

Curriculum Vitae

Dr. Sobha, L (Lalitha Devi).

Present Occupation: (2002 July to Present)

Program Director
AU-KBC Research Centre

Member Research Staff
Computational Linguistics Research Group (CLRG)
AU-KBC Research Centre, MIT Campus,
Anna University, Chromepet, Chennai-600044.
www.au-kbc.org

Scientist
KBC Research Foundation Pvt Ltd. www.kbcrf.co.in

(AU-KBC Research Centre is a Public Private Partnership (PPP) Research Lab where KBC Research Foundation as private entity and Anna University as the public entity)

Email sobha@au-kbc.org sobhanair@yahoo.com

Researcher and product developer in Natural Language Understanding and Processing

I work in the field of Computational Linguistics at theoretical and application level. At the theoretical level the area of research is in **Discourse Semantics (Anaphora, Connectives and other discourse markers, conversation analysis)**. The application level work is in the field of **Information Retrieval & Extraction** and basic pre-processing tools for Indic and Indo-European language families. More on the work of my group and me and our Publications are available at <http://au-kbc.org/nlp/index.html>

Professional / Occupational Information

- 2018 onwards **Program Director** , AU-KBC Research Centre
- 2002 July onwards working as **Scientist at KBCRF** and **Member Research Staff at AU-KBC Research Centre** , M.I.T. Campus of Anna University, Chrompet, Chennai.
- **Erasmus+ Teaching Fellow, December 2015- September 2017**, at Department of Information and Communication Sciences, **Artificial Intelligence (AI) Lab, University of the Aegean , Greece.**
- **Erasmus Mundus International scholar 2010-2012**, Taught Masters Programme 'International Masters in NLP & HLT'. The course is coordinated by the Université de Franche-Comté, France and taught as a joint activity with: Universidade do Algarve, Portugal, Universitat Autònoma de Barcelona, Spain, and **University of Wolverhampton, United Kingdom.** From **October 2010-December 2010**
- **Visiting Scientist, January 2011-March 2011** at the Dpto. De Sistemas Informaticos y Computacion, **Universidad Politecnica De Valencia, Valencia, Spain,**
- **Visiting Fellow, September 2010**, Computational Linguistics Group, **University of Wolverhampton, Wolverhampton, UK.**

- **Guest Researcher** (Postdoctoral Research) **2000**: Computational Linguistics Lab, **University of Wolverhampton**, United Kingdom with **Prof. Ruslan Mitkov**,
- **Research Scholar: December 1994-September 1999**. Thesis Title: **Anaphora Resolution in Malayalam and Hindi**, Supervisor: **Prof. B. N Patnaik, IIT Kanpur**, Kanpur, India.
- **Project Associate: July 1994- March 1995** A Computational Lexicon for **Juang**, a tribal language of India spoken in Orissa, **Prof. Achla Misri Raina, IIT Kanpur**, Kanpur, India
- **Project Associate: July 1991- June 1994**, *Foundations in Indian Linguistics and Logic*. With Prof **B.N. Patnaik and Prof. Rajeev Sangal** at Indian Institute of Technology Kanpur (IITK), Kanpur, India..
- **Project Assistant: January-April, 1991**, *Cross-cultural Studies: Anaphora Acquisition in Malayalam*. With **Prof. Barbara C Lust, Cornell University**, Ithaca, New York, USA ,

Education

Ph. D. Computational Linguistics	IIT Kanpur/MG University
M. Phil Computational Linguistics	Madurai Kamaraj University
M. A. Applied Linguistics	Kerala University
M.Sc Computer Science	Annamalai University
B.Sc. Mathematics	Kerala University

Select Journal Publications

Published

1. Vijay Sundar Ram and **Sobha Lalitha Devi. (2023)**. "Hindi to Dravidian Language Neural Machine Translation System". In Proceedings of Recent Trends in Natural Language Processing (RANLP), 2023.
2. Pattabhi RK Rao and **Sobha Lalitha Devi. (2023)**. "'ChemXtract' A System for Extraction of Chemical Events from Patent Documents' ". In Proceedings of Recent Trends in Natural Language Processing (RANLP), 2023
3. **Sobha Lalitha Devi**, Pattabhi R K Rao and Vijay Sundar Ram. **(2023)**. Workbench for Post-editing of Translations from English to Dravidian Languages. In: Proceedings of International Conference Human-informed Translation and Interpreting Technology (HiT-IT 2023), ISSN 2683-0078. pp. 315-324
4. Pattabhi RK Rao and **Sobha Lalitha Devi. (2023)**. Event Extraction From Social Media Text In Malayalam Using Neural Conditional Random Fields. In: Proceedings of FOSS Approaches towards Computational Intelligence and Language Technology (FOSS-CILT 23)
5. Vijay Sundar Ram R and **Sobha Lalitha Devi. (2023)**. Neural Machine Translation For English -Malayalam. In: Proceedings of FOSS Approaches towards Computational Intelligence and Language Technology (FOSS-CILT 23). (Awarded as Best Paper)
6. Treesa Cyriac and **Sobha Lalitha Devi, (2022)**. "Classification of Multiword Expressions in Malayalam", In. Proceedings of The WILDRE-6 Workshop within the 13th Language Resources and Evaluation Conference, France, pp 55–59

7. Sheeja S, Kumari and **Sobha Lalitha Devi**, (2022). “Automatic Identification of Explicit Connectives in Malayalam”, In. Proceedings of The WILDRE-6 Workshop within the 13th Language Resources and Evaluation Conference, France, pp 74–79
8. Nithya. J., Vijay Sundar Ram, R., and **Sobha Lalitha Devi**. (2021) Translation of Verb chunks using ML Techniques . In Tamil Computing - Tools and Applications Young Researchers' Conference 2021 (TaCTA-YRC2021)
9. C.S.Malarkodi and **Sobha Lalitha Devi**. (2021) Identification of Names from Tamil Novel . In Tamil Computing - Tools and Applications Young Researchers' Conference 2021 (TaCTA-YRC2021)
10. Sindhuja Gopalan and **Sobha Lalitha Devi**. (2021) 'Role of Concession Connectives in Sentiment Analysis'. In Tamil Computing - Tools and Applications Young Researchers' Conference 2021 (TaCTA-YRC2021) [Best Paper Award]
11. Kumari Sheeja Saraswathy & **Sobha Lalitha Devi** (2021), ‘Enhancement of sentiment analysis using clause and discourse connectives’, Computers, Materials & Continua, ISSN: 1546-2218. vol. 68, no. 2, pp. 1983-1999
12. C.S.Malarkodi, Pattabhi RK Rao and **Sobha Lalitha Devi**. (2020) “CLRG ChemNER: A Chemical Named Entity Recognizer @ ChEMU CLEF 2020”. In ChEMU: Cheminformatics Elsevier Melbourne University, Conference and Labs of the Evaluation Forum 2020 (CLEF 2020).
13. Jiayuan He, Dat Quoc Nguyen,Malarkodi C.S., Pattabhi RK Rao, **Sobha Lalitha Devi**,Trevor Cohn, Timothy Baldwin and Karin Verspoor (2020) “An Extended Overview of ChEMU 2020: Named Entity Recognition and Event Extraction of Chemical Reactions from Patents”. In: Arampatzis A. et al. (eds) Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2020. Lecture Notes in Computer Science, vol 12260. Springer, Cham. https://doi.org/10.1007/978-3-030-58219-7_18
14. **Sobha Lalitha Devi** (2020). "SocAnaRes-IL20: Anaphora Resolution from Social Media Text in Indian Languages @ FIRE 2020 - An Overview”, In the Forum for Information Retrieval and Evaluation-2020, IDRBT, Hyderabad, India.
15. Vijay Sundar Ram and **Sobha Lalitha Devi** (2020). “Handling Noun-Noun Coreference in Tamil”, In 5th Workshop on Indian Language Data: Resources and Evaluation (WILDRE), Organized under LREC-2020, Marseille, France.
16. Malarkodi C S and **Sobha Lalitha Devi**. (2020), “A Deeper Study on Features for Named Entity Recognition”, In 5th Workshop on Indian Language Data: Resources and Evaluation (WILDRE), Organized under LREC-2020, Marseille, France
17. C.S.Malarkodi and **Sobha Lalitha Devi**. (2019) "Generic Feature Selection Methodology to Named Entity Detection from Indian and European Languages", Advances in Electrical and Computer Engineering vol.19, no.1, pp.79-88, 2019, doi:10.4316/AECE.2019.01011
18. Pattabhi RK Rao and **Sobha Lalitha Devi**, (2018), "Enhancing Multi-Document Summarization using Concepts", Sādhana - Academy Proceedings in Engineering Sciences. vol 43:2 (27) . <https://doi.org/10.1007/s12046-018-0789-y>
19. Sindhuja Gopalan, and **Sobha Lalitha Devi**. (2018). "Cause and Effect Extraction from Biomedical Corpus", *Computacion y Sistemas vol 21(4)*
20. Vijay Sundar Ram and **Sobha Lalitha Devi**. (2017). "A Robust Coreference chain Builder for Tamil", In 18th International Conference on Computational Linguistics and Intelligent Text Processing, LNCS, Springer
21. Sindhuja Gopalan, and **Sobha Lalitha Devi**. (2017). “Cause and Effect Extraction from Biomedical Corpus”, In 18th International Conference on Computational Linguistics and Intelligent Text Processing, LNCS, Springer

22. Sindhuja Gopalan and **Sobha Lalitha Devi**. (2017). 'BNEMiner: mining biomedical literature for extraction of biological target, disease and chemical entities'. *International Journal of Business Intelligence, Data Mining*. 11(2), pp. 190-204. DOI: <https://doi.org/10.1504/IJBIDM.2016.081612>
23. Pattabhi RK Rao and **Sobha Lalitha Devi**, (2017), 'Patent Document Summarization Using Conceptual Graphs', *International Journal on Natural Language Computing (IJNLC)*, Vol. 6(3).
24. **Sobha Lalitha Devi** and Pattabhi RK Rao.(2016). "Mining of Social Networks from Literary Texts of Resource Poor Languages", In *LNCS, Springer, 17th International Conference on Intelligent Text Processing and Computational Linguistics*, Konya, Turkey.
25. Sindhuja Gopalan, Paolo Rosso and **Sobha Lalitha Devi**. (2016). "A Marker for Identifying Featured Articles in Biological Wikipedia", *In the Special Issue of Journal Research in computing Science (RCS)*.
26. Malarkodi C.S., Elisabeth Lex and **Sobha Lalitha Devi**.(2016). "Named Entity Recognition for the Agricultural Domain", *In the Special Issue of Journal Research in computing Science (RCS)*.
27. Vijay Sundar Ram and **Sobha Lalitha Devi**, (2016). "Two Layer Machine Learning Approach for Mining Referential Entities for a Morphologically Rich Language". *In Asian Journal of Information Technology*, Vol.15. pp.2831-2838.
28. Pattabhi R K Rao and **Sobha Lalitha Devi**. (2015) "Automatic Identification of Conceptual Structures Using Deep Boltzmann Machines" In Proceedings of Forum for Information Retrieval and Evaluation-2015, ACM DL, ISBN 978-1-4503-4004-5
29. Akilandeswari A. and **Sobha Lalitha Devi**. (2015), "Tamil Pronominal Resolution Boosted By Sentence Transformation" in **Australian Journal of Basic and Applied Sciences**, 9(23) July 2015,ISSN:1991-8178, Pages: 566-572.
30. **Sobha Lalitha Devi**, Vijay Sundar Ram, Pattabhi RK Rao. (2014). "A Generic Anaphora Resolution Engine for Indian Languages". In Proceedings of the 25th International Conference on Computational Linguistics (**COLING 2014**).
31. **Sobha Lalitha Devi**, Lakshmi S and Sindhuja Gopalan(2014). "Discourse Tagging for Indian Languages", In A. Gelbukh (ed), *Computational Linguistics and Intelligent Text Processing*, Springer **LNCS** Vol 8403, pp. 469-480.
32. Vijay Sundar Ram, Efstathios Stamatatoa, **Sobha Lalitha Devi** . (2014). "Identification of Plagiarism using Syntactic and Semantic Filters", In A. Gelbukh (ed), *Computational Linguistics and Intelligent Text Processing*, Springer **LNCS** 8404, pp. 495–506
33. Vijay Sundar Ram, R. and **Sobha Lalitha Devi**. (2013). "Pronominal Resolution in Tamil Using Tree CRFs", In Proceedings of 6th Language and Technology Conference, Human Language Technologies as a challenge for Computer Science and Linguistics - 2013, Poznan, Poland, LNAI pp. 333-337
34. Marimuthu K., Amudha K., Bakiyavathi T. and **Sobha Lalitha Devi**. (2013). "Word Boundary Identifier as a Catalyzer and Performance Booster for Tamil Morphological Analyzer", In Proceedings of 6th Language and Technology Conference, Human Language Technologies as a challenge for Computer Science and Linguistics - 2013, Poznan, Poland, LNAI
35. Akilandeswari A., and **Sobha Lalitha Devi**. (2013). "Conditional Random Fields Based Pronominal Resolution in Tamil ", In **International Journal on Computer Science and Engineering**, Vol. 5 Issue 6 pp 601 - 610
36. Vijay Sundar Ram and **Sobha Lalitha Devi**, (2012), "Coreference Resolution using Tree-CRF", In A. Gelbukh (ed), *Computational Linguistics and Intelligent Text Processing*, **Springer LNCS** Vol. 7181/2012 pp 285 - 296
37. Menaka S., Pattabhi R. K. Rao, **Sobha Lalitha Devi**, (2011), "Automatic Identification of Cause-Effect Relations in Tamil Using CRFs", In A. Gelbukh (ed), **Springer LNCS** Vol. 6608/2011 pp 316-327
38. **Sobha Lalitha Devi**, Pralayankar P., Kavitha V., Menaka S., (2011), "Translation of Hindi Se to Tamil in a MT syatem", In Chandan Singh .et .al (ed), *Information Systems for Indian Languages*, **Springer LNAI**, CCIS, 2011, Vol. 139, Part-2, ISBN - 978-3-642-19402-3 pp 246-249

39. **Sobha Lalitha Devi** and Menaka S., (2011), “Semantic Representation of Causality”, In Kommaluri Vijayanand, and L. Ramamoorthy, (ed), **LANGUAGE IN INDIA**, Vol. 11, ISSN 1930-2940
40. Sankar K., Vijay Sundar Ram R. and **Sobha Lalitha Devi**, (2011), “Text Extraction for an Agglutinative Language”, In Kommaluri Vijayanand, and L. Ramamoorthy, (ed), **LANGUAGE IN INDIA**, Vol. 11, ISSN 1930-2940
41. Vijay Sundar Ram R. and Pattabhi R.K. Rao and **Sobha Lalitha Devi**, (2011), “Identification of Different Feature Sets for NER Tagging Using CRFs and Its Impact”, In Kommaluri Vijayanand, and L. Ramamoorthy, (ed), **LANGUAGE IN INDIA**, Vol. 11, ISSN 1930-2940
42. **Sobha Lalitha Devi**, Vijay Sundar Ram R, Pravin Pralayankar, and Bakiyavathi T (2010) Syntactic Structure Transfer in a Tamil to Hindi MT System – A Hybrid Approach A. Gelbukh (Ed.): CICLING 2010, **Springer LNCS 6008**, pp. **438–450**.
43. **Sobha Lalitha Devi**, Sankar Kuppan, Kavitha Venkataswamy, and Pattabhi R.K. Rao (2009) “Identification of Similar Documents Using Coherent Chunks”, Anaphora Processing and Applications, **Springer LNAI**, volume **5847**, pp **54-68**.
44. Pralayankar P, Kavitha V and **Sobha L** (2009). "Case Marking Pattern from Hindi to Tamil MT", Volume-2, No-1, pp 26-31, **PIMT Journal of Research**, Ed.S. L Bharadwaj, Patiala, Punjab.
45. Menaka S and **Sobha L** (2009), "Optimising the Tamil Morphological Analyser", Volume-2, No-1, pp 13-16, **PIMT Journal of Research**, Ed.S. L Bharadwaj, Patiala, Punjab.
46. Vijay Sundar Ram and **Sobha L** (2009), “Tamil clause Identifier”, Volume-2, No-1, pp 42-46, **PIMT Journal of Research**, Ed.S. L Bharadwaj, Patiala, Punjab.
47. Vijay Sundar Ram and **Sobha Lalitha Devi** (2008). “**Clause Boundary Identification using Conditional Random Fields**”. Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science, LNCS, Volume 4919/2008, pp 140-150, **Springer Berlin / Heidelberg**.
48. Sobha, L. (2007). "Resolution of Pronominals in Tamil", **Computing Theory and Application**, the **IEEE Computer Society Press**, Los Alamitos, CA, pp. **475-79**.

Select Conference Papers (2016-2020)

1. C.S.Malarkodi and Sobha Lalitha Devi. (2021) “Identification of Names from Tamil Novel ”. In Tamil Computing - Tools and Applications Young Researchers' Conference 2021 (TaCTA-YRC2021)
2. Sinduja Gopalan and Sobha Lalitha Devi. (2021) “Role of Concession Connectives in Sentiment Analysis”. In Tamil Computing - Tools and Applications Young Researchers' Conference 2021 (TaCTA-YRC2021)
3. Nithya. J., Vijay Sundar Ram, R., and Sobha Lalitha Devi. (2021) “Translation of Verb chunks using ML Techniques”. In Tamil Computing - Tools and Applications Young Researchers' Conference 2021 (TaCTA-YRC2021)
4. C.S.Malarkodi, Pattabhi RK Rao and **Sobha Lalitha Devi**. (2020) “CLRG ChemNER: A Chemical Named Entity Recognizer @ ChEMU CLEF 2020”(using ML and Deep Learning). In ChEMU: Cheminformatics Elsevier Melbourne University, Conference and Labs of the Evaluation Forum 2020 (CLEF 2020).
5. Jiayuan He, Dat Quoc Nguyen,, Malarkodi C.S., Pattabhi RK Rao, **Sobha Lalitha Devi**,Timothy Baldwin and Karin Verspoor (2020) “An Extended Overview of ChEMU 2020: Named Entity Recognition and Event Extraction of Chemical Reactions from Patents”. In: Arampatzis A. et al. (eds) Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2020. Lecture Notes in Computer Science, vol 12260. Springer, Cham. https://doi.org/10.1007/978-3-030-58219-7_18

6. Vijay Sundar Ram and Sobha Lalitha Devi (2020). "Handling Noun-Noun Coreference in Tamil", In 5th Workshop on Indian Language Data: Resources and Evaluation (WILDRE), Organized under LREC-2020, Marseille, France.
7. Malarkodi C S and Sobha Lalitha Devi. (2020), "A Deeper Study on Features for Named Entity Recognition", In 5th Workshop on Indian Language Data: Resources and Evaluation (WILDRE), Organized under LREC-2020, Marseille, France.
8. Sobha Lalitha Devi (2019). "Resolving Pronouns for a Resource-Poor Language, Malayalam Using Resource-Rich Language, Tamil", In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), pp. 611–618., https://doi/10.26615/978954-452-056-4_072
9. Malarkodi C S and Sobha Lalitha Devi. (2019). "Fine-Grained Named Entity Recognizer for Tamil", In the Proceedings of 18th Tamil Internet Conference, Chennai
10. Sobha Lalitha Devi, Pattabhi RK Rao and Vijay Sundar Ram. (2018). "nigazaayvi" Event –Entity Profiling using Deep Syntactic and Semantic analysis of Tamil Documents, In proceedings of 17th Tamil Internet Conference, July 6 to 8, 2018, Coimbatore, India
11. Malarkodi C.S. and Sobha Lalitha Devi. (2018). "Twitter Named Entity Recognition for Indian Languages.", In proceedings of 18th International Conference on Computational Linguistics and Intelligent Text Processing, March 18 to 24, 2018, Hanoi, Vietnam
12. Vijay Sundar Ram and Sobha Lalitha Devi. (2018). "A Semi-automated Annotation of Co-reference Chains in Tamil", In proceedings of 18th International Conference on Computational Linguistics and Intelligent Text Processing, March 18 to 24, 2018, Hanoi, Vietnam
13. Vijay Sundar Ram and Sobha Lalitha Devi. (2017). "A Robust Coreference chain Builder for Tamil", In proceedings of 18th International Conference on Computational Linguistics and Intelligent Text Processing, April 17 to 23, 2017, Budapest, Hungary
14. Sindhuja Gopalan, and Sobha Lalitha Devi. (2017). "Cause and Effect Extraction from Biomedical Corpus", In proceedings of 18th International Conference on Computational Linguistics and Intelligent Text Processing, April 17 to 23, 2017, Budapest, Hungary
15. Vijay Sundar Ram and Sobha Lalitha Devi. (2016). "How to Handle Split Antecedents in Tamil?", In proceedings of Coreference Resolution Beyond OntoNotes co-located with NAACL 2016, San Diego, California
16. Pattabhi RK Rao and Sobha Lalitha Devi. (2016). CMEE-IL: Code Mix Entity Extraction in Indian Languages from Social Media Text @ FIRE 2016 – An Overview. In Working notes of FIRE 2016 - Forum for Information Retrieval Evaluation, Kolkata, India pp. 289-295.
17. Sobha Lalitha Devi and Pattabhi RK Rao.(2016). "Semantic Representation of Tamil Texts using Conceptual Graphs", In Proceedings of 3rd Workshop on Indian Language Data: Resources & Evaluation (WILDRE-3) under LREC 2016 in Portoroz, Slovenia
18. Sobha Lalitha Devi, Vijay Sundar Ram, Sindhuja Gopalan, Pattabhi RK Rao and Lakshmi S. (2016). "A Demo Proposal: Tamil – Hindi Automatic Machine Translation – A detailed Description", In Proceedings of 3rd Workshop on Indian Language Data: Resources & Evaluation (WILDRE-3) under LREC 2016 in Portoroz, Slovenia
19. Sobha Lalitha Devi, Pattabhi RK Rao, Vijay Sundar Ram and Malarkodi C.S.(2016). "A Demo Proposal: Tamil – English Cross Lingual Information Access (CLIA) System", In Proceedings of 3rd Workshop on Indian Language Data: Resources & Evaluation (WILDRE-3) under LREC 2016 in Portoroz, Slovenia
20. Sindhuja Gopalan and Sobha Lalitha Devi. (2016). "BioDCA Identifier: A System for Automatic Identification of Discourse Connective and Arguments from Biomedical Text". In Proceedings of

the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2016), pp 89–98, COLING 2016, Osaka, Japan, December 12th 2016.

Awards: (Recent)

2023- Best Active Consultant Award, Anna University, Chennai

(Achievement award for doing Research & Development Consultancy with Industry and Government)

2023-Best Paper Award, FOSS-CILT23 (Vijay Sundar Ram and Sobha L)

2021-Best Paper Award, TaCTA-YRC2021 (Sindhuja G and Sobah L)

Total Number of Publications:

SNo.	Type of Publication	Number of Publications
1	Journals	40
2	International Conferences	90
3	National Conferences	28
4	Books (edited)	5
5	Book Chapters	2

Edited Books

1. Rajiv Sangal, Dipti Misra Sharma, **Sobha L (Eds)** (2011) **Natural Language Processing (ICON2011)**.
2. Iris Hendrickx, **Sobha Lalitha Devi**, Antonio Branco, Ruslan Mitkov **(Eds.)** (2011) “Anaphora Processing and Application”. 8th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC2009. Volume 7099, **Lecture Notes in Artificial Intelligence, Springer Berlin/Heidelberg**.
3. Iris Hendrickx, **Sobha Lalitha Devi**, Antonio Branco, Ruslan Mitkov **(Eds.)** (2011) DAARC2009. Proceedings of 8th Discourse Anaphora and Anaphor Resolution Colloquium, Lisbon, Portugal
4. **Sobha Lalitha Devi**, Antonio Branco, Ruslan Mitkov **(Eds.)** (2009) “Anaphora Processing and Application”. 7th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC2009. Volume 5847, **Lecture Notes in Artificial Intelligence, Springer Berlin/Heidelberg**.
5. **Sobha Lalitha Devi**, Antonio Branco, Ruslan Mitkov **(Eds.)** (2009) DAARC2009. Proceedings of 7th Discourse Anaphora and Anaphor Resolution Colloquium, AU-KBC Research Centre, Anna University, Chennai, India

Tutorial

1. **Association of Computer Machines (ACM) SIGIR 2008, Singapore: Indian Language Information Retrieval: Dealing with Indian language Text Retrieval Issues**, Mandar Mitra, Prasenjit Majumder and **Sobha, L.** ,
2. **ICON2010, Indian Institute of Technology Khadakpur, Khadakpur , India. 2010 , Anaphora Resolution , An Introduction. Sobha Lalitha Devi.**

Students

SNo.	Programme	Number of students	Status of the programme
1	MS by Research	6	Completed
2	Ph.D	7	Completed
3	Ph.D	1	Submitted
4	B. E ,M.E/MCA Projects	25	Completed

Product:

Anil Chatbot (2019): Chatbot for e-governance. Artificial Intelligence (AI) Powered Natural Language Conversation Chatbot for Tamil Learning and E-Governance Application (Deployed at TNeGA)

Nigalazvi (2017): Event profiler: Extraction of various disaster events from the web for Tamil and English: Deployed at AUKBC site

Malayalam Spell Checker (2019) – An open source Malayalam Spell Checker integrated with Libre Office. *Deployed: in Libre Office, released on open source by ICFOSS- Govt. of Kerala*

Tamil NLP Tools Released for Research Purpose: *Released under open source, Available at AU-KBC Website*

searchKo (2007): searchko is a **Tamil portal**, which has a Tamil search engine as the main constituent. It has news aggregation and AdTrans an automatic advertisement translation system. More about searchko at www.searchko.in

Consultancy:

1. Oxford University Press, Oxford , UK, Tools for language computing.2018-2020
2. Microsoft , Hyderabad , India, Development of Tools for Dravidian Languages2017
3. ZowDow.Inc. Los Angeles , USA, Named Entity Recognizer for Indian Languages , Russian and Bhasa Indonesia.2017
4. Meltwater: Data Generation for Named Entity Recognition 2015
5. Malayalam University: Concise E-Dictionary for Malayalam Language2014-2017
6. Veda Semantics: Sentiment Analyser for Twitter data and Reviews April 2013 –April 2015
7. Emudra: Named Entity Recognizer for Twitter Data and Reviews April 2012- April 2013
8. Serendio Innovations: Entity Identification and Anaphora Resolution. January- April 2009.
9. Microsoft Research India (MSRI): Designing a Hierarchical Tagset for Indian Languages 2006-2008
10. Svapas Innovations, Consultant for BioAsk. 2006-2007
11. Burning Glass Inc, USA: Employment Trajectory analyzer (2005)
12. Information Extraction from classified documents for intelligence agency (2003-2005)

Projects

Duration	Projects	In Lakhs
March 2022-2025	Discourse Integrated Machine Translation between Dravidian Languages (DL-DISCO) Consortium Leader	1179.00

	6 other institutes in the consortium are (IIT Hyd, CDAC Noida, Central Univ Hyd, MIT Manipal, DAIICT Gandhinagar, ICFOSS Kerala) (MeitY ,MCIT), Govt of India, Research Project Under Bhashini Mission)	
March 2022-2025	E-IL Machine Translation system Consortium Leader CDAC Pune, (MeitY ,MCIT), Govt of India, Research Project Under Bhashini Mission)	100.30
March 2022-2025	Language Communicator Consortium Leader IIT BHU, (MeitY ,MCIT), Govt of India, Under Bhashini Mission)	54.00
December 2020	Chatbot for Tamil e- Learning, TVA, Tamil Nadu Government	4.00
March 2020 – August 2021	Speech to Speech Machine Translation, (Office of the Principal Scientific Advisor, Govt of India)	17.00
Feb 2020 –2021	“Natural Language Processing for e-Governance” -- A Smart Assistance for Website and Social ANIL Chatbot (TNeGA, Govt of Tamil Nadu)	15.00
July 2019 - 2021	Translating and sub-titling of video of UG and PG course Lecture content for SWAYAM Courses in 8 Indian Languages, (MHRD, Gov of India). Handling Tamil and Malayalam	250.00
August 2017- 2021	“A Platform for Cross Lingual and Multi Lingual Event Monitoring in Indian Language”; Under Imprint India Initiative, MHRD	99.96
March 2016- April 2017	Event-Entity Profiling from Tamil Documents– A Mobile Application , (Tamil Virtual Academy (TVA),Govt. of Tamil Nadu, Chennai,)	25.00
2010-2012 Extended upto2015	Indian Language-Indian Language Translation System, Tamil – Hindi Bilingual Translation System (Ministry of Communication and Information Technology (MeitY ,MCIT), Govt of India)	171.00
2010-2012 Extended upto 2017	Cross Lingual Information Access (Ministry of Communication and Information Technology (MeitY, MCIT), Govt of India)	138.00
2007	Testing and Evaluation of Tamil 16-bit All Character Encoding system for NLP (Tamil Virtual University (TVU), Chennai,)	5.50
2006-2008 Extended upto August 2010	Cross Lingual Information Access (Ministry of Communication and Information Technology (MeitY, MCIT), Govt of India)	98.00
2006-2008 Extended upto April 2010	Indian Language-Indian Language Translation System, Tamil – Hindi Biligual Translation System (Ministry of Communication and Information Technology (MeitY MCIT), Govt of India)	62.80
2005-2008	Information Extraction (General Domain) (Department of Science and Technology, (DST-Woman Scientist) Govt of India)	11.30
2004-2006	Development of Language Technology Tool to Aid Research in the Sociology of Environment (Funded by Indian Council for Social Sciences Research (ICSSR), India)	3.90
2003	Spread of Languages in the Internet: An Automatic Language Identifier(funded by Indiana University, USA, UNESCO Project of Prof. John Piolillo, USA, 2003)	\$6000 (Rs.3.00 L, Approx)
2005	Employment Trajectory analyzer (Burning Glass Inc, USA)	3.00

Consultancy		
--------------------	--	--

Total Funding from Projects and Consultancy

SNo	Type of Projects	No.	Total Amount INR
1	Government agencies funded Projects (DST, TVU Govt of TN., MeitY Govt. of India, Govt of Kerala.)	9	24.60 Cr
2	International Funding agency funded projects (European Union, FP7,through students exchange) and UNESCO	2	39 Lakhs
3	Consultancy Projects	10	1.00Cr

Teaching: Anna University

SNo	Course Title	Year
1	Natural Language Processing, Course Number (EL056)	2003, Fall
2	Natural Language Processing, Course Number (XE052)	2004, Spring
3	Natural Language Processing, Course Number (XE052)	2004, Fall
4	Natural Language Processing, Course Number (CN067)	2008, Fall
5	Discourse Analysis – Western and Eastern Perspectives (FI1973)	2012, Spring
6	Discourse Analysis – Western and Eastern Perspectives (FI1973)	2013, Spring

Workshops and Conference Organized

SNo	Conference/Workshop Title	Place	Year
1.	Workshop on Conversational Analysis , Co located with ICON2021	NIT Silchar, Assam	2021
2.	TACTA, Tamil Computing	Virtual	2020
3.	5 th Workshop on Indian Language Data: Resources and Evaluation (WILDRE) (Co organizer)	In conjunction with LREC 2022, France	2022
4	Tamil Internet Conference	Anna University, Chennai	2019
5	3 rd Workshop on Indian Language Data: Resources and Evaluation (WILDRE)	In conjunction with LREC 2016, Portorož (Slovenia)	2016 May
6	2 nd Workshop on Indian Language Data: Resources and Evaluation (WILDRE)	In conjunction with LREC 2014, Reykjavik, Iceland	2014
7	1 st Workshop on Indian Language Data: Resources and Evaluation (WILDRE)	In conjunction with LREC 2012, Istanbul, Turkey	2012
8	The iDravidian' 2017: Symposium on Natural Language Processing for Dravidian Languages	AU-KBC , Chennai	2017 May
9	The iDravidian' 2015: Symposium on Natural Language Processing for Dravidian Languages	co-located with ICON 2015, Trivandrum, India	2015 December
10	The iDravidian' 2014: Symposium on Natural	co-located with ICON	2014

	Language Processing for Dravidian Languages	2014, Goa, India	
11	Workshop on Machine Translation and Parsing in Indian Languages (MTPIL)	In conjunction with CoLing 2012, Mumbai, India	2012
12	The 8th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)	Faro, Portugal	2011
13	9th International Conference on Natural Language Processing (ICON-2011)	Chennai, India	2011
14	Workshop on Named Entity Recognition (under CLIA project)	Chennai, India	2011
15	The 7th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)	Goa, India	2009
16	Workshop on Named Entity Recognition	Chennai, India	2009
17	International Workshop on Referential Entity Resolution (RER)	Chennai, India	2008
18	Indo-German workshop on Technology transfer of VerbMobil project of DFKI, Phase I and Phase II (with Prof. C.N. Krishnan)	Chennai, India	2005

Shared Task Organization

1. **Social Media Anaphora Resolution(SocAnaRes2022) @ FIRE2022** Forum for Information Retrieval and Evaluation
2. **Social Media Anaphora Resolution(SocAnaRes2020) @ FIRE2020** Forum for Information Retrieval and Evaluation
3. **EventExtract-IL: Event Extraction from Newswires and Social Media Text in Indian Languages 2nd Edition @ FIRE2018** Forum for Information Retrieval and Evaluation - **2018**.
4. **EventExtract-IL: Event Extraction from Newswires and Social Media Text in Indian Languages, @FIRE2017** Forum for Information Retrieval and Evaluation - **2017**.
5. **CMEE-IL: Code Mix Entity Extraction for Indian Languages, @FIRE2016** Forum for Information Retrieval and Evaluation - **2016**.
6. **ESM-IL: Entity Extraction from Social Media Text for Indian Languages @ FIRE2015** Forum for Information Retrieval and Evaluation - **2015**.
7. **NER-IL: Named Entity Recognition for Indian Languages 2nd Edition, @FIRE2014** Forum for Information Retrieval and Evaluation - **2014**.
8. **NER-IL: Named Entity Recognition for Indian Languages, 1st Edition, FIRE2013** Forum for Information Retrieval and Evaluation - **2013**.
9. **CL!TR – Cross Lingual Indian Language Text Reuse, evaluation track @FIRE2011** Forum for Information Retrieval and Evaluation – **2011**.