Lakshika Balasuriya

M.Sc. Student in Computer Science – Kno.e.sis Center – Wright State University
377, Joshi Research Center, 3640 Colonel Glenn Hwy., Dayton, Ohio – 45435
+1-937-708-6516 | lakshika@knoesis.org | http://www.wright.edu/~balasuriya.3/

Research Interests

User Profile Identification, Emoji Analysis, Data/Text Mining & NLP, Machine/Deep Learning

Education

Present – January 2016
M.Sc. in Computer Science, **Wright State University**, Dayton, OH, USA
Thesis: "Finding Street Gang Members on Twitter: A Content-based Approach" | Advisors: Prof. Amit P. Sheth & Dr. Derek Doran
GPA: 3.93/4.0

December 2015 – September 2011
P.hD. in Computer Science, **Wright State University**, Dayton, OH, USA
Incomplete | Advisor: Prof. Guozhu Dong
GPA: 3.93/4.0

September 2009 – February 2005
B.Sc.(Sp)(Hons) in Information Technology *Second Upper Class Honours*
University of Moratuwa, Katubbedda, Sri Lanka
Thesis: "SATHKAARA – A Haematology Decision Supporting System"
GPA: 3.65/4.2

Research Internship Experience

August 2017 – May 2017
Applied Research Intern - Gracenote Inc., Emeryville, CA
Worked on creating a Named Entity Recognizer (NER) and an Entity Linker (EL) to identify and link Music Entities on the Web and in user generated text. DBpedia and Wikidata were used as resources to extract entity information for music artists, albums etc. A bi-directional LSTM model was trained using Keras/TensorFlow and Word Embeddings learned from a Wikipedia text corpus was used to improve the entity disambiguation task.

Graduate Research Experience

Present – January 2016
Research Assistant, Kno.e.sis Center – Wright State University

Finding Gang Members on Twitter

The goal of this project is to understand how street gang members (self-identified street gang members on Twitter) use social media. We have developed machine learning models to automatically identify street gang members' Twitter profiles using the content they share on social media (such as tweets and YouTube videos), profile descriptions, profile/cover images, and their emoji usage.

EmojiNet and Emoji Understanding

The goal of this project is to build tools and algorithms to improve machine understandability of emoji. We built the first machine-readable sense inventory for emoji called EmojiNet and currently working on emoji similarity and emoji sense disambiguation applications. Our datasets and REST APIs are available at http://emojinet.knoesis.org/

Context-Aware Harassment Detection on Social Media

The aim of this project is to develop comprehensive and reliable context-aware techniques to glean information about the people involved and their interconnected network of relationships to determine and evaluate potential harassment and harassers. My work focuses on automatically classifying aggressive tweets posted on Twitter.

December 2015 – September 2011

Contrast Pattern Mining Aided Clustering

I explored developing a contrast pattern-based clustering algorithm for Abbreviation Disambiguation in unlabelled clinical data.

Contrast Pattern Mining Aided Ontology Alignment

I explored using contrast pattern mining techniques to solve Ontology Alignment problem (focused on classes) in Linked Open Datasets.

Teaching and Software Development Experience

April 2016 – August 2012 | Graduate Teaching Assistant
Wright State University, Dayton, OH

Worked as a Teaching Assistant for CEG-2350 Operating Systems and Concepts course while teaching and grading all labs. The labs covered hands-on sessions on using Linux Operating System, Shell Scripting and Python programming.

July 2011 – November 2009 | Software Engineer
Lanka Communication Services (Pvt) Ltd, Colombo, Sri Lanka

Involved in project management, client communication, system analysis, design and development of web applications using PHP for online newspapers, and live election results portals for presidential and provincial elections of Sri Lanka.

April 2008 – October 2007 | Software Engineer Intern
Virtusa Corporation Pvt. Ltd, Colombo, Sri Lanka,

Aetna Pas Project: Was an active team member in Aetna Pas PMD team in migrating PMD tool from VB 6.0 to VB.NET. Imprima Project: Programmed extensively in SharePoint to successfully deliver the assigned modules.

Publicatons

- Sanjaya Wijeratne, Shreyansh Bhatt, **Lakshika Balasuriya**, Hussein S. Al-Olimat, Manas Gaur, Amir Hossein, Yazdavar, Amit Sheth. Feature Engineering for Twitter-based Applications. **Book Chapter** in Feature Engineering for Machine Learning and Data Analytics, Editors: Guozhu Dong and Huan Liu, CRC Press, USA, 2017 (In Press).
- Sanjaya Wijeratne, **Lakshika Balasuriya**, Amit Sheth, Derek Doran. A Semantics-Based Measure of Emoji Similarity. In 2017 IEEE/WIC/ACM International Conference on Web Intelligence **(Web Intelligence 2017)**. Leipzig, Germany; 2017.
- Sanjaya Wijeratne, **Lakshika Balasuriya**, Amit Sheth, Derek Doran. EmojiNet: An Open Service and API for Emoji Sense Discovery. In 11th International AAAI Conference on Web and Social Media **(ICWSM 2017)**. Montreal, Canada; 2017.
- **Lakshika Balasuriya**, Sanjaya Wijeratne, Derek Doran, Amit Sheth. Signals Revealing Street Gang Members on Twitter. In Workshop on Computational Approaches to Social Modeling (ChASM 2016) co-located with 8th International Conference on Social Informatics **(SocInfo 2016)**. Bellevue, WA, USA; 2016.
- Sanjaya Wijeratne, **Lakshika Balasuriya**, Amit Sheth, Derek Doran. EmojiNet: Building a Machine Readable Sense Inventory for Emoji. In 8th International Conference on Social Informatics **(SocInfo 2016)**. Bellevue, WA, USA; 2016.
- **Lakshika Balasuriya**, Sanjaya Wijeratne, Derek Doran, Amit Sheth. Finding Street Gang Members on Twitter. In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining **(ASONAM 2016)**. San Francisco, CA, USA; 2016
- Sanjaya Wijeratne, **Lakshika Balasuriya**, Derek Doran, Amit Sheth. Word Embeddings to Enhance Twitter Gang Member Profile Identification. In **IJCAI** Workshop on Semantic

Machine Learning **(SML 2016)**. New York City, NY: CEUR-WS; 2016.

- **Balasuriya L.**, Perera K., Bandara N., Perera S., Amarasena D., Dias D. : InforMate - A GIS for Diverse Mobile Devices, **eASiA 2009** conference, Sri Lanka.

Professional Activities and Services

- External Reviewer : WAIM(2013), KDD(2012)

Graduate Level Courses (selected)

- Data Mining
- Advanced Data Mining
- Machine Learning
- Deep Leaning (Coursera)
- Information Retrieval
- Semantic Web
- Web 3.0: Next Gen Web & Apps
- Distributed Computing
- Cloud Computing
- Programming Languages

Professional Certifications

- Sun Certified Programmer For The Java 2 Platform 1.4

Skills

Programming Languages:	Java
Scripting Languages:	Python, PHP, JSP, XML, JavaScript, HTML
Deep/Machine Learning Tools/Packages:	Keras, TensorFlow, Word2Vec, scikit-learn, Weka
Databases:	MySQL, MongoDB
Operating Systems:	Windows and Linux (Expert)
Cloud Computing:	Hadoop/MapReduce (Basic)

References

- Prof. Amit Sheth
  Director, Kno.e.sis Center,
  Wright State University,
  Dayton, OH, 45435 USA.
  amit@knoesis.org
  +1-937-775-5203
  http://knoesis.org/amit

- Dr. Derek Doran
  Director, Web & Complex Systems Lab,
  Kno.e.sis Center,
  Wright State University,
  Dayton, OH, 45435 USA.
  derek@knoesis.org
  +1-937-775-3728
  http://knoesis.org/people/derek/