Avinesh P.V.S.

Gruberstrasse 14,
Darmstadt Kranichstein,
64289, Hesse, Germany

E-mail: avinesh.pvs@gmail.com
http://avineshpvs.com/
Contact: +49-15217841446

EDUCATION

⬦ **PhD, Informatics**   Oct 2015 - Current
UKP Lab, Technische Universität Darmstadt, Germany
Supervisors: Dr. Christian M. Meyer, Prof. Iryna Gurevych

⬦ **MS by Research, Computer Science**   July 2008 - July 2011
LTRC, IIIT Hyderabad, India
Supervisors: Prof. Rajeev Sangal, Prof. Dipti Misra Sharma

⬦ **B.Tech with Honours, Computer Science**   Jul 2004 - Jul 2008
IIIT Hyderabad, India   CGPA: 8.54/10

RESEARCH
INTERESTS

⬦ NLP, ML, Lexical Acquisition, MT, Sentiment Analysis, Data Mining, IE & IR.

STRENGTHS

⬦ Ability to adapt and learn efficiently in different scenarios.

⬦ Strong research background and experience in solving practical problems in NLP and ML.

WORK
EXPERIENCE

⬦ **NLP & ML Developer**, IBM Watson Group (India)   (September 2013 - Feb 2015)
I worked on NLP Modeling team of the Watson Oncology Expert Advisor (OEA), the
Watson solution for the MD Anderson Cancer Center, the worlds leading cancer hospital.
We mined concepts from unstructured text in the medical domain and enhanced lexicon
using machine learning models as well as rule-based grammar for new concepts. As a part
of Watson services team, we built innovative products namely Watson-I and extraction of
embedded tables from PDF documents.

⬦ **Research Engineer**, Lexical Computing Limited (LCL)   (August 2010 - August 2013)
LCL is a research company which works at the intersection of the corpus and computa-
tional linguistics. The majority of tasks revolved around building web corpora for various
domains and languages. We collaborated in PRESEMT, a European funded MT project on
Comparable Corpora BootCat and Hierarchy of Domains. We also worked on TEDDCLOG
which is an automatic gap-fill question generation system for language teaching, under the
supervision of Dr. Simon Smith.

⬦ **Research Assistant**, LTRC, IIIT-Hyderabad (India)   (May 2006 - July 2011)
LTRC is one of the top NLP research institutes in India. I have been acquainted with
multiple areas of NLP and ML while working with different research groups. My work
at LTRC involved MT, tools for less privileged languages, dialog systems, part-of-speech
tagging and chunking for Indian languages. Some of my major contributions to LTRC
involves building a shallow parser for 9 Indian languages and playing a key role in SAMPARK
a Govt. of India project for Indian language MT.

⬦ **Teaching Assistant**, Artificial Intelligence, IIIT-Hyderabad (India) (Dec 2006 - April 2007)
Assist the Professor in managing a class of 180 students and plan tutorial & assignments
for the students. Correct the answer sheets, assignments and perform other tasks related to
the course website.

PUBLICATIONS **Avinesh P.V.S** and Christian M. Meyer. ***Joint Optimization of User-desired Content in Multi-document Summaries by Learning from User Feedback*** *in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017). Vancouver, Canada*

Ankur Parikh, ***Avinesh P.V.S****, Joy Mustafi, Lalit Agarwalla and Ashish Mungi.****ThinkMiners: Disorder Recognition using Conditional Random Fields and Distributional Semantics*** *in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland*

Avinesh P.V.S*, Diana McCarthy, Dominic Glennon, Jan Pomikálek.* ***Domain Specific Corpora from the Web*** *in Proceedings of the 15th EURALEX 2012 International Congress, Norway*

Adam Kilgarriff, ***Avinesh P.V.S****, Jan Pomikálek.* ***BootCatting Comparable Corpora*** *in Proceedings of 9th International Conference on Terminology and Artificial Intelligence, 2012, France*

Adam Kilgarriff, ***Avinesh P.V.S****, Jan Pomikálek.* ***Comparable Corpora BootCaT*** *in Proceedings of eLEX 2011, Slovenia*

Aswarth Dara, Prashanth Mannem, Hemanth Sagar Bayyarapu and ***Avinesh P.V.S****.* ***Transferring Syntactic Relations from English to Hindi Using Alignments on Local Word Groups*** *in Proceedings of IJCNLP-2011, Thailand*

Simon Smith, ***Avinesh P.V.S****, Adam Kilgarriff.* ***Gap-fill Tests for Language Learners: Corpus-Driven Item Generation*** *in Proceedings of ICON-2010: 8th International Conference on Natural Language Processing, Macmillan Publishers, India, 2010*

Avinesh P.V.S*, Ankur Parikh.* ***Phrase-Based Transliteration with Simple Heuristics*** *in Proceedings of the 2010 Named Entities Workshop, 48th ACL 2010, Sweden.*

Avinesh P.V.S*.* ***A Data Mining Approach to Learn Reorder Rules for SMT*** *in Proceedings of the NAACL-HLT 2010 Student Research Workshop, Los Angeles.*

Adam Kilgarriff, Siva Reddy, Jan Pomikálek, ***Avinesh P.V.S****.* ***A Corpus Factory For Many Languages,*** *in Proceedings of The Seventh International Conference on Language Resources and Evaluation (LREC), Malta, 2010.*

Prasanth Kolachina, Sriram Venkatapathy, Srinivas Bangalore, Sudheer Kolachina, ***Avinesh P.V.S****.* ***Phrase Based Decoding using a Discriminative Model,*** *in Proceedings of SSST-4, Fourth Workshop on Syntax and Structure in Statistical Translation, COLING, Beijing, August 2010.*

Taraka Rama, Karthik Gali, ***Avinesh P.V.S****.* ***Does Syntactic Knowledge help English-Hindi SMT?*** *in Proceedings of the NLP Tools contest, ICON 2008.*

Avinesh P.V.S*, Karthik G.* ***Part Of Speech Tagging Using Conditional Random Fields and Transformation Based Learning,*** *in Proceedings of Shallow Parsing for South Asian Languages (SPSAL) workshop, IJCAI 2007.*

PATENTS   Lalit Agarwalla, Ankur Parikh, **Avinesh P.V.S**, ***New Terms for Query Expansion using Unstructured Data***

Scott Carrier, Anil Omanwar, **Avinesh P.V.S**, Joy Bhakat,***Automated Correction of Natural Language Processing Systems***

Chinnappa Guggilla, **Avinesh P.V.S**, Prabhakar Majjiga, Praveen Kumar, ***Answering Natural Language Table Queries through Semantic Table Representation***

RESEARCH
PROJECTS   ◇ **NLP modelling, Watson Solutions:** Mine concepts from unstructured text in medical domain. Develop and enhance lexicon using machine learning models as well as rule-based grammar for new concepts.
Team Leader: Scott Carrier (IBM Watson)

⋄ **Watson-I - Information Extraction from Statistical Diagrams:** A solution for automatic data interpretation and answering analytical questions from tables, charts and statistical diagrams.
Team Coordinator: Joy Mustafi (IBM Watson)

⋄ **Extracting embedded tables from PDF:** As a part of Watson solution, Watson Engagement Advisors initial task is to process the unstructured text from documents. Our task was to extract embedded (without explicit borders) tables from the PDF documents.
Team Manager: Shreesha Kadambi (IBM Watson)

⋄ **TEDDCLOG: Testing Data Driven CLOze Generation:** Gap-Fill exercises have an important role in language teaching. Teddclog takes "a key" as input and automatically finds the distractors and a carrier sentence from the corpus.
Adviser: Prof. Simon Smith, Dr. Adam Kilgarriff

⋄ **Domain Specific corpora from the Web:** Domain specific corpora is needed for various NLP tasks. Here we present a BootCaT approach to collect large Domain specific corpora from the Web.
Adviser: Dr. Adam Kilgarriff, Dr. Jan Pomikàlek.

⋄ **CCBC: Comparable Corpora BootCaT:** The task here is to collect the comparable corpora i.e. a matching corpus in two languages in parallel using seed translations.
Adviser: Dr. Adam Kilgarriff, Dr. Jan Pomikàlek.

⋄ **Classifier for American/British English:** Classification of documents into American and British English using Support Vector Machine (SVM).
Advisors: Dr. Adam Kilgarriff

⋄ **Learning Reordering Rules In SMT:** Reordering plays pivotal role in MT. A learning algorithm named association rule mining, a popular data mining concept is used for learning the reorderings between source and target languages.
Adviser: Prof. Rajeev Sangal, Prof. Dipti Misra Sharma

⋄ **Shallow Parser for Indian Languages:** Shallow parser is the analysis of a sentence which identifies various constituents at the level of local word groups. This system was built for 9 Indian Languages as a part of SAMPARK (ILMT system).
Advisors: Prof. Rajeev Sangal, Prof. Dipti Misra Sharma

⋄ **Transfer Grammar Engine (TGE):** TGE captures the structural transfer between source and target language in MT.
Advisors: Prof. Rajeev Sangal, Prof. Dipti Misra Sharma

⋄ **HMM-based Tagging Engine:** This project involves implementation of a language independent Hidden Markov Model (HMM) based POS tagger. The core components of the project are smoothing algorithms and efficient data training models.
Advisors: Prof. Rajeev Sangal

⋄ **Sentiment Analysis Using NLP:** Analyzing opinions associated with an entity based on the polarities namely positive, negative and neutral.
Advisors: Kalyan Vasanth, Venkat Ramna (Valuepitch)

ADVANCED COURSES TAKEN ⋄ ML, NLP applications, PR, CL, IE & IR, AI, Operations Research, Linear Programming.

SKILLS ⋄ Python, Java, MySQL, NLTK, Scikit-learn, Tensorflow, RapidMiner, WEKA, OpenNLP.

ACHIEVEMENTS ⋄ Working prototype demo in ConnectIN 2014, at the Made in IBM Labs.

⋄ Deans merit list for the academic years 2005-2006, 2006-2007.

⋄ Our team secured 1st & 3rd place in SPSAL workshop IJCAI, 2007 & NLPAI, NLWAI workshop at Juhu, Mumbai, 2006 respectively.

REFERENCES Available upon request