Pavan Kapanipathi

1449 Newgate Ct Apt C
Fairborn, OH 45324

+1 713 435 9982 (phone)
pavan[at]knoesis.org
http://knoesis.org/researchers/pavan

Research Interests

- User Modeling; Personalization; Recommendation; Knowledge Graphs; Linked Data; Semantic Web; Text Mining; Information Retrieval; Applied Machine Learning; Social Data Analysis.

Education

- **Kno.e.sis Center, Wright State University**   Dayton, OH
  Doctor of Philosophy in Computer Science, Adviser: Dr. Amit P Sheth   2012 - Present
- **Kno.e.sis Center, Wright State University**   Dayton, OH
  Master of Science in Computer Science, Adviser: Dr. Amit P Sheth   2009 - 2012
- **Visvesvaraya Technological University**   Karnataka, India
  Bachelor of Engineering in Computer Science   2003 - 2007

Research and Industry Experience

- **Kno.e.sis Center, Wright State University**   Dayton, OH, USA
  Graduate Teaching/Research Assistant   January 2010 - Present

  - Thesis: Personalized and Adaptive Semantic Information Filtering for Social Media.

  - Led and contributed to multiple research projects, (1) Hazards SEES ▷ (2) Twarql (Open source) ▷, and (3) Twitris ▷.

  - Contributed to three award winning proposals from NSF and NIH.

  - Collaborating with co-Phd and Masters students on research on knowledge-graphs and social data. Research problems include: (1) Domain specific knowledge base generation; (2) Entity recommendation using hierarchical knowledge bases; and (3) Predicting location of a Twitter user. ▷

  - Research projects and proposals involved interdisciplinary, local and international collaborations such as Ohio State University, CITAR (Wright State University), Digital Enterprise Research Institute (now Insight, Galway), Fraunhofer IAIS (University of Bonn, Germany).

  - Teaching assistant for "Introduction to Java" for three semesters.

- **Samsung Research America**   San Jose, US
  Research Intern, Advanced Technology Labs   May 2014 - December 2014

  - Developed a semantic enrichment engine for trajectory data. The semantic enrichment engine harnessed relevant, open knowledge bases to enhance user experience (Patent pending).

  - Predicted location specific activities of interest from social data. The approach builds a probabilistic model leveraging Wikipedia as a source of background knowledge.

- **IBM TJ Watson Research Center**   New York, US
  Research Intern, Infosphere Streams   May 2013 - August 2013

  - Modelled Twitter users' preferences as a hierarchy, inferred from a knowledge graph (Hierarchical Interest Graph). The methodology utilizes an adaptation of spreading activation algorithm to score the concepts in the hierarchical interest graph that captures users' interests.

  - Prototyped a recommendation system that harnesses the hierarchical interest graph for recommending tweets to users. Evaluation of our approach against the state of the art tweet recommendation techniques such as Support Vector Machines, and Latent Dirichlet Allocation has shown superior performance. ▷

  - Internship resulted in a poster and research paper at top conferences.

- **Digital Enterprise Research Institute**   Galway, Ireland
  Research Intern, Social Software Unit   April 11 - August 11

  - Collaborated with Google for extending their PubSubHubbub protocol to a privacy aware Semantic Hub.

  - Semantic Hub disseminates information based on publisher/user's preferences. Semantic web technologies such as RDF, OWL, and SPARQL were utilized for determining and representing user preferences. This system was developed in Python requiring familiarity with Google app engine.

- Semantic Hub is used by (1) SMOB (an open source, distributed, semantic microblogging framework). (2) Personalized Filtering of social stream.
- Internship resulted in 4 publications at workshops and top conferences.

- **Accenture** — Bangalore, India
  Senior Programmer — *July 2007 - March 2009*
  - Primarily worked with two clients (1) Drugstore, and (2) SFR.
  - Drugstore.com: The objective of the project was to increase the performance of the web application running for drugstore.com. Eleven member team, we migrated the embedded SQL queries in the VC++ code to corresponding stored procedure calls.
  - SFR: Lead role in the development of new features and modules for CRM applications. Applications were developed in Java and required knowledge of IBM websphere (Application server) and Epiphany (CRM tool).

- **Robert Bosch** — Bangalore, India
  Project Trainee — *Feb 2007 - June 2007*
  - Developed a Test-program compiler in C#. The Test-program is used in the calibration process of fuel injection pumps.

Research Projects

- **Hazards SEES** ▷   Collaboration: Ohio State University

Infrastructure systems are a cornerstone of civilization. Damage to infrastructure from natural disasters such as an earthquake (e.g. Haiti, Japan), a hurricane (e.g. Katrina, Sandy) or a flood (e.g. Kashmir floods) can lead to significant economic loss and societal suffering. Human coordination and information exchange are at the center of damage control. This project seeks to radically reform decision support systems for managing rapidly changing disaster situations by the integrated exploitation of social, physical and hazard modeling capabilities.

I play the role of the lead co-ordinator for the project. I am contributing to design novel, multi-dimensional cross-modal aggregation and inference methods to compensate for the uneven coverage of sensing modalities across an affected region. I am also working on using machine learning with knowledge base driven techniques to continuously adapt the filters for crawling social data relevant to dynamically evolving events. Preliminary evaluation shows that our technique can improve the recall of the tweets being streamed by maintaining good precision relevant to the event of interest ▷.

Twitris+: 360 degree Social Media Analytics platform ▷

Social media has had unprecedented growth in the recent times. This gives opportunity to decision makers– from corporate analysts to coordinators during emergencies, to answer questions or take actions related to a broad variety of activities and situations: who should they really engage with, how to prioritize posts for actions in the voluminous data stream, what are the needs and who are the resource providers in emergency event, how is corporate brand performing, and does the customer support adequately serve the needs while managing corporate reputation etc. We demonstrate these capabilities using Twitris+ by multi-faceted anlaysis along dimensions of Spatio-Temporal-Thematic (STT), People-Content-Network (PCN), and Subjectivity: Emotion-Sentiment-Intent (ESI). Twitris' diversity and depth of analysis is unprecedented.

My work in this project involved porting the Twarql infrastructure to Twitris and allowing complex querying to answer interesting questions on Twitter data. The complex query answering enables SPARQL queries with DBpedia as the background knowledge on event relevant tweets.

Twarql ▷   Collaboration - DERI

Twitter has become a prominent medium to share opinions, observations and suggestions in real-time. Insights from these microposts ("Wisdom of the Crowd") has proved to be invaluable for businesses and researchers around the world. However, the microblog data published is increasing in numbers with the popularity and growth of Twitter. This has induced challenges in filtering these microblog data to cater the needs for aggregation and collective analysis for sense-making. Twarql addresses these challenges by leveraging Semantic Web technologies to enable a flexible query language for filtering microblog posts.

I was involved in designing and developing end to end Twarql system. The system includes (1) a pipeline of information extraction techniques on tweets to extract metadata, (2) an annotation module to transform the raw metadata to RDF, and (3) a flexible filtering and broadcasting component for users. Technologies used include Java, Storm, RDF, SPARQL, Hadoop, and PIG. We published two papers with one winning the Triplification challenge in 2010.

SMOB   Collaboration - DERI, Google

Users of traditional microblogging platforms such as Twitter face drawbacks in terms of (1) Privacy of status updates as a followee – reaching undesired people, and (2) Information overload as a follower – receiving uninteresting microposts from followees. In this project we have implemented a privacy-aware version of google's PuSH protocol (Semantic Hub) for distributed and user-controlled dissemination of microposts using SMOB (semantic microblogging framework). The approach leverages users' Social Graph to dynamically create group of followers who are eligible to receive micropost. The restrictions to create the groups are provided by the followee based on the hastags in the micropost. Both SMOB and Semantic Hub are available as open source.

Publications and Patents

- **Published Work**

    - Raghava Mutharaju, and **Pavan Kapanipathi**. Are We Really Standing on the Shoulders of Giants? 1st Workshop on Negative or Inconclusive Results in Semantic Web 2015, ESWC, 2015.

    - Siva Kumar Chekula, **Pavan Kapanipathi**, Derek Doran, Amit Sheth. Entity Recommendations Using Hierarchical Knowledge Bases. 4th International Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data, 2015.

    - **Pavan Kapanipathi, Revathy Krishnamurthy** (Joint first author), Amit Sheth, Krishnaprasad Thirunarayan. Knowledge Enabled Approach to Predict the Location of Twitter Users. In Extended Semantic Web Conference, 2015. **(acceptance rate 23%)**.

    - **Pavan Kapanipathi**, Prateek Jain, Chitra Venkataramani, Amit Sheth. User Interests Identification on Twitter Using a Hierarchical Knowledge Base. In Extended Semantic Web Conference 2014, Crete Greece. **(acceptance rate 23%)**

    - **Pavan Kapanipathi**, Prateek Jain, Chitra Venkataramani, Amit Sheth. Hierarchical Interest Graph from Twitter. 23rd International conference on World Wide Web companion 2014 (WWW companion 2014), Seoul, South Korea.

    - Fabrizio Orlandi, **Pavan Kapanipathi**, Alexandre Passant, Amit Sheth. Characterising concepts of interest leveraging Linked Data and the Social Web. The 2013 IEEE/WIC/ACM International Conference on Web Intelligence, Atlanta, USA, United States, 2013.

    - **Pavan Kapanipathi**, Fabrizio Orlandi, Amit Sheth, Alexandre Passant. Personalized Filtering of the Twitter Stream. 2nd workshop on Semantic Personalized Information Management at ISWC 2011, September 2011.

    - **Pavan Kapanipathi**, Julia Anaya, Amit Sheth, Brett Slatkin, Alexandre Passant. Privacy-Aware and Scalable Content Dissemination in Distributed Social Network. 10th International Semantic Web Conference 2011, Bonn, Germany, September 2011. **(acceptance rate 22%)**

    - **Pavan Kapanipathi**, Julia Anaya, Alexandre Passant . SemPuSH: Privacy-Aware and Scalable Broadcasting for Semantic Microblogging. 10th International Semantic Web Conference 2011, Bonn, Germany, September 2011.

    - Alexandre Passant, Julia Anaya, Owen Sacco, **Pavan Kapanipathi**. SMOB: The Best of Both Worlds. Federated Social Web Europe Conference, Berlin, June 3rd -5th 2011.

    - Alexandre Passant, Owen Sacco, Julia Anaya, **Pavan Kapanipathi**. Privacy-By-Design in Federated Social Web Applications, Websci 2011, Koblenz, Germany June 14-17, 2011.

    - Pablo Mendes, **Pavan Kapanipathi**, Alexandre Passant. Twarql: Tapping into the Wisdom of the Crowd. Triplification Challenge 2010 at 6th International Conference on Semantic Systems (I-SEMANTICS), Graz, Austria, 1-3 September 2010. **(Winner of Triplification Challenge 2010)**

    - Pablo Mendes, Alexandre Passant, **Pavan Kapanipathi**, Amit Sheth. Linked Open Social Signals.WI2010 IEEE/WIC/ACM International Conference on Web Intelligence (WI-10), Toronto, Canada, Aug. 31 to Sep. 3, 2010. **(16.6% acceptance)**

    - Pablo Mendes, **Pavan Kapanipathi**, Delroy Cameron, Amit Sheth. Dynamic Associative Relationships on the Linked Open Data Web. In Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, April 26-27th, 2010, Raleigh, NC: US.

- **Work in Progress**

  - **Pavan Kapanipathi**, Krishnaprasad Thirunarayan, Fabrizio Orlandi, Amit Sheth, Pascal Hitzler. A Real-Time #approach for Continuous Crawling of Events on Twitter by Leveraging Wikipedia. Technical Report.
  - **Pavan Kapanipathi**, Siva Kumar, Derek Doran, Prateek Jain, Chitra Venkataramani, Amit Sheth. Hierarchical Knowledge Base enabled Twitter User Modeling and Recommendation. (Journal).
  - Siva Kumar, **Pavan Kapanipathi**, Derek Doran, Prateek Jain, Amit Sheth. Exploring Taxonomical Interests for Entity Recommendations. Technical report, 2015.
  - Sarasi Sarangi, **Pavan Kapanipathi**, Amit Sheth. Domain-specific Sub graph Generation. Technical report, 2015.

- **Patents**

  - Joakim Soderberg, Edwin A. Heredia, **Pavan Kapanipathi**, Glenn Algie, Alan Messer. Semantic Enrichment of Trajectory Data. (Pending)

Research Grants and Proposals

- NIH-R01 - (3 years, $1.5Million) *significant participation*
  - Title: Trending: Social media analysis to monitor cannabis and synthetic cannabinoid use.
  - Awardees: Wright State University (Kno.e.sis, CITAR), University of Massachusetts - Amherst
  - Project: EdrugTrends

- NSF - (4 years, $2.0Million) *limited participation*
  - Title: Social and Physical Sensing Enabled Decision Support.
  - Awardees: Wright State University (Kno.e.sis), Ohio State University
  - Project: Hazards SEES

- NSF - (3 years, $200.000) *limited participation*
  - Title: Market Driven Innovations and Scaling up of Twitris.
  - Awardees: Wright State University (Kno.e.sis)
  - Project: Twitris Commercialization

Skills

Programming: Java, Python, R, SQL, PIG, SPARQL.
Tools & Softwares: Database (MySQL), Graphs (Virtuoso, Jena, Gephi), Machine Learning (Weka, scikit-learn), Natural Language Processing (Stanford CoreNLP), Version Management (SVN, GIT), Big Data (Hadoop, Storm).
Open Source Contributions: Twarql, SMOB.

Professional Activities and Achievements

- Invited to present on Semantic Filtering at the Big Data tutorial at Wright State University. ▷
- Invited talk at EMC CTO Office, Bangalore. (as a part of Invited Speaker Series Lecture). ▷
- Invited to present my research for Wright State University, Graduate School External Advisory Board. ▷
- Invited talk at Frontiers of Cloud Computing and Big Data Workshop, 2014, IBM TJ Watson Research Center. ▷
- Awarded scholarship of $1500 to attend the 10th International Semantic Web Conference at Bonn, Germany.
- Winner of the Triplification Challenge: "Twarql: Tapping into the Wisdom of the Crowd" won the Triplification Challenge in the open track at 6th International Conference on Semantic Systems in September 2010. ▷
- External Reviewer: AAAI2016, ICWSM2015, WWW2015, AAAI2013, ESWC2013, WWW2013, ISWC2011, IJSWIS.
- PC Member: ESWC2016, ICWS2015, HT2015, ESWC2015, ESWC2014, WI2014, WoLE2013(WWW Workshop), WI2013, DERIVE2012 (ISWC Workshop).

References

Prof. Amit Sheth (Adviser)	Prof. T K Prasad	Prof. Pascal Hitzler	Dr. Chitra Venkataramani
Director	Professor	Director	Program Director
Kno.e.sis Center	Kno.e.sis Center	DaSe Lab	Cognitive Systems Performance Lab
Wright State University	Wright State University	Wright State University	IBM T.J. Watson Research Center
amit@knoesis.org	tkprasad@knoesis.org	pascal.hitzler@wright.edu	chitrav@us.ibm.com

More references available on request.