# The Group of Artificial Intelligence-Natural Language Processing-Machine Learning (http://www.iitp.ac.in/~ai-nlp-ml/)

# Department of Computer Science & Engineering
# IIT Patna
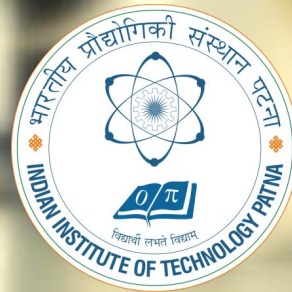
# History: AI-NLP-ML Group

- Was set up in the Computer Science and Engineering Dept. on June 03, 2015

- Inaugurated on December 07, 2015

# About Our Lab, at a Glance..

## People

**Associated faculty:**
**---Prof. Pushpak Bhattacharyya**
**--Dr. Asif Ekbal**
**--Dr. Sriparna Saha**
**Students:**
**PhD :** Graduated-5; Ongoing-22
**M.Tech** (so far): 30+
**B.Tech** (so far): 62+
**Linguists & Staffs: 4**

## Research & Outreach

Publications in top NLP & AI Journals and Conferences
**Conferences**: ACL, EMNLP, EACL, COLING, ICPR etc.
 **Journals**: Knowledge based System, Gene Reports, Expert Systems with Application, ACM TKDD, BMC Bioinformatics etc.

## Collaboration

**Sponsorships:** Elsevier, Accenture, Samsung, ezDI, LG, CDOT, MeiTY, MHRD etc.

**Associations** with universities (IIT Bombay, University of Trento- Italy, TU Darmstadt- Germany, University of Mainz- Germany, L3S Research Centre-Germany, University of Hamburg- Germany, University of Caen-France, Macquarie University-Australia, WSU-USA , Kyoto University-Japan)

# The Premise of our AI-NLP-ML: Research

**Areas**
NLP
Bioinformatics
Text mining
Information Extraction

**Techniques**
Supervised classifiers
Probabilistic models
Deep learning
Clustering
Multi objective optimization

**Languages**
English
Indian languages: Hindi,
Bengali etc.

**Courses**
Artificial Intelligence
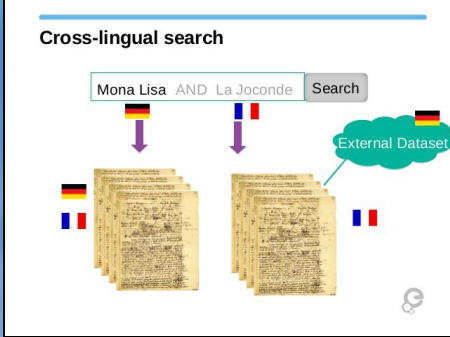Natural Language
Processing
Pattern Recognition
Machine Learning

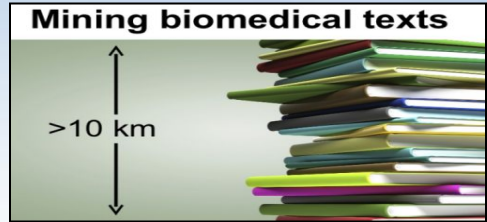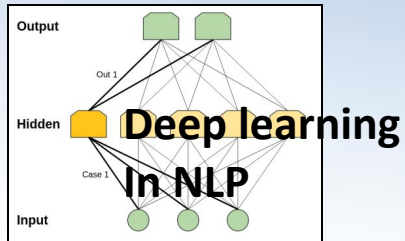**Facilities**

High performance clusters,
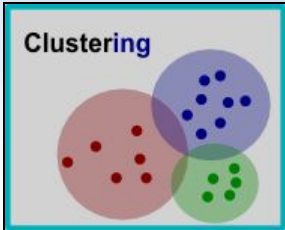*Workstations etc.*

**Synergy**
Fortnightly AI-NLP-ML
'entertainment' talks

Regular participation in
NLP shared tasks

Machine Translation

Cross-lingual search

Mona Lisa AND La Joconde   Search

External Dataset

SENTIMENT ANALYTICS

Bioinformatics

Temporal IR

Ongoing research areas of AI-NLP-ML lab

Artificial Intelligence and Machine Learning in Healthcare

NEWS → Information Extraction → who; what; when?

Clustering

Output

Out 1

Hidden

Case 1

Input

Deep learning In NLP

Mining biomedical texts

>10 km
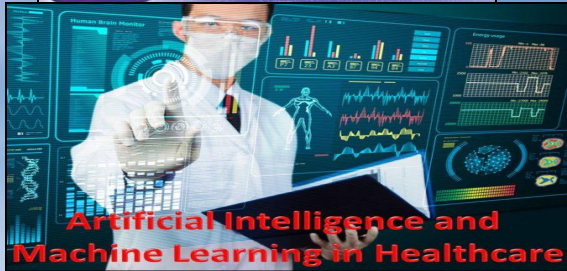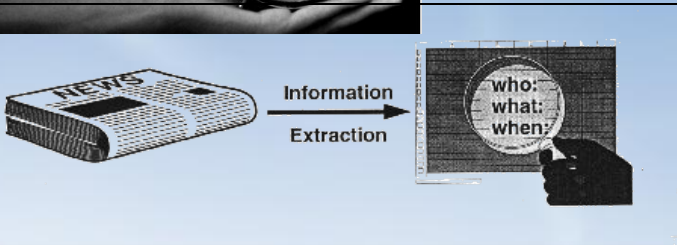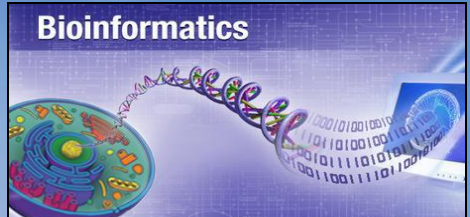
# Ongoing research areas

- **Machine Translation (E-IL, IL-E):** Neural MT, SMT
- **Information Retrieval:** Novelty Detection in Texts, Temporal Sense Detection, Event Monitoring (Cross-lingual and Multi-lingual)
- **Temporal Information Processing:** Temporal Sense Detection, Users' psychological behavior (Emotion, Sentiment etc.) analysis w.r.t to time
- **Sentiment and Emotion Analysis:** Aspect based sentiment analysis, Sentiment analysis in Indian languages, Sentiment analysis in cross-lingual and multilingual environment, Sentiment analysis in code-mixed environment, Emotion Analysis in English and Indian languages
- **Information Extraction:** NER, Coreference Resolution, Relation Extraction, Question-Answering , Summarization, Dialogues etc.
- **Bioinformatics:** Gene expression data clustering, Micro-RNA classification, Biclustering of gene expression data, Gene selection, Protein-Protein Interaction
- **Text Mining in Biomedical/Health Care Domain:** NER, Relation Extraction, Coreference resolution, Social Health etc.
- **Multi-modality**: NLG, Summarization, Sentiment, Emotion, QA

# Representative List of Projects

- Centre of Excellence for Natural Language Processing-Reed Elsevier Publishing Company
  - Some novel R&D problems in the broad areas of AI, NLP and ML  (AI Assisted Peer Review System, Email Classification etc. )
- ATL IIT AI Lab- Accenture Pvt. Limited
  - Research on QA, Multilingual Support and Virtual Agent
- Shusrut: ezDI Research Lab on Health Informatics-ezDI (Ahmadabad, India; Head office-USA)
- MeiTY, Govt. of India (with IITB, IIT Delhi, IIITH, IIITA, CDAC Pune)
  - **Hindi to English Machine Translation for Judicial Domain**

- Imprint India, MHRD  (with IIT Kharagpur, DAICT, Au-KBC and IITB)
  - **A Platform for Cross-lingual and Multilingual Event Monitoring in Indian Languages**
- **Imprint-II, MHRD (With IIIT-Hyderabad, Jadavpur University)**
  - **Sevak:** An Intelligent Indian Language Chatbot
  - **Domains:** Railways, Judicial and Health
  - **Languages:** English, Hindi, Telugu, Bengali

Activities are under the "**Centre of Excellence on Artificial Intelligence**"

# Other Projects with Industries

- LG Soft, Samsung Research, Skymap Global

Problems:

Information Extraction, SMS classification,  Recommendation System, Image Recognition, Multimodal Sentiment and Emotion Analysis, Dynamic Natural Language Generation, Knowledge graph enabled conversations

# Publications/resources from group

- Since June 2015
    - Number of conference paper: Approx. 120 (ACL, EMNLP, COLING, EACL, NAACL ICPR, IJCNN, IEEE TEC etc.)
    - Number of Journal papers : 65 (PLOS ONE, ACM TALLIP, Knowledge based System, Gene Reports, Expert Systems with Application, BMC Bioinformatics, ACM TKDD etc.)
- Created various lexical resources and tools for
    - Named Entity Recognition (Bengali)
    - Sentiment Analysis (Hindi)
    - Temporal Information Processing (English and Hindi)
    - Novelty Detection (English)
    - Severity Detection in Medical Text (Medical Blogs)
    - Anaphora Resolution (Bengali and Hindi)

*Some of these resources have already been made available*

# Selected Publications(Journals: Since June 2015)

1. S. Kamila, M. Hasanuzzaman, **A. Ekbal** and P. Bhattacharyya (2019). Resolution of grammatical tense into actual time, and its application in Time Perspective study in the tweet space. *PLOS ONE*, https://doi.org/10.1371/journal.pone.0211872.

2. Md. Shad Akhtar, Palaash Sawant, Sukanta Sen, **Asif Ekbal**, Pushpak Bhattacharyya (2019). Improving Word Embedding Coverage in Less-Resourced Languages Through Multi-Linguality and Cross-Linguality: A Case Study with Aspect-Based Sentiment Analysis. *ACM Trans. Asian & Low-Resource Lang. Inf. Process*. 18(2): 15:1-15:22.

3. Sabyasachi Kamila, Mohammed Hasanuzzaman, **Asif Ekbal**, Pushpak Bhattacharyya (2019). Tempo-HindiWordNet: A Lexical Knowledge-base for Temporal Information Processing. *ACM Trans. Asian & Low-Resource Lang. Inf. Process*. 18(2): 19:1-19:22

4. Shweta Yadav, **Asif Ekbal**, Sriparna Saha, Ankit Kumar, Pushpak Bhattacharyya (2019). Feature assisted stacked attentive shortest dependency path based Bi-LSTM model for protein-protein interaction. *Knowledge Based System*. 166: 18-29, Elsevier.

5. Debajyoty Banik, Asif Ekbal, Pushpak Bhattacharyya (2019). Machine Learning Based Optimized Pruning Approach for Decoding in Statistical Machine Translation. *IEEE Access* 7: 1736-1751.

# Selected Publications(Journals: Since June 2015)

6.  Debajyoty Banik, **Asif Ekbal**, Pushpak Bhattacharyya, Siddhartha Bhattacharyya (2019). Assembling translations from multi-engine machine translation outputs. *Appl. Soft Comput*. 78: 230-239, Elsevier.

7.  S. Akhtar, D. Gupta, A. Ekbal and P. Bhattacharyya. Feature Selection and Ensemble Construction: A Two-step Method for Aspect Based Sentiment Analysis, Knowledge-Based Systems, Elsevier, 2017.

8.  S. Acharya, S. Saha. Importance of Proximity Measures in Clustering of Cancer and miRNA Datasets: Proposal of an Automated Framework. In Molecular BioSystems, 2016, DOI: 10.1039/C6MB00609D.

9.  U. K. Sikdar, A. Ekbal and S. Saha (2016). A Generalized Framework for Anaphora Resolution in Indian Languages, in Knowledge based Systems, Elsevier

10. S. Saha, S. Mitra, and R. K. Yadav (2016): A Multiobjective based Automatic Framework for Classifying Cancer-microRNA Biomarkers, Gene Reports

11. S. Saha, A. Alok, A. Ekbal (2016): Brain Image Segmentation using Semi-supervised Clustering, Expert Systems with Applications, Elsevier

# Selected Publications(Conferences: Since June 2015)

1. Sukanta Sen, Kamal Kumar Gupta, **Asif Ekbal** and Pushpak Bhattacharyya (2019). Multilingual Unsupervised NMT using Shared Encoder and Language-Specific Decoders. In *Proceedings of Association for Computational Linguistics* **(ACL).** (*Accepted*, Core-A*)

2. Tirthankar Ghosal, Rajeev Verma, **Asif Ekbal** and Pushpak Bhattacharyya (2019). DeepSentiPeer: Harnessing Sentiment in Review Texts To Recommend Peer Review Decisions. In *Proceedings of Association for Computational Linguistics* **(ACL)**. (*Accepted*, Core-A*)

3. Hardik Chauhan, Mauzama Firdaus, **Asif Ekbal** and Pushpak Bhattacharyya (2019). Ordinal and Attribute Aware Response Generation in a Multimodal Dialogue System. In *Proceedings of Association for Computational Linguistics (ACL)*. (*Accepted*, Core-A*)

4. S. Yadav, **A. Ekbal**, S. Saha and P. Bhattacharyya (2019). A Unified Multi-task Adversarial Learning Framework for Pharmacovigilance Mining. In *Proceedings of Association for Computational Linguistics (ACL)*. (*Accepted*, Core-A*)

5. Mohammed Hasanuzzaman, Sabyasachi Kamila, Mandeep Kaur, Sriparna Saha and Asif Ekbal (2017). Temporal Orientation of Tweets for Predicting Income of User. In Association for Computational Linguistics **(ACL** 2017).

6. Tirthankar Ghosal, Ashish Raj, **Asif Ekbal**, Sriparna Saha and Pushpak Bhattacharyya (2019). A Deep Multimodal Investigation To Determine the Appropriateness of a Scholarly Submission. In *Proceedings of ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. June 2-6, 2019.  Urbana-Champaign, Illinois (CORE rank A*)- *Accepted*.

# Selected Publications(Conferences: Since June 2015)

7.  Tirthankar Ghosal, **Asif Ekbal**, Sriparna Saha, Pushpak Bhattacharyya and Ravi Sonam (2019). Is the Paper Within Scope? Are You Fishing in the Right Pond? *Proceedings of ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, June 2-6, 2019.  Urbana-Champaign, Illinois (CORE rank A*)-*Accepted*.

8.  Tirthankar Ghosal, Debomit Dey, Avik Dutta, **Asif Ekbal**, Sriparna Saha and Pushpak Bhattacharyya (2019). A Multiview Clustering Approach To Identify Out-of-Scope Submissions. *Proceedings of ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, June 2-6, 2019. Urbana-Champaign, Illinois. (CORE rank A*)- *Accepted*.

9.  Tirthankar Ghosal, Ananya Chakraborty, Ravi Sonam, **Asif Ekbal**, Sriparna Saha and Pushpak Bhattacharyya (2019). Incorporating Full Text and Bibliographic Features to Improve Scholarly Journal Recommendation. *Proceedings of ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, June 2-6, 2019.  Urbana-Champaign, Illinois (CORE rank A*)- *Accepted*.

10. Tirthankar Ghosal, Rajeev Verma, **Asif Ekbal** and Pushpak Bhattacharyya (2019). A Sentiment Augmented Deep Architecture to Predict Peer Review Outcomes. *Proceedings of ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, June 2-6, 2019.  Urbana-Champaign, Illinois  (CORE rank A*)-Accepted.

11. Dushyant Singh Chauhan, Md Shad Akhtar, **Asif Ekbal,** and Pushpak Bhattacharyya (2019). Context-aware Interactive Attention for Multi-modal Sentiment and Emotion Analysis. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2019); Hong Kong; 2019. (*Accepted*, Core-A, h5-index: 88)

# Selected Publications(Conferences: Since June 2015)

12. Deepanway Ghosal, Md Shad Akhtar, Dushyant Chauhan, Soujanya Poria, **Asif Ekbal** and Pushpak Bhattacharyya (2018). Text, Visual and Acoustic are Friends! A Multi-Modal Attention Framework for Utterance-Level Sentiment Prediction. In *Proceedings of EMNLP, PP. 3454-3464*, November 2- 4, 2018. Brussels, Belgium (Core-A, h5-index: 88).

13. Md Shad Akhtar, Dushyant Chauhan, Deepanway Ghosal, Soujanya Poria, **Asif Ekbal**, and Pushpak Bhattacharyya (2019). Multi-task Learning for Multi-modal Emotion Recognition and Sentiment Analysis. June 2 -7, 2019. Minneapolis, USA (accepted, Core-A, h5-index: 61).

14. Shweta, Asif Ekbal, Sriparna Saha and Pushpak Bhattacharyya (2017). Entity Extraction in Biomedical Corpora: An Approach to Evaluate Word Embedding Features with PSO based Feature Selection. In European Chapter of the Association for Computational Linguistics **(EACL 2017).**

15. Md Shad Akhtar, Ayush Kumar, Asif Ekbal and Pushpak Bhattacharyya (2016). A Hybrid Deep Learning Architecture for Sentiment Analysis. In proceeding of the 26th International Conference on Computational Linguistics **(COLING 2016),**

16. Shweta, Asif Ekbal, Sriparna Saha and Pushpak Bhattacharyya (2016). A Deep Learning Architecture for Protein-Protein Interaction Article Identification. In 23$^{rd}$ International Conference of Pattern Recognition **(ICPR 2016).**

17. M. S. Akhtar, A. Ekbal and P. Bhattacharyya (2016). Aspect based Sentiment Analysis in Hindi: Resource Creation and Evaluation. In proceeding of the 10th edition of the Language Resources and Evaluation Conference (**LREC 2016).**

18. Dipawesh Pawar, Mohammed Hasanuzzaman and Asif Ekbal (2016). Building Tempo-HindiWordnet-A resource for temporal information access in Hindi. *In LREC*

# Awards and Achievements of group

- Eminent Engineer Award: Prof. Pushpak Bhattacharyya
- ACL President: Prof. Pushpak Bhattacharyya
- INAE Fellow: Prof. Pushpak Bhattacharyya
- NASI-Young Scientist Platinum Jubilee Awards (2016)-Dr. Sriparna Saha
- IEI Young Engineers award 2015-Dr. Sriparna Saha
- JSPS Fellowship, Japan: Dr. Asif Ekbal
- Visvesvaraya Young Faculty Research Fellowship Award, Govt. of India : Dr. Asif Ekbal
- GOOGLE, MICROSOFT Support: Sabyasachi, Shweta Yadav (PhD Student)
- Selected for Google NLP Summit, Zurich: Shad, Shweta
- One paper from AI-NLP-ML group's research received best paper award in CLINICAL NLP workshop collocated with COLING-16
- Most downloadable article recognition, Data and Knowledge Engineering, Elsevier

# Workshop/Courses organized

- Prof. Sadao Kurohasi- Kyoto University, Japan
  - GIAN Workshop on NLP in summer (May 2-8, 2016)
- Prof. Carlos A Coello Coello- CINVESTAV/IPN, Mexico

  -GIAN Workshop on Multi-objective Optimization during December 15-22,2016
- Prof. Massimo Poesio-University of Essex, UK and University of Trento, Italy
  - Offered a course on Cognitive Science for Computational Linguistics during March 27-31, 2017
- Prof. Andy Way, Dublin City University, Ireland
  - Offered a a course on Neural Machine Translation during December 4-10, 2017
- Prof. Erik Cambria, NTU, Singapore
  - Offered a course on Big Social Data Analysis during Feb 26-March 02, 2018

**AI-NLP-ML Lab**
**Room Number: 510, 5th Floor, Block III,**
**Department of Computer Science and Engineering**
**Indian Institute of Technology Patna**
**Patna 801103, India**
**Web: http://www.iitp.ac.in/~ai-nlp-ml/**

Contact Number
+91-612-3028338