

## Summary

I am a 4th-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) *Bridging the Last Mile* and *ICICLE AI Institute* (ICICLE: Intelligent CI with Computational Learning in the Environment) projects – focusing on building artificially intelligent systems to assist researchers in carrying their experiments productively.

## Education History

<b>Doctor of Philosophy</b> (Computer Science and Engineering) The Ohio State University (Computer Science and Engineering) Columbus, OH	Aug, 2019 - Present GPA: 3.93/4.0
<b>Master of Science</b> (Computer Science and Engineering) The University of Minnesota Duluth, MN	Aug, 2016 - Jul, 2018 GPA: 3.95/4.0
<b>Bachelor's in Technology</b> (Computer Science and Engineering) Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering & Technology (VNRVJiet) Affiliated to <i>Jawaharlal Nehru Technological University Hyderabad (JNTU-Hyd)</i> Hyderabad, Telangana, India	Sept, 2009 - May, 2013 GPA: 3.90/4.0

## Work History

<b>Teaching Research Fellow (Summer India)</b> Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering & Technology (VNRVJiet)	May, 2022 – July, 2022
<b>Research Intern</b> Oak Ridge National Laboratory (remote)	June, 2021 – Aug, 2021
<b>Web Application Developer</b> Spherexx.com Tulsa, OK	Aug, 2018 – July, 2019
<b>IT Intern</b> CHART Industries, Tulsa, OK	May, 2017 – July, 2017
<b>Software Engineer</b> CA Technologies (or Computer Associates International, Inc.). Hyderabad, Telangana, India	June, 2013 – July, 2016

## 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**

The genome sequencing pipeline consists of identifying an appropriate genome with sufficient coverage and processing it for quality by 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 fitting the synthesized data's accuracy against the original data could be used for predicting the hyperparameters for a target sequencing genome.

- **Information Extraction from Domain-Specific 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 research almost doubles 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 as the research problem is addressed. The gene worked on, the tools and settings used, and results from these documents in the form of a knowledge graph which researchers can then query. Thus, instead of reading entire documents, the researchers can query the Knowledge graph and learn about recent developments in Gene sequencing.

- **Predicting resource allocations (walltime) 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 real or scaled-down synthetic data sets generated by simulating a few low-resource consumptions runs – and predict 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 analysis 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 2006, and since many new words are created, 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*.
- Vallabhajosyula, Manikya Swathi, and Rajiv Ramnath. "Towards Practical, Generalizable Machine-Learning Training Pipelines to build Regression Models for Predicting Application Resource Needs on HPC Systems." *Practice and Experience in Advanced Research Computing*. 2022. 1-5.
- S. Vallabhajosyula and R. Ramnath, "Establishing a Generalizable Framework for Generating Cost-Aware Training Data and Building Unique Context-Aware Walltime Prediction Regression Models," 2022 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Melbourne, Australia, 2022, pp. 497-506, doi: 10.1109/ISPA-BDCloud-SocialCom-SustainCom57177.2022.00070.
- Vallabhajosyula, Manikya Swathi, and Rajiv Ramnath. " Insights from the HARP Framework: Using an AI-Driven Approach for Efficient Resource Allocation in HPC Scientific Workflows"  
[http://camps.aptaracorp.com/ACM\\_PMS/PMS/ACM/PEARC23/70/3a84c88c-f988-11ed-b37c-16bb50361d1f/OUT/pearc23-70.html](http://camps.aptaracorp.com/ACM_PMS/PMS/ACM/PEARC23/70/3a84c88c-f988-11ed-b37c-16bb50361d1f/OUT/pearc23-70.html)

## MS Thesis/Project Reports:

1. (MS Thesis) "Hypernym Discovery over WordNet and English Corpora - using Hearst Patterns and Word Embeddings", S. Vallabhajosyula, T. Pederson <https://hdl.handle.net/11299/200144>
2. "Information Extraction from Gene Sequencing Scientific Documents", Chanana, R., co-advisors: Bryan Carstens, Jian Chen, May 2021. \*\*
3. "Generating Knowledge graphs on scientific execution workspace", Gulati, R., co-advisors: Bryan Carstens, Jian Chen, December 2020. \*\*
4. "HARP - HPC Application Runtime Predictor", Prasanna, Miriyala, R., co-advisors: Bryan Carstens, Jian Chen, December 2022. \*\*

(\*\* These are the MS projects I have worked with and provided guidance. Please request the CSE department at The Ohio State University to get access to these documents)

## POSTERS Presented:

1. Vallabhajosyula, Manikya Swathi (2018): Identifying Hypernym for a New Sense in WordNet. figshare. Poster. <https://doi.org/10.6084/m9.figshare.22565089.v1>
2. Vallabhajosyula, Manikya Swathi; Ramnath, Rajiv (2020): EAGER: Bridging the Last Mile. figshare. Poster. <https://doi.org/10.6084/m9.figshare.11777808.v3>
3. Vallabhajosyula, Manikya Swathi; Ramnath, Rajiv (2021): Modeling A Framework To Estimate Resource Requirements For Scientific Workflows. figshare. Poster. <https://doi.org/10.6084/m9.figshare.22363183.v1>
4. Chanana, Rishabh; Vallabhajosyula, Manikya Swathi; Ramnath, Rajiv (2021): Using synthesized data to train machine learning models used in genome engineering pipeline. figshare. Poster. <https://doi.org/10.6084/m9.figshare.22565113.v1>
5. Vallabhajosyula, Manikya Swathi; Ramnath, Rajiv (2022): Building an AI-powered Assistant for Computational Scientists. figshare. Poster. <https://doi.org/10.6084/m9.figshare.11777796.v1>
6. Vallabhajosyula, Manikya Swathi; Ramnath, Rajiv (2023): Modeling A Framework To Estimate Resource Requirements For Scientific Workflows. figshare. Poster. <https://doi.org/10.6084/m9.figshare.22363183.v1>
7. Vallabhajosyula, Manikya Swathi; Ramnath, Rajiv (2023): owards Characterizing DNNs to Estimate Training Time using HARP (HPC Application Resource (runtime) Predictor [http://camps.aptaracorp.com/ACM\\_PMS/PMS/ACM/PEARC23/82/138a150e-f8dd-11ed-b37c-16bb50361d1f/OUT/pearc23-82.html](http://camps.aptaracorp.com/ACM_PMS/PMS/ACM/PEARC23/82/138a150e-f8dd-11ed-b37c-16bb50361d1f/OUT/pearc23-82.html)

## Industrial Projects

### **Web Application Developer**

(Aug'18 – Aug'19)

#### **Spherexx.com, TULSA-OK**

- Developing a **full-stack application** – a voice bot - which helps create new leads, booking appointments, and provide information about the requested property – in the 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, and 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, and discussions and evaluating students' progress 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 that 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 and 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 working 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**, and **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

<b>Programming/testing</b>	:	C, C++, C#, Core JAVA, J2EE, TestNG, Selenium, IA custom code, Batch, Shell, Perl, Python.
<b>Web Development</b>	:	HTML 5 with CSS, JavaScript.ASP.NET
<b>Servers and Frameworks</b>	:	Tomcat, JBOSS, Product Automation Framework (CA Internal Framework), Selenium Webdriver
<b>Operating Systems</b>	:	Windows Servers, UNIX, AIX
<b>Tools and IDEs</b>	:	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
<b>Databases</b>	:	MySQL, Oracle 11g,12c, DB2 LUW and DB2 z/os, MS-SQL
<b>Application Life Cycle</b>	:	Agile Scrum, Scrumban, Waterfall

## Awards & Certifications

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