Manikya Swathi Vallabhajosyula

https://orcid.org/0000-0002-9094-1722

1466 Michigan Ave, Apt 5, Columbus, Ohio – 43201

Email ID:

vallabhajosyula.2@buckeyemail.osu.edu

swathi.vallabhajosyula@gmail.com

Ph No: +1 (480)-553-0455

Summary

I am a 3rd year Computer Science Ph.D. graduate at The Ohio State University with 4 years of professional experience in developing and testing web and server applications. I am working as a Graduate Research Assistant with Dr. Rajiv Ramnath at The Ohio state University on an NSF-EAGER (Early-concept Grants for Exploratory Research) project – *Bridging the Last Mile* which focuses on building an artificially intelligent system to assist researchers to carry their experiments productively.

Education History

Doctor of Philosophy (Computer Science and Engineering)   Aug, 2019 - Present

The Ohio State University (Computer Science and Engineering)   GPA: 3.95/4.0

Columbus, OH

Master of Science (Computer Science and Engineering)   Aug, 2016 - Jul, 2018

University of Minnesota   GPA: 3.94/4.0

Duluth, MN

Bachelor's in technology (Computer Science and Engineering)   Sept, 2009 - May, 2013

Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering &Technology (VNRVJIET)   GPA: 3.90/4.0

Affiliated to *Jawaharlal Nehru Technological University Hyderabad (JNTU-Hyd)*

Hyderabad, Telangana, India

Work History

Research Intern   June, 2021 – Aug, 2021

Oak Ridge National Laboratory (remote)

Web Application Developer   Aug, 2018 – July, 2019

Spherexx.com

Tulsa, OK

IT Intern   May, 2017 – July, 2017

CHART Industries,

Tulsa, OK

Software Engineer   June, 2013 – July, 2016

CA Technologies (or Computer Associates International, Inc.).

Hyderabad, Telangana, India

Academic Coursework and Projects

Key-Courses: *Natural Language Processing (Introduction and Advanced), Intermediate Studies in Artificial Intelligence, Advance Artificial Intelligence, Computational Linguistics (Introduction and Advanced), Software Engineering, Algorithms, Operating Systems, Database Management System, Programming Languages and Programming using JAVA, PERL, Python, C/C++, and/or C# to develop stand-alone or web full stack applications.*

EAGER Bridging the Last Mile - Project Description   Aug'19-present

Relevant Coursework: *Natural Language Processing (regular expression parsing, part-of-speech tagging, dependency parsing), Artificial Intelligence (statistical models to build baselines for task specific classifiers and regressors), Operating Systems (batch processing, hyper-parameters settings for performance optimizations), Research project capstone (application of research methods and best practices in research; teamwork, written and oral communication)*

With the onset of any research/project, most of the time is spent on data, pre-processing and tool exploration. Some preliminary analysis along with small-scale computations is needed to compare the tools' results while adjusting relevant software parameters and modal parameters. During this entire process, one has to explore various resources or seek ad-hoc advice from colleagues, collaborators, advisors. This project proposes the use of artificial intelligence to build a cyberinfrastructure tool that assists by utilizing past experiences and other resources to cater to individual researcher's needs.

- **Building Knowledge Triples on System Trace**

     With the availability of such multi-format/multi-dimensional data, we need to build a knowledge base which could represent the information in one specific format. Formatting the information to a standard triple syntax in machine processing. The

purpose of this work is to create a knowledge base by implementing a compatible processing pipeline for pre-processing the data, identifying log structures and assigning appropriate pre-defined class labels,

- **Synthesizing Genome Sequencing Data for Machine Learning**
Genome sequencing pipeline consists of identifying an appropriate genome with sufficient coverage and processing it for quality by a pre-processing and assembling it. The tools in the pipeline need to be hyperparameter tuned with respect to the input genome and Machine Learning (ML) algorithms could be trained to predict these settings. There isn't sufficient training data to predict these hyperparameters, so synthesizing multiple genomes from a single genome can help detect appropriate features for several ML algorithms. The synthesized training data set is manually and programmatically validated against the original sequencing data. The algorithm that fits the accuracy of synthesized data against the original data could be used for predicting the hyperparameters for a target sequencing genome.

- **Information Extraction from Domain Specific (Genome Sequencing) Scientific Documents**
In recent years there has been a huge growth in the amount of scientific literature that is being published. It has also been found that the amount of research almost doubles in every ten to fifteen years. Around 2.2 million new scientific documents were published in 2016 alone. With this rapid growth it becomes increasingly difficult for researchers to do a literature review in their field of study. This work proposes to develop a system that extracts useful information from Gene sequencing scientific documents and then structures information like the research problem addressed, the gene worked on, the tools and settings used and results from these documents in the form of a knowledge graph which can then be queried by researchers. Thus, instead of reading entire documents the researchers can then query the Knowledge graph and learn about recent developments in the field of Gene sequencing.

- **Predicting optimal settings to run applications on High Performance Computing (HPC) systems for optimal resource utilization**
Applications like genome sequencing or mapping climate patterns require heavy processing speed and memory to obtain results within practical windows. These applications run on shared HPC systems and thus need appropriate resource allocations to run the job till completion with maximum allocated resource utilization. It might take a few runs to identify these settings and if the target environment changes, these settings change and additional runs are required to re-tune these settings. This process of identifying optimal settings is both recourse and time intensive. In this project, we train an ML model on sample training data – comprising of real or scaled down synthetic data sets generated by simulating a few low resource consumption runs – and predict on real time execution data.

Discovering Hypernym for a New Sense in WordNet (Natural Language Processing -NLP)   Aug'17-May'19

Relevant Coursework: *Natural Language Processing (regular expression parsing, word vectors, part-of-speech tagging, dependency parsing, statistical alanyls of data distribution to obtain Hearst Patterns, word similarity analysis with vector and WordNet representations)*

WordNet is a free and open-source lexical database used in many NLP applications like machine translation (translating sentences from one language to another). WordNet was last updated in the year 2006 and since many new words are created and hence these applications would generate inaccurate results. Hence updating WordNet automatically without manual intervention is important. [As part of **SemEval 2016** – Task 14 and **SemEval 2018** – Task 9]

- My research uses a well-established pattern recognition algorithm **(Hearst Patterns & regular expressions)** and the state-of-art **vector space models (word2vec)** to predict an apt location to insert the new word-sense into the WordNet.
- Published a paper on this research in **SemEval@NAACL-HLT 2018**.

Papers Published:

Hassan, A. Z., Vallabhajosyula, M. S., & Pedersen, T. (2018). Umduluth-cs8761 at semeval-2018 task 9: Hypernym discovery using hearst patterns, co-occurrence frequencies and word embeddings. *arXiv preprint arXiv:1805.10271*.

Industrial Projects

Web Application Developer   (Aug'18 – Aug'19)
Spherexx.com, TULSA-OK

- Developing a **full-stack application** – a voice bot - which helps creating new-leads alongside booking appointments and providing information about the requested property – in lease management system.
- The application is developed using **C#**, **.NET MVC** Framework and a backend **python (NumPy, Pandas)** classifier (**Machine Learning-Scikit learn**).

Teaching Assistant   (Aug'19 – Dec'19)
The Ohio State University, Columbus.

- Assisted the lead professor in **conducting** regular classes, examinations and class exercises. **Conducting** lab sessions, discussions and **evaluating** the progress of students in their projects, exercises and examinations.
- Was a TA for **CSE 1223 Introduction to programming in JAVA.**

Teaching Assistant (Aug'16 – May'18)
University of Minnesota, Duluth.
- Assisted the lead professor in **conducting** regular classes, examinations and class exercises. **Conducting** lab sessions, discussions and **evaluating** the progress of students in their projects, exercises and examinations.
- **Tutored** the students in **computer programming** using **C++, HTML, CSS, JavaScript and Natural Language Processing.**

IT Intern (May'17 – July'17)
CHART Industries, TULSA-OK
- **Analyzing and fixing bugs in the in-house software application called CACHE build using ASP.NET.** CACHE is a server based application which automates majority of the work needed to produce a quotation for the company's products.

Software Engineer (June'13 – July'16)
CA Technologies (or Computer Associates International, Inc.).
During my tenure with CA, I have worked on various Products in different roles during their product cycles.
- Worked as a **developer** to build shippable deployments using **JAVA** and **Install anywhere** for various products.
- Along with the **Quality Assurance Engineers** designed implement automation suits using **Selenium** to execute the test cycles of each build for the products in a jiff.
- Was responsible for unit testing the code using **TestNG** framework, maintaining the integrity and secrecy of code using **CA Harvest** Tool and worked on tools like **LoadRunner** to test and improve the performance of web based products by fine tuning the product, database and operating system parameters.
- Tested the **compatibility** of the product on databases like **Oracle**, **DB2**, **MySQL** and Operating systems like **Windows**, **UNIX**, **AIX** in both standalone and distributed modes.
- Gained experience in product development life cycle by working alongside Senior Software Engineers and Architects in various phases of Product Development Life Cycle (**AGILE-SCRUM**) such as **building product backlog**, **planning** the product development, **testing** (performance, load and alpha) cycles, executing the cycles in **sprints**, **customer demonstrations** and **training** the services employees.

Software Proficiency

Programming/testing	:	C, C++, C#, Core JAVA, J2EE, TestNG, Selenium, IA custom code, Batch, Shell, Perl, Python.
Web Development	:	HTML 5 with CSS, JavaScript.ASP.NET
Servers and Frameworks	:	Tomcat, JBOSS, Product Automation Framework (CA Internal Framework), Selenium Webdriver
Operating Systems	:	Windows Servers, UNIX, AIX
Tools and IDEs	:	Install Anywhere, HP Application Lifecycle Management (ALM), HP Load Runner & Performance Center (PC), Rally, VersionOne, Microsoft Office, Harvest, Embedded Entitlement Manager (CA), Salesforce Charter, Maven. Eclipse, Rational Rose, Rest Client, Selenium IDE with FireBug
Databases	:	MySQL, Oracle 11g,12c, DB2 LUW and DB2 z/os, MS-SQL
Application Life Cycle	:	Agile Scrum, Scrumban, Waterfall

Awards & Certifications

Professional Awards	:	Above & Beyond September 2014 @ CA Technologies, Above & Beyond March 2016 @ CA Technologies
Academic Awards	:	Received Assistantship with Tuition Fee waiver from University of Minnesota, Third best Academic Project 2013
Soft Skill Certifications	:	Leaders at All Levels @ **CA Technologies**, JAWS