ASHA HEGDE

TECHNICAL CONSULTANT
Mangalore University

hegdekasha@gmail

+91 94811 64881

4-41/5, Green Valley, Konaje

SKILL

Problem Solving

Adaptability

Collaboration

Strong Work Ethic

Time Management

Critical Thinking

Handling Pressure

Leadership

EDUCATION

DEGREE NAME / MAJOR

Pursuing **Ph.D.** in the Department of Post-Graduate Studies and Research in Computer Science, Mangalore University

M.Sc. With distinction (82%) in Computer Science
Mangalore University
2011, August

B.Sc.with distinction (80%)
(Mathematics, Electronics, Computer Science)
Karnataka University Dharwad
2009, September

RESUME

AREA OF INTEREST

My interest of research is in computational linguistics and Natural Language Processing. I am interested in developing resources and tools for under-resources languages. Most of my work involve with focus on building dedicated models for under-resourced languages. Currently, I am pursuing **Ph.D.**under the supervision of Prof. H.L. Shashirekha and my thesis title is **"Machine Translation of Under-Resourced Languages: A Case Study Translation from Kannada to Sanskrit and Tulu"**. The motivation behind this research work is to understand the impact of shared vocabulary in machine translation of morphologically rich languages.

EXPERIENCE

Technical Consultant|Mangalore University
January 2022 - Present

- Installed operating system
- Installed system updates to address vulnerabilities and reduce security issues.
- Enhanced interfaces to promote better functionality for users.
- Troubleshot systems comprised of security alarms and Internet connectivity.
- Investigated system issues and implemented resolutions to reduce downtime.
- Enhanced and reconstructed computer systems to increase efficiency

Worked as Lecturer - Mangalore University|Mangalore
August 2018 - May 2019

- Led projects and analyzed data to identify opportunities for improvement.
- Created plans and communicated deadlines to complete projects on time.
- Participated in continuous improvement by generating suggestions, engaging in problem-solving activities to support teamwork.
- Identified issues, analyzed information and provided solutions to problems.
- Conducted research, gathered information from multiple sources and presented results.

Lecturer for B.Com, GFGC, Sirsi|First Grade College, Sirsi
June 2012 - April 2013

Lecturer and Project Guide for BCA |JMJ College, Sirsi
June 2012 - April 2013

Lecturer and Project Guide for BCA |HiraWomenscollege
June 2011 - April 2012

LANGUAGE SKILLS

English – Professional
Hindi – Professional
Kannada – Native
Tulu – Professional
Tamil – Basic
Telugu – Professional
Malayalam – Professional

WORKED AS VOLUNTEER

- Shared Task on Machine Translation in Dravidian Languages – 2022
- EACL - 2021

Github page:
https://github.com/hegdekasha

WORKSHOP/WEBINARS

- "Multilingual and Code-Switching ASR Challenges for Low Resource Indian Languages(MUCS) 2021"
- 'Practical Applications of Sanskrit' **(participated as speaker)**
- PFA Certificate of presenting a paper in **ICON 2020**.

References

H.L. Shashirekha
Professor
Department of Computer Science
Mangalore University
India
hlsrekha@gmail.com

Bharathi Raja Asoka Chakravarthi
Adjunct Lecturer at School of Computer Science, National University of Ireland, Galway
bharathi.raja@insight-centre.org

Anand Kumar M
Assistant Professor, IT
NITK-Suratkal, India
m_anandinfo@gmail.com

RESEARCH EXPERIENCE

Mangalore University **2019 - Present**

- Worked as co-advisor for the post graduate projects which were dedicatedly on developing resources for under-resourced languages.
- The work focus was on developing resources for Kannada and Tulu.
- Participated in many shared tasks organized by different organizers. The research focus was on experimentation oftrending technologies for under-resourced languages. By that investigating effective and dedicated models for under-resourced languages.
- Worked as reviewer for many workshops and conferences.

TECHNICAL SKILLS

- Several experiments are conducted in different areas of NLP, namely sequence labelling tasks (POS tagging and NER), Text classification at sentence level (sentiment analysis, Hate speech and Offensive content identification), and Text classification at Document level (Fake news detection).
- Further, exhaustive experiments are conducted on machine translation specifically for under-resourced languages.

PUBLICATIONS

- **Papers Published:**
  - MUCS@ - Machine Translation for Dravidian Languages using Stacked Long Short Term Memory (2020), In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*.
  - MUCS@Adap-MT 2020: Low Resource Domain Adaptation for Indic Machine Translation, In *Proceedings of the 17th International Conference on Natural Language Processing (ICON): Adap-MT 2020 Shared Task.*
  - MUM at ComMA@ICON: Multilingual Gender Biased and Communal Language Identification using Supervised Learning Approaches (2021), In *Proceedings of the 18th International Conference on Natural Language Processing: Shared Task on Multilingual Gender Biased and Communal Language Identification.*
- **Papers Accepted:**
  - Corpus Creation for Sentiment Analysis in Code-Mixed Tulu Text (SIGUL at LREC-2022)
  - Ensemble Based Machine Learning Models for Hate Speech and Offensive Content Identification, (working notes in FIRE 2021)
  - Urdu Fake News Detection Using Ensemble of Machine Learning Models, (working notes in FIRE 2021)
  - Overview of the Shared Task on Machine Translation in Dravidian Languages, (Dravidianlangtech-2022 shared task in ACL-2022)
  - MUCS@Text-LT-EDI@ACL 2022: Detecting Sign of Depression from Social Media Text using Supervised Learning Approach (LT-EDI workshop at ACL -2022)
  - MUCS@DravidianLangTech@ACL2022: Ensemble of Logistic Regression Penalties to Identify Emotions in Tamil Text. (Dravidianlangtech-2022 workshop at ACL-2022)

Brief summary:

Greetings!
I am Asha Hegde. I'm currently pursuing my Ph.D. in Natural language Processing (NLP) at Mangalore University, India. My thesis title is **"Machine Translation of Under-Resourced Languages: A Case Study Translation from Kannada to Sanskrit and Tulu"**. The motivations behind selecting this topic are:

- Exploring the impact of shared vocabulary on Machine Translation (MT) using under-resourced morphologically rich agglutinative languages

- Developing e-resources for Kannada and Tulu

I have made a detailed study on machine translation and impact of under-resourceness, morphological richness, and agglutinativeness on machine translation. My research focus is on developing effective pre-processing and postprocessing techniques which include language dependent techniques to language independent techniques dedicatedly for under-resourced languages like Kannada, Sanskrit and Tulu. As part of my research, several experiments are being carried out to identify challenges in handling under-resourced morphologically rich languages for translation. By developing baseline translation models, these challenges were empirically studied and research is in progress to tackle these issues. During these experiments, I have figured out the challenges with translation of under-resourced languages and as a resolution I started my research from scratch ie., by developing parallel corpus for Kannada-Tulu and Kannada-Sanskrit as there is a lack of parallel corpora for these language pairs when this research title was finalized. The parallel corpus construction includes two components, namely, manual corpus construction by consulting experts and synthetic data generation using grammar. Then the newly constructed parallel corpus is used to construct MT models by considering trending technologies in MT. My research work also involved developing e-resources like POS tagger (for Tulu only) and word segmentation, different word representations for Kannada and Tulu, etc.

In the domain of NLP, I have participated in many shared tasks and a brief summary of few of them are given below:

- The models are submitted to a shared task on "Machine Translation in Dravidian languages", at EACL 2021. In this work, I used the seq-2-seq model with stacked LSTM to conduct MT [1].
- A model is submitted to a shared task on "AdapMT 2020: Low Resource Domain Adaptation for Indic Machine Translation". The model was built based on extracting the sentences from a large corpus for domain adaptation [2].
- I have worked on milticlass classification task and the models are submitted to a shared task on Multilingual Gender Biased and Communal Language Identification at ICON 2021. I used pretrained transformer embeddings ie., mBERT [3].
- I have worked on hate speech detection in Indo-Aryan languages using ensemble of ML algorithms. The models were submitted on a shared task HASOC-2021@FIRE-2021 [4].

Also, I have worked on corpus creation of Tulu-Kannada-English code-mixed text for sentiment analysis and the paper is under communication with SIGUL workshop at LREC-2022. This work aims at developing an automatic sentiment analysis model for code-mixed Tulu-Kannada-English which is the first work that deals with Tulu-Kannada-English a trilingual dataset for sentiment analysis. The model submitted can be found at https://github.com/hegdekasha/Tulu-sentiment-analysis and the dataset can be shared based on the request. In this paper, I have scrapped YouTube comments and each comment is manually tagged by native speakers. Then inter annotator agreement is also used to check the agreement between annotators. Baseline ML classifiers are implemented to set up benchmarked results for sentiment analysis using Tulu-Kannada-English code-mixed text.

With regards to the professional experience, I am also working as a technical assistant for the past few months and handling a few projects. Currently, I am working at Computer Centre, Mangalore University, where my job involved to write clean and maintainable code for web applications. I worked on PHP based front end and mysql as the backend. My future goal is to conduct research in the areas of Natural language processing, specifically in the domain of multimodal NLP using under-resourced languages.

Lastly, I would like to say that I believe it is my calm demeanor, structured manner of thinking, patience and determination that have helped me in carrying out my research successfully. I have developed the ability to identify the root problems in a research statement and come up with new and efficient solutions to it, and given the opportunity to pursue an internship/collaborative research would apply my skills to carry out meaningful and impactful research.

References:

[1] Asha Hegde and H.l. Shashirekha. 2020. MUCS@Adap-MT 2020: Low Resource Domain Adaptation for Indic Machine Translation. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON): Adap-MT 2020 Shared Task*, pages 24–28, Patna, India. NLP Association of India (NLPAI).

[2] Asha Hegde, Ibrahim Gashaw, and Shashirekha H.l.. 2021. MUCS@ - Machine Translation for Dravidian Languages using Stacked Long Short Term Memory. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 340–345, Kyiv. Association for Computational Linguistics.

[3] Hegde, Asha, Mudoor Devadas Anusha, Sharal Coelho, and Hosahalli Lakshmaiah Shashirekha. "MUM at ComMA@ ICON: Multilingual Gender Biased and Communal Language Identification using Supervised Learning Approaches." In *Proceedings of the 18th International Conference on Natural Language Processing: Shared Task on Multilingual Gender Biased and Communal Language Identification*, pp. 64-69. 2021.

[4] Mandl, Thomas, Sandip Modha, Gautam Kishore Shahi, Hiren Madhu, Shrey Satapara, Prasenjit Majumder, Johannes Schäfer et al. "Overview of the HASOC subtrack at FIRE 2021: Hate speech and offensive content identification in English and Indo-Aryan languages." *arXiv preprint arXiv:2112.09301* (2021).

[5] Amjad, Maaz, Grigori Sidorov, Alisa Zhila, Alexander Gelbukh, and Paolo Rosso. "UrduFake@ FIRE2021: Shared Track on Fake News Identification in Urdu." In *Forum for Information Retrieval Evaluation*, pp. 19-21. 2021. https://doi.org/10.1145/3503162.3505240