

# Shibamouli Lahiri

## Home Address

1000 N Huron River Dr Apt 112  
Ypsilanti, MI 48197  
(814) 321-7351  
shibamouli@yahoo.com

## Office Address

1161 North Ingalls Building  
300 North Ingalls Street  
Ann Arbor, MI 48109  
lahiri@umich.edu

## **Broad Research Interests**

- Natural Language Processing (NLP) and Computational Linguistics
- Data Mining

## **Specific Research Interests**

- Computational Sociolinguistics, Psycholinguistics and Anthrop linguistics
- Computational Stylistics and Stylometry
- Culturomics, Longitudinal Analysis, Time Series data.
- Sentiment and Emotion Analysis. Affect Mining
- Complex Network Properties of Text
- NLP in Indian Languages (esp. Bengali)
- Summarization and Keyword Extraction

## **Education**

- **Doctor of Philosophy** Fall 2014 - Winter 2018  
Computer Science and Engineering  
University of Michigan  
**Dissertation:** Keywords at Work: Investigating Keyword Extraction in Social Media Applications  
**Degree:** August 17, 2018  
**GPA: 3.89/4.0**
- **Ph.D. Student** Fall 2012 - Summer 2014  
Computer Science and Engineering  
University of North Texas  
**GPA: 3.727/4.0**
- **Master of Engineering** Fall 2008 - Summer 2012  
Computer Science and Engineering  
(Minors in Statistics and Computational Science)  
The Pennsylvania State University  
**GPA: 3.24/4.0**
- **Bachelor of Engineering** Fall 2004 - Spring 2008  
Computer Science and Engineering  
Jadavpur University, Kolkata, India  
**GPA: 8.83/10.0**  
**Class Rank: 7th (First Class Honors)**

## **Past Projects**

- p **Predicting Student Grade and Student Stress from Piazza Discussions**
  - Advisor: Dr Rada Mihalcea
  - Collaborator: Dr Carmen Banea
- p **Product Usage Behavior Mining**
  - Advisors: Dr Rada Mihalcea, Dr V. G. Vinod Vydiswaran
- p **Keyword Extraction from Emails**

- Advisor: Dr Rada Mihalcea
  - Collaborator: Dr Po-Hsiang Lai, Samsung Research America
- p **Matching Graduate Applicants with Faculty**
- Advisors: Dr Rada Mihalcea, Dr Carmen Banea
  - Collaborators: Joseph Zimmer, Lauren Molley
- p **Designing a Formality Score at Sentence Level**
- Collaborator: Dr Xiaofei Lu
- p **ACL ARC Style Browser**
- p **Applying Dynamic Games and Reinforcement Learning to Community Consensus-building and Leader Selection**
- Collaborators: Partha Mukherjee (Project Leader), Dr Jens Grossklags (Faculty Mentor), Dr Kalyan Chatterjee (Faculty Mentor)
- p **Multi-document Summarization in Persian**
- Collaborator: Saeid Parvande

#### Side Projects

- s **Culturomics and Stylistics in Bengali Language**
- Collaborators: Shanta Phani, Dr Arindam Biswas

#### Selected Journal Publications

1. Shanta Phani, **Shibamouli Lahiri**, Arindam Biswas, *A Supervised Learning Approach for Authorship Attribution for Bengali Language Literary Texts*, ACM Transactions on Asian and Low-Resource Language Information Processing (TAL-LIP), 2017.
2. **Shibamouli Lahiri**, Rada Mihalcea, Po-Hsiang Lai, *Keyword Extraction from Emails*, Journal of Natural Language Engineering (JNLE), 2016.
3. Saeid Parvande, **Shibamouli Lahiri**, Fahimeh Boroumand, *PerSum: Novel Systems for Document Summarization in Persian*, International Journal of Asian Language Processing (IJALP), 2016.

#### Selected Conference Publications

1. **Shibamouli Lahiri**, V. G. Vinod Vydiswaran, Rada Mihalcea, *Identifying Usage Expression Sentences in Consumer Product Reviews*, IJCNLP 2017.
2. **Shibamouli Lahiri**, Carmen Banea, Rada Mihalcea, *Matching Graduate Applicants with Faculty Members*, Socinfo 2017.
3. Shanta Phani, **Shibamouli Lahiri**, Arindam Biswas, Sudipta Dutta, *An Indic Language N-gram Viewer*, FIRE 2016.
4. Shanta Phani, **Shibamouli Lahiri**, Arindam Biswas, *A Machine Learning Approach for Authorship Attribution for Bengali Blogs*, IALP 2016.
5. Shanta Phani, **Shibamouli Lahiri**, Arindam Biswas, *Authorship Attribution in Bengali Language*, ICON 2015.
6. Vanessa Loza, **Shibamouli Lahiri**, Rada Mihalcea, Sean Lai, *Building a Dataset for Summarization and Keyword Extraction from Emails*, LREC 2014.
7. Shanta Phani, **Shibamouli Lahiri**, Arindam Biswas, *Inter-rater Agreement Study on Readability Assessment in Bengali*, ICONACC 2014.
8. Shanta Phani, **Shibamouli Lahiri**, Arindam Biswas, *Culturomics On A Bengali Newspaper Corpus*, IALP 2012.

9. **Shibamouli Lahiri**, Prasenjit Mitra, Xiaofei Lu, *Informality Judgment at Sentence Level and Experiments with F-score*, CICLING 2011.
10. Sumit Bhatia, **Shibamouli Lahiri**, Prasenjit Mitra, *Generating Synopses For Document-element Search*, CIKM 2009.

**Selected  
Workshop  
Publications**

1. Shanta Phani, **Shibamouli Lahiri**, Arindam Biswas, *Personality Recognition in Source Code Working Note: Team BESUMich*, Shared Task on Personality Recognition in SOurce COde (PR-SOCO), held in conjunction with FIRE 2016 (Forum for Information Retrieval Evaluation).
2. Shanta Phani, **Shibamouli Lahiri**, Arindam Biswas, *Sentiment Analysis of Tweets in Three Indian Languages*, 6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP 2016).
3. **Shibamouli Lahiri**, *Complexity of Word Collocation Networks: A Preliminary Structural Analysis*, EACL 2014 Student Research Workshop.
4. **Shibamouli Lahiri**, Rada Mihalcea, *Using N-gram and Word Network Features for Native Language Identification*, BEA 2013 Workshop.

**Selected  
arXiv  
e-prints**

1. Saeid Parvande, **Shibamouli Lahiri**, Fahimeh Boroumand, *PerSum: Novel Systems for Document Summarization in Persian*, ArXiv e-print (2016).
2. **Shibamouli Lahiri**, *SQUINKY! A Corpus of Sentence-level Formality, Informativeness, and Implicature*, ArXiv e-print (2015).
3. **Shibamouli Lahiri**, Sagnik Ray Choudhury, Cornelia Caragea, *Keyword and Keyphrase Extraction Using Centrality Measures on Collocation Networks*, ArXiv e-print (2014).
4. **Shibamouli Lahiri**, Rada Mihalcea, *Authorship Attribution Using Word Network Features*, ArXiv e-print (2013).
5. **Shibamouli Lahiri**, Xiaofei Lu, *Inter-rater Agreement on Sentence Formality*, ArXiv e-print (2011).
6. **Shibamouli Lahiri**, Juan Pablo Fernández Ramírez, Shikha Nangia, Prasenjit Mitra, C. Lee Giles, Karl T. Mueller, *ChemXSeer Digital Library Gaussian Search*, ArXiv e-print (2011).

**Summer  
Internship**

- |  |             |
|--|-------------|
| <b>NEC Labs</b> , Princeton, NJ  | Summer 2011 |
| <ul style="list-style-type: none"> <li>• Project: Multiple-choice Question Answering</li> <li>• Mentors: Dr Christopher Malon, Dr Bing Bai</li> </ul>  |             |
| <b>IBM India Research Lab (IBM IRL)</b> , Delhi, India   | Summer 2010 |
| <ul style="list-style-type: none"> <li>• Project: Topic Modeling of Call Center Chats</li> <li>• Mentor: Sachindra Joshi</li> <li>• Collaborators: Avinava Dubey, Dr Shantanu Godbole</li> </ul> |             |

**Research  
Assistantship**

- |   |                              |
|---|------------------------------|
| <b>Computer Science and Engineering</b> , University of Michigan  | Winter 2017 -<br>Winter 2018 |
| <ul style="list-style-type: none"> <li>• Project: LEAP (Analytics for <u>L</u>earners <u>A</u>s <u>P</u>eople)</li> <li>• Supervisor: Dr Rada Mihalcea</li> </ul>   |                              |
| <b>Computer Science and Engineering</b> , University of North Texas   | Summer 2013 -<br>Summer 2014 |
| <ul style="list-style-type: none"> <li>• Project: Email Summarization and Keyword Extraction</li> <li>• Supervisor: Dr Rada Mihalcea</li> <li>• Sponsor: Samsung Research America, Richardson TX</li> </ul> |                              |

	<b>Architectural Engineering</b> , Penn State University	Spring 2012
	<ul style="list-style-type: none"> <li>• Project: DAYSIM (Java programming)</li> <li>• Supervisor: Dr Richard Mistrick</li> </ul>	
	<b>Political Science</b> , Penn State University	Fall 2011
	<ul style="list-style-type: none"> <li>• Project: Political opinion extraction from parliamentary speeches</li> <li>• Supervisor: Dr Burt Monroe</li> </ul>	
	<b>Information Sciences and Technology</b> , Penn State University	Spring 2011,
	<ul style="list-style-type: none"> <li>• Project: Cypriot Digital Antiquities Repository (A digital library of Ancient Greek manuscripts collected from Cyprus)</li> <li>• Advisor: Dr Prasenjit Mitra</li> <li>• Collaborators: Dr Zoe Stamatopoulou, Partha Mukherjee</li> </ul>	Spring-Fall 2010
	<b>Information Sciences and Technology</b> , Penn State University	Fall 2009
	<ul style="list-style-type: none"> <li>• Project: ChemXSeer Gaussian Search</li> <li>• Advisor: Dr Prasenjit Mitra</li> <li>• Collaborator: Dr Shikha Nangia</li> </ul>	
	<b>Computer Science and Engineering</b> , Jadavpur University	Spring 2008,
	<ul style="list-style-type: none"> <li>• Project: Cross-lingual (Bengali <math>\iff</math> English) and monolingual information extraction</li> <li>• Advisor: Prof Sivaji Bandyopadhyay</li> </ul>	Fall 2007
<b>Teaching Assistantship</b>	<b>Computer Science and Engineering</b> , University of Michigan	Fall 2016
	<ul style="list-style-type: none"> <li>• Course: EECS 498 (Natural Language Processing)</li> <li>• Instructor: Dr Rada Mihalcea</li> </ul>	
	<b>Computer Science and Engineering</b> , University of Michigan	Winter 2016,
	<ul style="list-style-type: none"> <li>• Course: EECS 498 (Information Retrieval and Web Search)</li> <li>• Instructor: Dr Rada Mihalcea</li> </ul>	Winter 2015
	<b>Computer Science and Engineering</b> , University of Michigan	Fall 2015
	<ul style="list-style-type: none"> <li>• Course: EECS 492 (Introduction to Artificial Intelligence)</li> <li>• Instructor: Dr Benjamin Kuipers</li> </ul>	
	<b>Computer Science and Engineering</b> , University of Michigan	Fall 2014
	<ul style="list-style-type: none"> <li>• Course: EECS 183 (Elementary Programming Concepts)</li> <li>• Instructors: Dr Mary Lou Dorf, Jeremy Gibson, William Arthur</li> </ul>	
	<b>Computer Science and Engineering</b> , University of North Texas	Spring 2013,
<ul style="list-style-type: none"> <li>• Course: CSCE 1020 (Program Development)</li> <li>• Instructor: Dr Robert Gary Goodman</li> </ul>	Fall 2012	
<b>Information Sciences and Technology</b> , Penn State University	Spring 2009	
<ul style="list-style-type: none"> <li>• Course: IST 451 (Computer Security)</li> <li>• Instructor: Dr Sencun Zhu</li> </ul>		
<b>Information Sciences and Technology</b> , Penn State University	Fall 2008	
<ul style="list-style-type: none"> <li>• Course: IST 220 (Computer Networks)</li> <li>• Instructor: Dr Prasenjit Mitra</li> </ul>		

<b>Graduate Mentoring</b>	g Shanta Phani, Ph.D. from Indian Institute of Engineering Science and Technology, Shibpur	Fall 2011 - Summer 2019
	g Saeid Parvande, Ph.D. from University of Tulsa	Summer 2012 - Summer 2016
	g Soumik Mandal, Ph.D. student at Ohio State University	Spring-Summer 2016
	g Keshavan Ravi, Formerly Master's student at University of North Texas	2013, Fall 2012
	g Debashis Ganguly, Ph.D. student at University of Pittsburgh	Spring-Fall 2011
	g Kheyali Mitra, Master's graduate from Simon Fraser University	Summer-Fall 2011
<b>Undergraduate Mentoring</b>	u Joseph Zimmer, University of Michigan	Summer-Fall 2014
	u Lauren Molley, University of Michigan	Summer-Fall 2014
	u Saranya Roy, BPPIMT Kolkata	Spring-Summer 2008
	u Ritanjana Adhikary, BPPIMT Kolkata	Spring-Summer 2008
<b>Invited Talk</b>	Visva-Bharati University, Santiniketan, West Bengal, India One-day International Webinar	
	• State of the Art and Future of Text Mining	Fall 2020
	• Methods for Clinical Information Extraction - by Scott L. DuVall, PhD	Fall 2020
<b>Guest Lecture</b>	University of Michigan Computer Science and Engineering Natural Language Processing Reading Group	
	• Detecting Promotional Content in Wikipedia - by Shruti Bhosale, Heath Vinicombe, Raymond Mooney	Winter 2014
	• A Just-In-Time Keyword Extraction from Meeting Transcripts - by Hyun-Je Song, Junho Go, Seong-Bae Park, Se-Young Park	Fall 2013
	APLNG 578 Corpus Analysis (graduate course at Penn State) • Inter-rater Agreement on Sentence Formality	Spring 2011
<b>Book Editor</b>	Network and Application Security: Fundamentals and Practices <i>by Debashis Ganguly (Science Publishers, Enfield, New Hampshire)</i>	
<b>Reviewer</b> <i>(inter alia)</i>	<ol style="list-style-type: none"> <li>1. Conferences – <i>AAAI, CIKM, EDBT, ICDM, SIGMOD, CICLING 2013, FLAIRS 2013, ACM SAC 2013 (Recommender Systems Track), ICONACC 2014, EMNLP 2014 (Semantics Track), IHCI 2014, FLAIRS 2015, FLAIRS 2017, FLAIRS 2018, FLAIRS 2019, CODS-COMAD 2020</i></li> <li>2. Journals – <i>Computational Linguistics, Computer Speech and Language, Journal of Information Science (Sage Publications), Arabian Journal for Science and Engineering (AJSE), Computational Intelligence (Wiley), Journal of the Association for Information Science and Technology (JASIST), Information Processing and Management (Elsevier)</i></li> </ol>	

**PC Member**  
(*inter alia*)

- NAACL-HLT 2018
- LREC 2018
- EMNLP 2017 Stylistic Variation Workshop
- COLING 2016 (Lexical Semantics, Ontologies Track)
- AAAI 2016 Workshop on Scholarly Big Data: AI Perspectives, Challenges, and Ideas
- Keyphrase 2015 (ACL 2015 International Workshop on Novel Computational Approaches to Keyphrase Extraction)
- ICCCNT 2014 (5th International Conference on Computing, Communication and Networking Technologies)

**Co-Chair**

NAACL Student Research Workshop (SRW) 2015

**Events Attended**  
(*inter alia*)

- e 4th Annual MIDAS (Michigan Institute for Data Science) Symposium, Ann Arbor, MI, USA October 2018
- e Beyond the Admin Shell: Proposal Editing 101, Ann Arbor, MI, USA August 2018
- e SIGIR 2018, Ann Arbor, MI, USA July 2018
- e Socinfo 2017, Oxford, UK September 2017
- e IC2S2 2016, Evanston, IL, USA June 2016
- e MSLD 2016, Bloomington, IN, USA May 2016
- e NAACL 2015, Denver, CO, USA June 2015
- e ICWSM 2014, Ann Arbor, MI, USA June 2014
- e EACL 2014, Gothenburg, Sweden April 2014
- e NAACL 2013, Atlanta, GA, USA June 2013
- e CICLING 2011, Tokyo, Japan February 2011

**Participated In**  
(*inter alia*)

1. 2018 n2c2 Track 2 Summer 2018
  - Mentor: Dr V. G. Vinod Vydiswaran
  - Collaborators: Dr Tanmay Basu, Nabarup Maity, Farhan Siddiqui, Jinghui Liu, Xinyan Zhao, Hyeon Joo
2. SIGIR 2018 as volunteer Summer 2018
3. MIDAS (Michigan Institute for Data Science) Research Forum Fall 2017
4. a2-dlearn: The Ann Arbor Deep Learning Event Fall 2017
5. Ann Arbor Vedanta Symposium Fall 2017
6. 3rd Annual MIDAS (Michigan Institute for Data Science) Symposium Fall 2017
7. Vedanta Study Circle at the University of Michigan Fall 2016 - Summer 2018
8. PR-SOCO (Personality Recognition in Source Code) Summer-Fall 2016
  - Collaborators: Shanta Phani, Dr Arindam Biswas
9. a2-dlearn: The Ann Arbor Deep Learning Event Fall 2015
10. Ann Arbor/Detroit NLPers (A2D-NLP) Winter 2015 onwards
11. University of Michigan NLP Reading Group Fall 2014 onwards
12. Possession Expression Identification in Free Text Summer 2014

	<ul style="list-style-type: none"> <li>• Mentor: Dr Rada Mihalcea</li> <li>• Collaborator: Jon Juett</li> </ul>	
	13. University of Michigan Engineering Graduate Symposium	Fall 2013
	14. Native Language Identification Shared Task 2013 (a part of BEA 2013 Workshop in Atlanta, GA)	Spring 2013
	<ul style="list-style-type: none"> <li>• Mentor: Dr Rada Mihalcea</li> </ul>	
	15. Penn State College of Engineering Research Symposium (CERS 2011)	Spring 2011
	16. Kaggle Chess Ratings Contest	Fall 2010
	<ul style="list-style-type: none"> <li>• Mentor: Dr Dan Kifer</li> <li>• Collaborators: Jingchen Liu, Mu Qiao, Pradeep Teregowda</li> </ul>	
<b>Community Service</b> ( <i>inter alia</i> )	c Help with Graduate Admission at University of Michigan Computer Science and Engineering Department (Under Artificial Intelligence Discipline)	Winter 2017 Winter 2016 Winter 2015
<b>Student Award</b>	1. UMich Rackham Conference Travel Grant 2. UNT Toulouse Graduate School Travel Grant	Summer 2017 Summer 2013
<b>Academic Achievements</b>	1. Penn State Candidacy Exam – passed four courses (required only two) 2. All India Rank 130 in GATE (Graduate Aptitude Test in Engineering) 3. GRE English 750, Math 800 (both out of 800) 4. All India Rank 10 in ICFAI Entrance Exam 5. WBJEE Rank 89 in Engineering, 6 in Medical	2009 2008 2007 2004 2004
<b>Courses Taken at Michigan</b>	1. EECS 583 (Advanced Compilers) 2. EECS 592 (Advanced Artificial Intelligence) 3. EECS 595 (Natural Language Processing)	
<b>Courses Taken at UNT</b>	1. CSCE 5250 (Introduction to Game Programming) 2. CSCE 5290 (Natural Language Processing) 3. CSCE 6933.002 (Human-centered Computing) 4. CSCE 6933.002 (Learning from Social Networking)	

**Courses Taken  
at Penn State**

1. CSE 511 (Operating System Design)
2. CSE 520 (Science of Computer Programming)
3. CSE 530 (Fundamentals of Computer Architecture)
4. CSE 550 (Numerical Linear Algebra)
5. CSE 555 (Numerical Optimization Techniques)
6. CSE 565 (Algorithm Design and Analysis)
7. CSE 583 (Pattern Recognition)
8. CSE 597A (Learning Theory)
9. CSE 598A (Machine Learning)
10. CSE 598D (Bioinformatics II)
11. CSE 598F (Probability and Limit Theorems)
12. STAT 463 (Applied Time Series Analysis)
13. STAT 500 (Applied Statistics)
14. STAT 501 (Regression Methods)
15. STAT 505 (Applied Multivariate Statistical Analysis)
16. STAT 506 (Sampling Theory and Methods)
17. STAT 557 (Data Mining I)
18. IST 441 (Information Retrieval and Search Engines)
19. IST 597B (Analysis and Design of Serious Games)
20. APLNG 578 (Computational and Statistical Methods for Corpus Analysis)
21. PHYS 580 (Elements of Network Science and Its Applications)

**Copyright**

Shibulipi, a constructed writing system for Indic Languages. Copyright Registration  
Number: TXU001685803

**Computer  
Skills**

		<b>Last Exposure</b>	<b>Proficiency</b>
Languages:	Java	April 2015	Moderate
	C++	December 2014	Moderate
	Java Swing	May 2012	Moderate
	Windows C++	December 2012	Beginner
Scripting:	Shell Script, Awk, Python	Current	Good
	Perl	June 2015	Moderate
	Lua	August 2011	Beginner
Web Design:	HTML	Current	Good
	CSS	Current	Good
	PHP	Current	Good
	JSP	April 2012	Good
	Javascript	April 2014	Beginner
Packages:	R, Weka, Mallet, NLTK, igrph	Current	Good
	MATLAB	April 2015	Good
	Apache Solr, Apache Ant, Apache Maven	April 2015	Moderate
	Amazon Mechanical Turk Java API	March 2015	Moderate
	PyGame	October 2015	Moderate
	Minitab, SPSS, Indri	April 2012	Beginner

Software:	CRFTagger, Stanford Parser/NER	Current	Good
	SVMLight	April 2012	Moderate
	Visual Studio, DirectX	December 2012	Beginner
	LanguageTool	April 2012	Beginner
	LLVM	December 2015	Beginner
Repositories:	ShareLatex	Current	Good
	Bitbucket	December 2013	Good
	Subversion (SVN)	December 2012	Good
	GitHub	Current	Moderate
Platform:	Linux (Ubuntu 18.04 and lower)	Current	Good
	PC (Windows 8 and lower)	Current	Good

---

**Crowdsourcing** Amazon Mechanical Turk job requester and worker

**References** Provided on request

**Last Updated** November 1, 2020