. Shobha started the lecture by talking about the shallow parsing. It can be carried out on full sentences or partial sentences. From the sentences we want to find out the subject, verb and objects . Then She explained the concept discourse analysis. After that the topic moved on to the analysis in Malayalam sentences. But when it comes to Malayalam sentences, first it is required to find out the word boundary, then perform the morphanalysis to find the out the pos, only after that we can extract the subject, verb and object. Then discussed about preprocessing, which involves Sentence Splitter to separate text into sentences, tokenizer to separate sentence into words , Morphological analyzer to analyse words, POS Tagger to find out the ambiguity in words, – clause identifier, semantic role identifier, Multi word Identifier etc,. Then she introduced various systems NLP, such as Machine translation, Information extraction, information retrieval, Question answering and Semantic analysis.

Then the session moved on to discuss about depth parsing, which is defined as the process of assigning structural description to sequences of words in a NLP. She explained with some example sentences. Next topic of discussion was part of speech tagging, she introduced penn treebank, 61-tag C5 tagset, 146-tag C7 Tagset for tagging english sentences, Brown corpus etc,. Then about BIS tagset for tagging Malayalam sentences.

The session dispersed to a tea break at 11:25 am then renewed at 11:55 am. It started with discussing different POS taggers which is classified into 2 main categories such as rule based and stochastic. Rule based methods are mainly based on hand-written rules, on the other hand stochastic methods adopt frequency, probability, HMM metrics. Then mentioned TBL, Open Source POS tagger for English which uses supervised learning, then discussed advantages and disadvantages of TBL. Next topic of discussion was chunking, explained chunking with examples. A typical chunk consists of a content word surrounded by a constellation of function words. She explained variou uses of chunkers. Chunking is a different stage process, which involves word identification, chunk identification (NP, VP, text chunking). Then explained various chunk tagset such as NP, VGF, VGNF, CCP, VGNN etc,. The last topic of discussion was clause boundary identifier. Morning session wound up at 1:00 pm.

The session covered one of the most important concept of the language technology area. Parts of speech tagging, the basic requirement for language processing was discussed in detail.

Session 2 - Day 5 Afternoon:

Afternoon session started at 2:15 pm and was addressed by Dr. Deepa P Gopinath Training Officer at Directorate of Technical Education. Topic of the session was Speech Synthesis. Session started with introducing various general problems regarding malayalam speech synthesis, which is followed by Text to speech (TTS) synthesis. She discussed about pronunciation of vowels and consonants and gave a detailed explanation of Text to speech synthesis functional diagram, which is enhanced by a sample example video. TTS is categorised to 3 broad classes: Concatenate synthesis, Formant Synthesis and Articulatory Synthesis.

She explained the history and current status of speech synthesis with the help of a video which is followed by familiarizing festival and swaram applications. The class dispersed for tea break at 3:45 pm and renewed at 4:10 pm. The topic of discussion was speech synthesis and recognition and various challenges in the process. Session then moved on to discussing identification of phonemes and syllables. The last topic of discussion was various TTS systems. Such as phoneme based, diphone based , unit selection based tts systems. The session ended at 4.50 pm.

The session discussed about speech synthesis, another research area related to the language processing. The students got idea about recognizing and processing speech with the help of machines and how it can be used for the society and mankind.

Day 6:21-01-2019

Sixth day of winter school started with the talk of Dr. Gopakumar, Assistant professor at NIT Calicut. The topic of discussion was neural networks. The session started with explaining the concept of neural networks in analogy to the neurons in the human brain. Which is followed by describing Artificial Neural Networks(ANN) , its formation and functionalities. The topic covered brain's characteristics along with neural networks such as massive parallelism, learning ability, adaptivity, generalization ability, fault tolerance, low energy consumption etc. Then the topic moved on to explain the concept of pattern recognition which is an application of neural networks. This included mathematical model of neuron, explaining neural network structure etc. which is followed by introducing various activation functions like, logistic function, sigmoid function, linear, non-linear etc. Next topic of discussion was confusion matrix which explained concept true positive, true negative and so on. Then he explained recurrent networks, feed forward networks which are usually arranged in layers, such that each unit receives input only from units in the immediately preceding layers, Single Layer network – Every unit connects directly from the network's inputs to its outputs. Then explained multilayer neural networks and single layer feed forward networks. He enhanced the concept of perceptron using a mathematical model of 2 bit adder then mentioned the concept of majority function. The session dispersed to tea break at 11:30 am and renewed at 11:50 am. It started with the concept of multi layer feedforward networks which covered

Gradient-descent loss-minimization method, Nonlinear Regression, learning in multilayer networks, back propagation networks etc. Next topic of discussion was learning in neural networks which included over fitting , cross validation, further explained optimal brain damage algorithm and tiling algorithm. Lunch break was at 1:00 pm and the session renewed at 2:pm.

Session 2 - Afternoon :

Afternoon session was about the case study of neural networks including practical sessions. It covered deep learning , RNN and CNN. RNN is mainly used for training series prediction and time series anomaly detection. CNN is mainly for image classification. Tea break was at 3.30 pm. After the break the session continued by explaining diagram of machine learning and deep learning. Three of the most common layers are: Convolution, Activation or ReLu and Pooling. Convolution puts the input images through a set of convolution filters, each of which activates certain features from the images. Installed tensorflow along with anaconda navigator to show more practical applications of neural networks. Which includes downloading data and training using networks. The session completed at 5:00 pm.

The relatively new and interesting topic of neural networks were discussed and participants gained more knowledge through practical sessions. The session was very informative and useful as participants could experiment what they learned.

Day 7 : 22-01-2019

Session 1 - Forenoon :

7th day of Women winter school began at 10:00 am. The forenoon session was delivered by Dr. Elizabeth Sherly, Professor at IIITMK Trivandrum and the topic of discussion was Speech recognition. She started with importance in speech recognition in real life problem such as uber eats, etc and mentioned that speech recognition of continuous sentences are more complex than that of simple sentences, which was followed by discussing the history of ASR(Automatic Speech Recognition). She stated that Voice Recognition in computers is not a recent tech, it has a history of way back from 1952 and mentioned early speech recognition systems such as Shoe Box , Radio Rex etc. She mentioned the revolution in speech recognition in the last decade with the example of google voice search. She discussed neural network, HMM based ANN for learning and training. Next topic of discussion was brain computing, and she explained how the neuro computing is done in analogy to the human brain. She further discussed the challenges in ASR such as different speakers having different accents, slags, pronunciation, dialects, high viability of speech signals etc. Then she compared how humans and computers recognize speech by explaining words like digitization, quantization, sampling. The session dispersed for a tea break at 11:10 am and renewed at 11:25 am.

Session started with giving a detailed description of ASR architecture. This can be categorized in to many models, such as rule based systems which again can be categorized into two according to the words and phonemes it identifies, Hidden markov model systems, Statistical systems etc. She gave detailed description of HMM in speech processing, then discussed gaussian mixtures and different language models such as unigram, bigram and n-gram models. She discussed how to incorporate neural networks to speech recognition, explained different types of neural networks such as CNN, deep belief networks, recursive neural networks etc. Then introduced supervised and unsupervised training in deep neural networks.

Topic then moved onto various speech recognition applications which was presented by Ms.Lekshmi , Research scholar, IIITMK. She started with explaining how to save wave files with labels and noise profile in speech samples. She introduced praat a free scientific software for phonetic analysis and explained praat with screenshots. Then explained another software called jAudio, a complete java based feature extraction tool. Then discussed about Hidden Markov Model ToolKit (HTK) and gave a overview of how it works and how to install it. Next topic of discuss was Building ASR using Kaldi. Kaldi is a Toolkit for Speech recognition with Apache Licence. Gave a detailed description about its working principle, also gave

information about how to download kaldi from Github. The session came to an end at 1.45 pm and dispersed for lunch.

Automatic speech recognition was a relatively new topic for the whole team of participants. They were very enthusiastic to know about the technology as it is has various applications in the current research and development environment.

Session 2 - Afternoon:

Afternoon session started at 2:30 pm , and the session was handled by Dr. Shoba Lalitha Devi. She talked about Named entity extraction. It started with Natural Language and Discourse. Then moved onto relation identification and Information extraction .She stated that Entity Identification and Entity Extraction is a task that locate and classify atomic elements in text into predefined categories such as name of a person, Organization, locations, expression of times, quantities, monetary values, percentages etc. She explained the terms with examples. Next she discussed about problems in named entity extraction , ambiguity. Classified named entities into noun phrase, rigid designator and co-referential. She explained various ambiguities such person vs location and person vs organization. Session dispersed to a tea break at 3.45 pm and renewed at 4.00pm. She discussed different approaches to NER. first one was dictionary based approach. She noted some advantages and disadvantages regarding dictionary lookup. Introduced new concept of maximum entropy theory. Then discussed about lexicon based approach. Which is followed by Deep explanation of NER using HMM and Maximum Entropy theory methods. The session came to an end at 5.00 pm.

After the session there was a photo session including all the winter school participants and icfoss members at the icfoss office in Thejaswini building.

NER, one of the challenging task for a linguist researching in any language discussed in detailed. The participants got to know about struggles and practical issues faced by a researcher in identifying extracting named entities while processing language.

Day 8 : 23-01-2019

Session 1: Forenoon

Eighth day of winter school started with the presentation of Dr. Rajeev R R, who is also the chair of women winter school, regarding the subject of Automatic Malayalam - Tamil Machine Translation System, which enable translation of any text in Malayalam to tamil and vice versa. Then he explained the entire architecture of translation system. He talked about machine translations in Indian languages, stated that Indian languages are highly inflectional, with a rich morphology, relatively free word order, and default sentence structure subject-object-verb (SOV) and majority of works are carried out in rule based method. Then he stated that both Malayalam and Tamil are highly agglutinative languages and had a comparison between both languages. He explained code mixing using examples such as English-Malayalam mixed Sentence and Social Media chat or message languages. Then he discussed about the difficulties in Malayalam tokenization and difference between translation and transliteration. Which was followed by explaining some basic NLP tasks such as morphological analysis, pos tag, chunking, pruning which is the Task of removing additional abbreviation feature structure. The session dispersed for tea break at 11.10 am.

Session after tea break was handled by Dr Vishal Goyal, professor at punjabi university, through skype. The topic of discussion was Natural Language Processing – Machine Translation. He talked about NLP and NLU(Natural Language understanding), need for NLP, and applications of NLP. He gave detailed explanation about spell checker and grammar checker, speech processing. Synthesis, information extraction, machine translation, question answering system etc. He stated that ambiguity is the most difficult problem in processing natural languages. Then he discussed Gurmukhi to Shahmukhi transliteration, sentiment analysis: classifying sentiments to positive, negative and neutral etc. which was followed by different methods of machine translation such as rule based, statistical based, hybrid based etc. He introduced various projects that are going on in nlp such as corpus analysis, Extraction of proverbs from hindi sentence, word sentence disambiguation, automatic sentence alignment in parallel corpus, automatic word alignment in parallel corpus, font converter etc. which was followed by discussing Hindi to Punjabi Machine Translation System and various challenges in that project. Then there was demonstration of works he had done. The session dispersed for lunch break.

The session gave insight about how the translation problem from one language to another can be easily tackled. As translation is an important task which help to communicate the people over the world, the session was so useful for some of the researchers among the participants who works in the language processing area.

Afternoon session was taken by Mr. Jitesh Pubreja, he gave detailed introduction NLTK tool kit. Gave instructions about how to download NLTK, and made participants to download NLTK. He introduced stemmers in stemmer in NLTK, such as Porter, most commonly used stemmer, Lancaster, Snowball etc. then introduced lemmatizer, WordNet Noun Lemmatizer and WordNet Verb Lemmatizer. Then there was a practical session which demonstrated stemming, word tokenization, sentence tokenization, lemmatization, pos tagging etc.

Next topic of discussion was Plagiarism detection tool for Indian languages with special focus on Punjabi & Hindi Documents. He stated that plagiarism is the unauthorised use or imitation or reproduction of existing work. Plagiarism is a serious problem for researchers. Lots of plagiarism detector available in English Language, but not good one for Indian Languages. He talked about existing plagiarism detection tools such as Eve2. The session came to an end at 5:00 pm.

Participants got introduced to various packages in Python-NLTK library and how it can be utilized for processing language. The session got some practical experiments with the packages.

Day 9: 24-01-2019

The day started with the presentation of Dr. Niladri Shekhar Dash at 10:00 am, and the topic of discussion was digital dictionaries. He started with his idea about the linguistics, development of dictionaries in digital platform with various technologies etc, explained lexicography and various methods. He mentioned several dictionaries and sister domains such as wikipedia.org, dictionary.com. Which was followed by explaining criterias for dictionary classification such as density of entries, nature of inclusion, number of languages, nature of entries etc. He discussed different types of dictionaries: historical, academic, descriptive, monolingual, bilingual etc,. Then he mentioned some special dictionaries with specific understanding. That are dialect dictionaries, dictionary of technical terms, Pronunciation Dictionaries, Grammatical Dictionaries, Dictionary of collocations, dictionary of borrowed words. He described how to collect data to populate dictionaries, and input to the various dictionaries mentioned above, and the type of entries dictionaries can have: single word entries and multi word entries. Then he discussed about the organization of words in dictionary. Session dispersed for lunch break at 1.00 pm.

Session 2: Afternoon

Session renewed at 2.00 pm after lunch break. Topic of discussion was digital dictionaries. He introduced various forms of digital dictionaries: E- Dictionary , Electronic Dictionary,

Multimedia, Multi- Model Dictionary, Online Dictionary etc. He pointed out uses of digital dictionaries, such as better definition ,better illustration, appropriate quotation, audio- video animation facility, better Linguistics, contextualized linguistics information etc. Then discussed about various forms of digital dictionary such as Lexipedia, Flexipedia, Digital Lexical profile, wordNet, Machine Readable Dictionary, Digital Lexical Database, TermBank. Topic then moved on to explain characteristics of modern DD's. There was a Malayalam typing session using swanalekha writing tool after break. The session wrapped up at 5.00 pm.

The participants were enthusiastic to know about how dictionaries used in language processing. As dictionaries plays an important role in processing natural languages. The participants got to know about various types and construction of dictionaries.

Day 10: 25-01-2019

Session 1: Forenoon

Final day of women winter school was mainly intended for presenting new project proposals from participants, and evaluation of the same. The evaluation committee comprised of Dr Rajeev R R, Mr. Jayadev, Mr. Sampath Kumar, Programming Heads of ICFOSS. Participants were divided to 7 groups, each containing at least 4 members.

The below table trace the Groups and the proposal presented

Group	Area	Proposed Project	Members
Group 1	Cyber Security	Malicious URL Detection using NLP & Machine Learning	Remya Odan Valappil Anaya Anson Anu Unnikrishnan Deepa Vijay
Group 2	Medicine		Chithra Raj N Anjali M K Surya U Mini U
Group 3	Forensics	Auto Forensic Text Analyzer	Divya Saju Junaida M K Reshma Prasad Soumya Jacob
Group 4	Social Media Analytics	About Kerala -A Twitter Sentiment Analysis	Dr. A Thamizharasi Manju G. Diliya M Khan Vidya S Dandagi

Group 5	Disaster Manangement	FLOOD PREDICTION AND RESCUE	Renu S Nishy Reshmi S Nadeekha P V Veena Vijayan
Group 6	Education	Edu-Assistance	Athira B U Ajusha P V Merlin George Shagi G U
Group 7	Assistive Technology	Third Hand	Sangeetha Jamal Neetha Thomas Hrudya K P Sythalakshmi C.S

Each Team was allowed a maximum of 20 minutes for the presentation, followed by a Q & A session.

Session dispersed for lunch break at 1:45 pm and renewed at 2.50pm.

Valedictory

Afternoon session began with addressing note of Dr S Chitra IAS, Director of Kerala IT Mission. She talked about the role of women in different regions of society and the problems they face in workplaces, home etc., which was followed by the talk of Mrs. Mridul Eapen , Member of Planning Board.

She stated that women empowerment should start from family itself, that family should treat both the girl child and boy child equally, she also mentioned that platforms like winter school provide great opportunities to women to get exposed to technical world.

They presented momento to the participants for the best project proposal done during the final day of the event. There was a feedback session after the talk. Some of the participants shared their experience in the winter school. The session closed by 5.00 pm.

Feedback

The event was a success. The feedback we received was overwhelmingly positive, with all the participants commenting favourably on the experience.
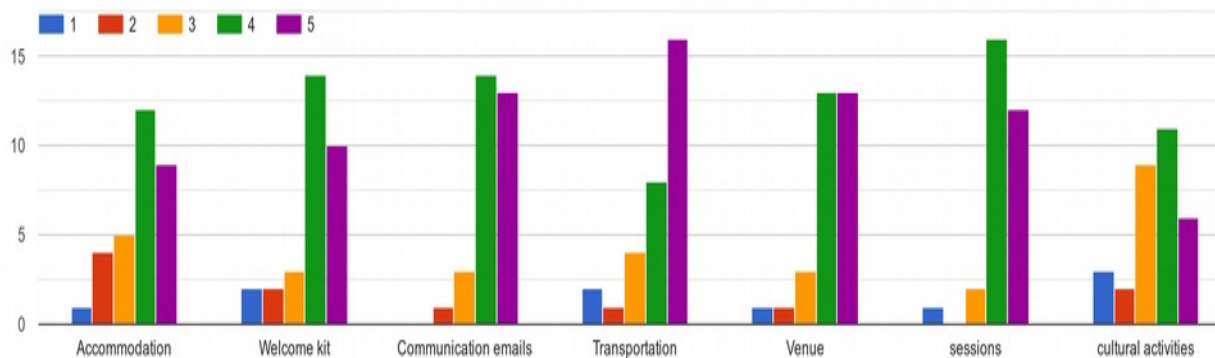
The rating system for the questionaaire prepared for the evaluation of the overall programme was as follows:

1-Poor , 2-Fair, 3-Average, 4-Good, 5-Excellent

The relevant charts from the analysis of the feedback received from the participants are provided below.
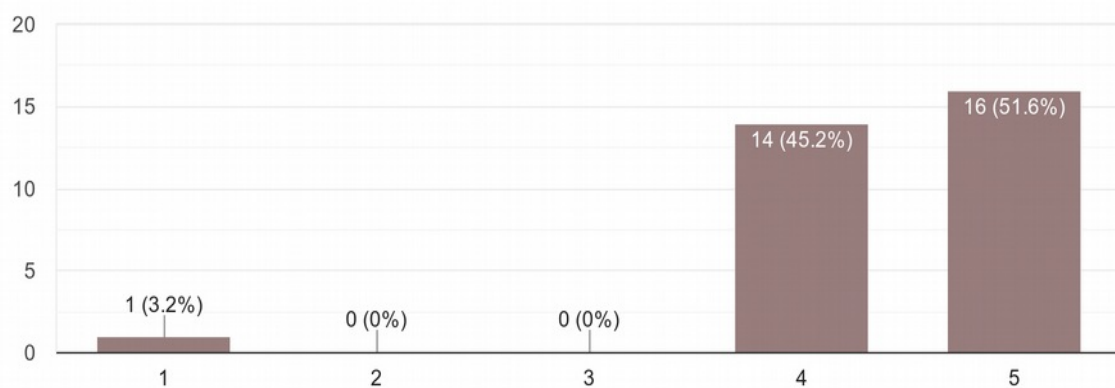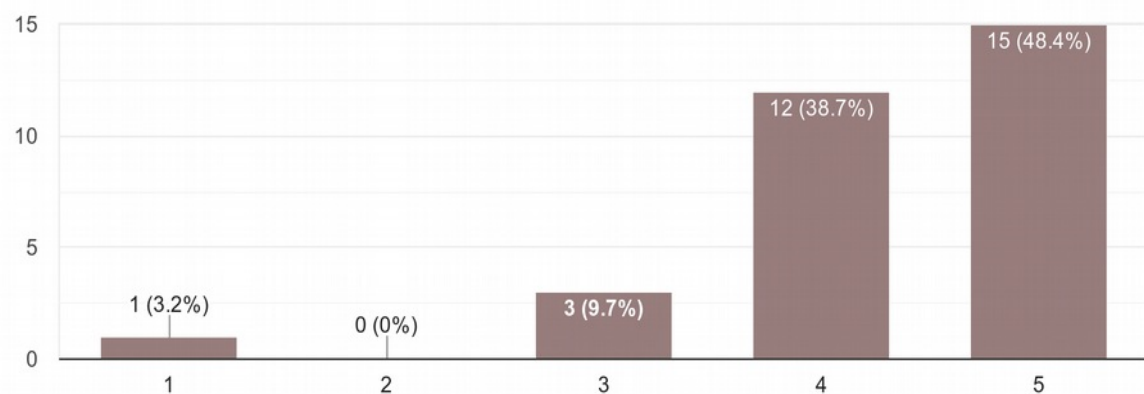
How satisfied were you with the logistics?

How satisfied were you with the event?

31 responses

How relevant and helpful do you think it was for your career?

31 responses

Event Outcome

The event involved the women community focussing on the academic and research development in the area of language processing. The participants were keen on presenting the project proposals and they positively took the suggestion to keep in contact with each other to develop the projects they came up with.