Sudha Rao

609-933-0409
raosudha@cs.umd.edu

Education

University of Maryland, College Park
PhD - Computer Science *2013 - 2018 (expected)*
Focus: Natural Language Processing
Advisors: Hal Daumé III, Philip Resnik

Veermata Jijabai Technological Institute Mumbai
Bachelor of Technology - Computer Engineering; CPI:9.4/10 *May 2011*

Publications

- **Sudha Rao**, Yogarshi Vyas, Hal Daumé III, and Philip Resnik. **Parser for Abstract Meaning Representation using Learning to Search** *Under submission*
- **Sudha Rao**, Allyson Ettinger, Hal Daumé III, and Philip Resnik, **Dialogue focus tracking for zero pronoun resolution**, NAACL 2015, Denver, Colorado, June 2015.
- Suzanne DSilva, Neha Joshi, **Sudha Rao**, Sangeetha Venkatraman, and Seema Shrawne. **Improved Algorithms for Document Classification & Query-based Multi-Document Summarization**, *International Journal of Engineering and Technology vol. 3, no. 4, pp. 404-409 , 2011*

Research Experience

- **Semantics for biology** *Summer 2015*
  - Worked on identifying protein interactions in biology texts using **Abstract Meaning Representation (AMR)** of sentences.
  - This work is a part of the ***Big Mechanism*** project and I am working on this with **Dr. Daniel Marcu** and **Dr. Kevin Knight** at **ISI, USC**.
- **Semantic Parsing** *Fall 2014*
  - Worked on automatically learning **Abstract Meaning Representation (AMR)** for English sentences using **SEARN** (Search based Structured Prediction) technique.
- **Machine Translation** *Spring 2013*
  - Worked on developing a **Sequence Labeling** model for restoring **dropped pronouns** when translating SMS text from Chinese to English.
  - This work was done as a part of **IBM's BOLT** project.
- **Query based Multi-document Summarization** *Spring 2011*
  - Analyzed and **improved a Hypergraph based technique** for summarization to make it less processor intensive.
  - Designed a simple algorithm for query based multi-document summarization using **k-means clustering** with results comparable to the Hypergraph based technique.

Work Experience

- **NVIDIA Graphics Pvt. Ltd.** Pune
  System Software Engineer *July 2011 – April 2013*
  - Designed and implemented a test infrastructure for Tegra Driver on varied mobile operating systems and platforms.
  - Worked on in-house mobile tools being developed as a part of *Planning, Infrastructure & Operations* team.
- **Microsoft India Development Center** Hyderabad
  Software Development Engineer - Intern *May 2010 – July 2010*
  - Worked on Data Protection Management.

Teaching Experience

CMSC 131 - Object Oriented Programming Fall 2013

Projects

- **Predicting clinical depression** *Spring 2013*
  - We used **Facebook status updates** to predict neuroticism among people.
  - This project was done as a part of the **Computational Linguistics** course.
- **Farmers Buddy** *Fall 2009*
  - We developed a first of its kind portal to be accessible via Internet allowing users, especially farmers, to obtain information regarding various activities involved in farming and providing a platform for interaction between the different users of the system.
  - This project was developed as part of **IBM's The Great Mind Challenge**, 2009, and was among the **Top 20** projects all over India.

Honours/Awards

- Scholarship to attend the **Grace Hopper Celebration** conference 2014 and 2015.
- Recipient of **Dean's Fellowship** at University of Maryland, College Park.
- Travel fund to attend **NIPS** conference 2013.
- Scholarship to attend the **Women in Theory** conference 2012 at Princeton University.
- Ranked **3rd** out of 70 students in the Computer Science department of VJTI.
- Recipient of scholarship award from **Sir Ratan Tata** Trust (2008, 2009 and 2010).
- **56th** rank in Maharashtra in MHT-CET examination held in May 2007 in Engineering Course. Score: **(194 / 200)**

Technical Skills

- **Languages:** Python, MATLAB, Java, C++, C#, SQL, Perl
- **Web Technologies:** J2EE(JSP, Servlet), HTML, CSS, Javascript, AJAX

- **Databases:** Microsoft SQL, IBM DB2
- **Operating Systems:** Linux, OS X, Microsoft Windows

Relevant coursework

- **Graduate**: Computational Linguistics 1 & 2, Multi-lingual NLP, Prediction: Brain vs Machines, Analysis of Algorithms.
- **Undergraduate**: Advanced Algorithms, Approximation algorithms, Intelligent systems.

UNOFFICIAL TRANSCRIPT
FOR ADVISING PURPOSES ONLY
As of: 07/23/15

Name: Rao, Sudha
E-Mail: raosudha@cs.umd.edu
Major: Computer Science
Graduate Graduate Degree Seeking

Transcripts received from the following institutions:

University of Mumbai on 12/12/12

Historic Course Information is listed in the order:
Course, Title, Grade, Credits Attempted, Earned and Quality Points

Fall 2013
MAJOR: COMPUTER SCIENCE COLLEGE: GRADUATE SCHOOL

CMSC651	ANALYSIS OF ALGORITHMS	A	3.00	3.00	12.00
CMSC723	COMPUTATIONAL LING I	A+	3.00	3.00	12.00
CMSC798E	GRAD SEM CMPTR SCI	A	1.00	1.00	4.00

Semester: Attempted 7.00; Earned 7.00; GPA 4.000
Grad Cumulative: 7.00; 7.00; 4.000

Spring 2014
MAJOR: COMPUTER SCIENCE COLLEGE: GRADUATE SCHOOL

CMSC773	COMPUTATIONAL LING II	A-	3.00	3.00	11.10
CMSC798F	GRAD SEM CMPTR SCI	A	1.00	1.00	4.00
CMSC828W	LINGUISTIC PREDICTION	A-	3.00	3.00	11.10

Semester: Attempted 7.00; Earned 7.00; GPA 3.742
Grad Cumulative: 14.00; 14.00; 3.871

Fall 2014
MAJOR: COMPUTER SCIENCE COLLEGE: GRADUATE SCHOOL

CMSC754	COMPUTATIONAL GEOMETRY	B+	3.00	3.00	9.90
LING848	SEM IN COMPUTAT LINGUIST	A	3.00	3.00	12.00

Semester: Attempted 6.00; Earned 6.00; GPA 3.650
Grad Cumulative: 20.00; 20.00; 3.805

Spring 2015
MAJOR: COMPUTER SCIENCE COLLEGE: GRADUATE SCHOOL

CMSC828I	MLTLNG NTRL LANG PRCSSNG	A	3.00	3.00	12.00
CMSC878O	SPARSITY & MCHN LRNG	A	3.00	3.00	12.00

Semester: Attempted 6.00; Earned 6.00; GPA 4.000

Grad Cumulative: 26.00; 26.00; 3.850

Grad Cumulative Credit: 26.00
Grad Cumulative GPA : 3.850

**** Current Course Information ****

1507

Course	Sec	Credits	Core /Div	Grd/ Meth	Drop /Add	Add Date	Drop Date	Modified Date
UNIV099	CP21	0.00		S-F	A	05/21/15		05/21/15

1508

Course	Sec	Credits	Core /Div	Grd/ Meth	Drop /Add	Add Date	Drop Date	Modified Date
CMSC660	0101	3.00		REG	AW	06/08/15		06/08/15
CMSC631	0101	3.00		REG	A	04/16/15		04/16/15
CMSC701	0101	3.00		REG	A	04/16/15		04/16/15

Telegram : "VIJETINST" Matunga, Mumbai 400019
Tel: +91-22-24198102-6 (5 lines)
website: http://www.vjti.ac.in/
FAX : 91- 022 -24152874

VEERMATA JIJABAI TECHNOLOGICAL INSTITUTE

[Central Technological Institute, Maharashtra State]
Matunga, Mumbai 400019

An autonomous Institute Affiliated to University of Mumbai

Ref: April 03, 2012

Official Transcript of Academic Record of

RAO SUDHA VASUDEVA

Reg No. **071071045**

Year of admission: 2007-2008
Admitted to: Bachelor of Technology in Computer Engineering
Duration of the Program: 8 Semesters (4 years)

Year of Graduation: 2010-2011
Cumulative Performance Index: **9.4**

- Veermata Jijabai Technological Institute (known as Victoria Jubilee Technical Institute prior to January 26, 1997) was established in 1887 and is one of the premier Institutes in India. The Institute conducts various diploma, post-diploma, graduate and post graduate programmes in the fields of Civil, Computer, Electrical, Electronics, Production, Mechanical, Textile and Information Technology and related fields.

- The admission to the diploma programmes are strictly on merit based on their Secondary School Certificate examination.
  The admissions to the under graduate degree programmes are strictly on merit based on the state level test conducted by the State Government of Maharashtra after their Higher Secondary Certificate examination.
  Candidates who complete their Diploma, a three year programme after Secondary School Certificate examination, are also admitted directly to the second year of the under graduate degree programmes strictly on merit.
  The admissions to the post graduate degree programmes are also done for few seats strictly on merit based on their Diploma Examination Secondary Certificate examination.
  The admissions to the post graduate programmes are strictly on merit based on GATE examination conducted by IIT.

- The Institute has been awarded status of Autonomous Institute since 2004-2005. The Institute is affiliated to University of Mumbai (known as University of Bombay prior to 1996). The Degree is conferred by the University of Mumbai.

- The medium of instruction at this institute is English.

Administrative Officer/Registrar

Telegram : "VIJETINST" Matunga, Mumbai 400019
Tel: +91-22-24198102-6 (5 lines)
website: http://www.vjti.ac.in/
FAX : 91- 022 -24152874

VEERMATA JIJABAI TECHNOLOGICAL INSTITUTE

[Central Technological Institute, Maharashtra State]
Matunga, Mumbai 400019

RAO SUDHA VASUDEVA **REG. NO. 071071045**

Semester-wise performance:

Semester	I	II	III	IV	V	VI	VII	VIII
CPI	9.2	9.5	9.6	9.5	9.5	9.4	9.4	9.4
SPI	9.22	9.77	9.67	9.51	9.24	8.80	9.40	9.58

Sem	Course Code	Course Title	L	P/T	Tot Al	Cr	Evaluation Weightages			ESE Hours	Obtained Grade	
							TWA	MST	ESE		Grade	Grade Points
I	100011	Engineering Chemistry I	2	1	3	5	10	15	75	2	AA	10
I	100021	Engineering Physics I	2	1	3	5	10	15	75	2	AB	9
I	100031	Engineering Mathematics I	3	2	5	8	10	15	75	3	AB	9
I	100040	Engineering Graphics	2	4	6	8	20	20	60	4	BB	8
I	100051	Elements of Engineering I	3	2	5	8	20	20	60	3	AA	10
I	100061	Computer Programming I	3	2	5	8	20	20	60	3	AA	10
I	100071	Workshop Practice I	-	3	3	3	100	-	-	-	BB	8
II	100012	Engineering Chemistry II	2	1	3	5	10	15	75	2	AA	10
II	100022	Engineering Physics II	2	1	3	5	10	15	75	2	AA	10
II	100032	Engineering Mathematics II	3	2	5	8	10	15	75	3	AA	10
II	100052	Elements of Engineering II	3	2	5	8	20	20	60	3	AB	9
II	100062	Computer Programming II	3	2	5	8	20	20	60	3	AA	10
II	100072	Workshop Practice II	-	3	3	3	100	-	-	-	AB	9
II	100080	Engineering Mechanics	4	2	6	10	20	20	60	3	AA	10
III	200033	Engineering Mathematics III	3	1	4	7	15	15	70	3	AA	10
III	307010	Electronic Devices and Circuits	3	2	5	8	15	15	70	3	AA	10
III	307020	Electrical Network	4	2	6	10	15	15	70	3	AA	10
III	307030	Data Structure	3	2	5	8	15	15	70	3	AB	9
III	307040	Digital Logic Design and Application	3	2	5	8	15	15	70	3	AB	9
III	307050	Discrete Structures	3	1	4	7	15	15	70	3	AA	10
IV	200034	Engineering Mathematics IV	3	1	4	7	15	15	70	3	AB	9
IV	307060	Principles of Analog and Digital Communication	3	2	5	8	15	15	70	3	AB	9
IV	307070	Computer Organization and Architecture	3	2	5	8	15	15	70	3	AB	9
IV	307080	Database Systems	4	2	6	10	15	15	70	4	AA	10
IV	307090	Analysis Of Algorithms	3	2	5	8	15	15	70	3	AA	10
IV	307100	Industrial Economics and Management	3	-	3	6	15	15	70	3	AA	10
V	200020	Technical Communication and Presentation Skills	2	2	4	6	15	15	70	3	BB	8
V	200040	Engineering Statistics	3	1	4	7	15	15	70	3	AA	10
V	307110	Object Oriented Analysis and Design	3	2	5	8	15	15	70	3	AA	10
V	307120	Computer Network	3	2	5	8	15	15	70	3	AB	9
V	307130	Microprocessors	4	2	6	10	15	15	70	3	AB	9
V	307140	Theoretical Computer Science	4	-	4	8	15	15	70	3	AB	9
V	307150	Computer Programming Lab	-	3	3	3	15	15	70	3	AA	10
VI	307160	System Programming	4	2	6	10	15	15	70	3	AB	9
VI	307170	Operating Systems with UNIX	4	2	6	10	15	15	70	3	AB	9
VI	307180	Web Technology	4	2	6	10	15	15	70	3	AA	10
VI	307190	Computer Graphics	4	2	6	10	15	15	70	3	BC	7
VI	307200	Advanced Database	4	2	6	10	15	15	70	3	AB	9

Administrative Officer/Registrar

Telegram : "VIJETINST" Matunga, Mumbai 400019
Tel: +91-22-24198102-6 (5 lines)
website: http://www.vjti.ac.in/
FAX : 91- 022 -24152874

VEERMATA JIJABAI TECHNOLOGICAL INSTITUTE

[Central Technological Institute, Maharashtra State]
Matunga, Mumbai 400019

RAO SUDHA VASUDEVA **REG. NO. 071071045**

VII	307210	Advanced Microprocessors	4	2	6	10	15	15	70	3	AB	9
VII	307220	Intelligent Systems	4	2	6	10	15	15	70	3	AA	10
VII	307230	Digital Signal Processing	4	2	6	10	15	15	70	3	AB	9
VII	307240	Software Engineering	4	2	6	10	15	15	70	3	AA	10
VII	407060	Elective - I: Advanced Computer Networks	4	2	6	10	15	15	70	3	AB	9
VIII	307250	System Security	4	2	6	10	15	15	70	3	BB	8
VIII	307260	Distributed Computing	4	2	6	10	15	15	70	3	AA	10
VIII	307270	Multimedia Systems	4	2	6	10	15	15	70	3	AA	10
VIII	407150	Elective – II: Advanced Algorithms	4	2	6	10	15	15	70	3	AA	10

Sem	Course Code	Course Title	L	P/T	Total	Cr	Evaluation Weightages			Obtained Grade	
							Report	Stage I	Stage II Seminar /Viva	Grade	Grade Points
VII - VIII	407900	Project	-	8	8	8	25	25	50	AA	10

1. Abbreviations:

L - one hour lectures per week	P/T - Practical / Tutorial hour per week
Cr - Credits	TWA - Term Work Assessment
MST - Mid Semester test	ESE- End Semester Examination
SPI – Semester Performance Index	CPI – Cumulative Performance Index

2. Credit System

Each course has a credit associated with it. One lecture hour carries 2 credits and two practical / tutorial carry 1 credit.

3. Performance Evaluation System

VJTI uses relative grading system. The grades are awarded on the basis of continuous assessment of the student. The grades awarded are as under:

AA	AB	BB	BC	CC	CD	DD	EE	FF
10.0	9.0	8.0	7.0	6.0	5.0	4.0	2.0	0.0
Outstanding	Excellent	Very Good	Good	Satisfactory	Average	Marginal	Unsatisfactory	Very Weak

Other grades like Au: Audit, PP: Pass, NP: Fail, DR: Dropped out, II: Incomplete due to non-appearance in ESE, RR: Fail due to lack of attendance are also used.

Most recent grade for the course shall be taken into account for the computation of SPI / CPI

Semester Performance Index (SPI) measures the performance of a student in a particular semester, which is the weighted average of the grades secured in all the courses taken in that semester

Example:
A student has registered for six courses in a Semester, out of which two are 8 credits, three 6 credits and one is 10 credit course i.e. a total of (2x8 + 3x6 + 10) 44 credits.
If the student secures BB, BC, AA, DD, BC, BB grades respectively in these courses, SPI will be calculated as;
SPI = (8x8 + 8x7 + 6x10 + 6x4 + 6x7 + 10x8) / 44 = 7.40

Cumulative Performance Index (CPI) is calculated at the end of every semester, taking into account the performance in all courses cleared by the student up to the semester for which the CPI is to be calculated.

A CPI of 6.5 and above may be considered as equivalent to First Class (60%) of similar other programmes of Mumbai University

Administrative Officer/Registrar

GRE® Examinee Score Report

Note: This report is not valid for transmission of scores to an institution.

All dates are formatted as MM/DD/YYYY.

Print Date: 11/28/2012

Print

Examinee Information

Name:	Sudha Rao
Address:	C/8 Shree Shivkirti CHS Ltd, Chincholi Bunder Rd, Malad West Mumbai, 400064 India
Email Address:	raosudha89@gmail.com
Phone Number:	8237209416
Date of Birth:	04/08/1989
Social Security Number:	
Gender:	Female
Intended Graduate Major Code:	0402
Intended Graduate Major:	Computer and Information Sciences - Computer Science
Most Recent Test Date:	10/22/2012
Registration Number:	4512878

General Test Scores

Test Date	Verbal Reasoning*				Quantitative Reasoning*				Analytical Writing	
	Prior Format		Current Format		Prior Format		Current Format			
	Scaled Score	Estimated Current Score	Scaled Score	% Below	Scaled Score	Estimated Current Score	Scaled Score	% Below	Score	% Below
10/22/2012			152	53			165	92	3.5	30

NS — No Score. Indicates that no questions were answered.
Only reported score will be available for display.
* The GRE Verbal Reasoning and Quantitative Reasoning score scales changed in August 2011. For tests taken August 2011 or later, scores are printed in the "Current Format" columns. For tests taken before August 2011, scores on the prior scales and the corresponding estimated scores on the current scales are printed in the "Prior Format" columns.

Subject Test Scores

Test Date	Test Name / Subscore Name	Scaled Score	% Below

Score Recipient(s)

Your score reporting history is shown below. "Pending" indicates your scores are not yet available, or your order has not yet been processed.

Undergraduate Institution				
Report Date	**Institution (Code)**	**Department (Code)**	**Test Type**	**Test Date**

Score Recipient(s)				
Report Date	**Institution or Fellowship Sponsor (Code)**	**Department (Code)**	**Test Type**	**Test Date**
Pending	Carnegie Mellon U (2074)	Computer Science (0402)	General Test	10/22/2012
10/31/2012	Georgia Inst Technology (5248)	Computer Science (0402)	General Test	10/22/2012

10/31/2012	U Illinois Urbana (1836)	Computer Science (0402)	General Test	10/22/2012
10/31/2012	U Washington (4854)	Computer Science (0402)	General Test	10/22/2012
10/31/2012	U MD Coll Park (5814)	Computer Science (0402)	General Test	10/22/2012

Requested scores are sent only if available on report date.
* Undergraduate Institution does not wish to receive scores
** Score recipient not valid/active

QUESTIONS ABOUT THIS GRE REPORT OF SCORES

Information to help you interpret your GRE scores is available at www.ets.org/gre/stupubs. If you have any questions concerning this GRE Report of Scores, call ETS at 1-609-771-7670 or 1-866-473-4373 (toll free for test takers in the U.S., U.S. Territories*, and Canada) between 8:00 a.m. and 7:45 p.m. EST or email gre-info@ets.org. **Includes American Samoa, Guam, Puerto Rico, and U.S. Virgin Islands*

SCORE REPORTING

Policies pertaining to score reporting and use are periodically reviewed and revised by the GRE Board. The policies and procedures explained in the 2012-13 *GRE Information and Registration Bulletin* are effective only for the time period of August 1, 2012, to June 30, 2013, and supersede previous policies and procedures in previous Bulletins. GRE scores are reportable for five (5) years following the testing year (July 1 to June 30) in which you tested. Currently, GRE scores earned after July 1, 2008, are available.

PERCENTILE RANK (% BELOW)

The percentile ranks in this report indicate the percentage of examinees who scored below your score. Note that these percentile ranks may be different from those that applied when the scores were originally reported to you if the scores were earned prior to July 2012. This reflects annual updating of these data to permit admissions officers to compare scores, whenever earned, with those for a recent reference group.

- Interpreting Your GRE Scores
- Order Additional Scores Reports
- Have you taken the GMAT®? Learn how you can participate in a special study. (PDF)

Graduate Management Admission Test and GMAT are registered trademarks of the Graduate Management Admission Council.

Note: ***This report is not valid for transmission of scores to an institution.***

Candidate Profile
Update Contact Information
Update Identifying Information
Update TOEFL Search Service Preferences
Change Password

Test Registration and Orders
Register for a Test
Order Score Reports
View Order(s)

Scores
View Scores

Store

Bulletin

My Home Page

View Scores

Registration Number: 0000000016203339

Request additional score reports

Test	Test Date	Reading	Listening	Speaking	Writing	Total
TELXML	Sat Nov 10 09:36:24 EST 2012	29	28	27	30	114

How to interpret scores

SUDHA RAO
14, Shalini apartment,
Sanghvi Nagar,
Aundh
Pune, Maharashtra
411007
India
raosudha89@gmail.com
(91) 8237209416
ETS ID: 7469229

Reading Skills	Level	Your Performance
Reading	**High(22-30)**	Test takers who receive a score at the **HIGH** level, as you did, typically understand academic texts in English that require a wide range of reading abilities regardless of the difficulty of the texts. Test takers who score at the **HIGH** level, typically • have a very good command of academic vocabulary and grammatical structure; • can understand and connect information, make appropriate inferences, and synthesize ideas, even when the text is conceptually dense and the language is complex; • can recognize the expository organization of a text and the role that specific information serves within the larger text, even when the text is conceptually dense; and • can abstract major ideas from a text, even when the text is conceptually dense and contains complex language.

Listening Skills	Level	Your Performance
Listening	**High(22-30)**	Test takers who receive a score at the **HIGH level**, as you did, typically understand conversations and lectures in English that present a wide range of listening demands. These demands can include difficult vocabulary (uncommon terms, or colloquial or figurative language), complex grammatical structures, abstract or complex ideas, and/or making sense of unexpected or seemingly contradictory information. When listening to lectures and conversations like these, test takers at the **HIGH** level typically can • understand main ideas and important details, whether they are stated or implied; • distinguish more important ideas from less important ones; • understand how information is being used (for example, to provide evidence for a claim or describe a step in a complex process); • recognize how pieces of information are connected (for example, in a cause-and-effect relationship); • understand many different ways that speakers use language for purposes other than to give information (for example, to emphasize a point, express agreement or disagreement, or convey intentions indirectly); and • synthesize information, even when it is not presented in sequence, and make correct inferences on the basis of that information.

Speaking Skills	Level	Your Performance
Speaking about familiar topics	**Good(3.5 - 4.0)**	Your responses indicate an ability to communicate your personal experiences and opinions effectively in English. Overall, your speech is clear and fluent. Your use of vocabulary and grammar is effective with only minor errors. Your ideas are generally well developed and expressed coherently.
Speaking about campus situations	**Good(3.5 - 4.0)**	Your responses indicate an ability to speak effectively in English about reading material and conversations typically encountered by university students. Overall, your responses are clear and coherent, with only occasional errors of pronunciation, grammar, or vocabulary.
Speaking about academic course content	**Fair(2.5 - 3.0)**	Your responses demonstrate that you are able to speak in English about academic reading and lecture material, with only minor communication problems. For the most part, your speech is clear and easy to understand. However, some problems with pronunciation and intonation may occasionally cause difficulty for the listener. Your use of grammar and vocabulary is adequate to talk about the topics, but some ideas are not fully developed or are inaccurate.
Writing Skills	**Level**	**Your Performance**
Writing based on reading and listening	**Good(4.0 - 5.0)**	You responded well to the task, relating the lecture to the reading. Weaknesses, if you have any, might have to do with • slight imprecision in your summary of some of the main points and/or • use of English that is occasionally ungrammatical or unclear.
Writing based on knowledge and experience	**Good(4.0 - 5.0)**	You responded with a well-organized and developed essay. Weaknesses, if you have any, might have to do with • use of English that is occasionally ungrammatical, unclear, or unidiomatic and/or • elaboration of ideas or connection of ideas that could have been stronger.

Return To Your Home Page

Statement of Funding

My current funding during the nine months of the academic year amounts to $18,360, drawn either from the Computer Science department (if I am a TA) or from external funding sources like DARPA, NSF, etc (if I am a RA) for the 20 hours/week of my TA/RA duties. During the three months of summer, I can work for 40 hours/week either at UMD (as TA/RA) or at an external internship. If I continue working at UMD, my funding for summer amounts to $12,240. If I go for an internship, the funding varies a lot from place to place. For e.g. my funding for this summer internship at USC amounts to $15,000.

Integrated Research Plan

In the field of computational linguistics, solving many problems requires transforming natural language text into a representation that is closer to how machines interpret input. Syntactic representations like constituent parse trees and dependency parse trees have been successfully used for solving many problems in this field for a long time now. But an equivalent sentence level semantic representation has been less explored. In recent few years, the field is seeing an increasing need for more semantic level understanding of text for solving more challenging problems. Currently most of the approaches to learning any kind of representation lay their foundation on techniques that need huge amounts of annotated data to work well. However obtaining such huge amounts of semantically annotated data is difficult and costly since it requires a linguistic expert in the loop. But humans do not need such huge amounts of data to be able to efficiently parse the meaning of a sentence. They can generalize much better. So how do they do that? There has been lot of work in the linguistic field that tries to answer this question. But the field of computational linguistics has not fully made use of this research. In this work I make an attempt to bridge this gap. In particular I am interested in developing linguistically motivated methods to automatically parse a sentence into one kind of semantic representation called Abstract Meaning Representation (Banarescu et al., 2013) (AMR). The two areas of linguistics that I wish to borrow ideas from are semantics and pragmatics. Using semantics I hope to understand how humans learn to compose meaning of a sentence using meaning of its parts and thus learn to identify the right relations between the parts. This understanding will be helpful for AMR parsing where one of the key challenges right now is getting the relations right between the concepts. Pragmatics on the other hand will help me understand how humans make use context and common sense knowledge while parsing a sentence. I plan to incorporate these ideas into AMR parsing using the distant supervision paradigm (described later). I believe that my prior experience in AMR parsing puts me in a good starting position to explore the computational aspect of this work. For gaining more insights into the linguistic aspect, I plan to take relevant courses and also collaborate with people working in Semantics and Pragmatics in UMD's strong linguistics group.

What is AMR?

The key motivation behind developing AMR was to have a comprehensive and broad-coverage semantic formalism that puts together the best insights from a variety of semantic annotations (like named entities, co-reference, semantic relations, discourse connectives, temporal entities, etc.) in a way that would enable it to have the same kind of impact that syntactic treebanks (e.g. Penn treebank) have on natural language processing tasks. AMR tries to capture the notion of 'who did what to whom' in a sentence. It abstracts away from syntactic idiosyncrasies yet preserves the meaning of the sentence. It is composed of concepts and relations, not nouns and verbs. In fact the creators go as far as to say that there are no nouns, verbs, adjectives, adverbs, affixes or zero pronouns in AMR. This allows AMR to have the same representation for a huge number of variations of a sentence (see fig 1)

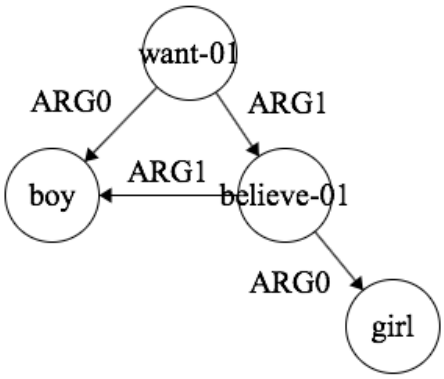
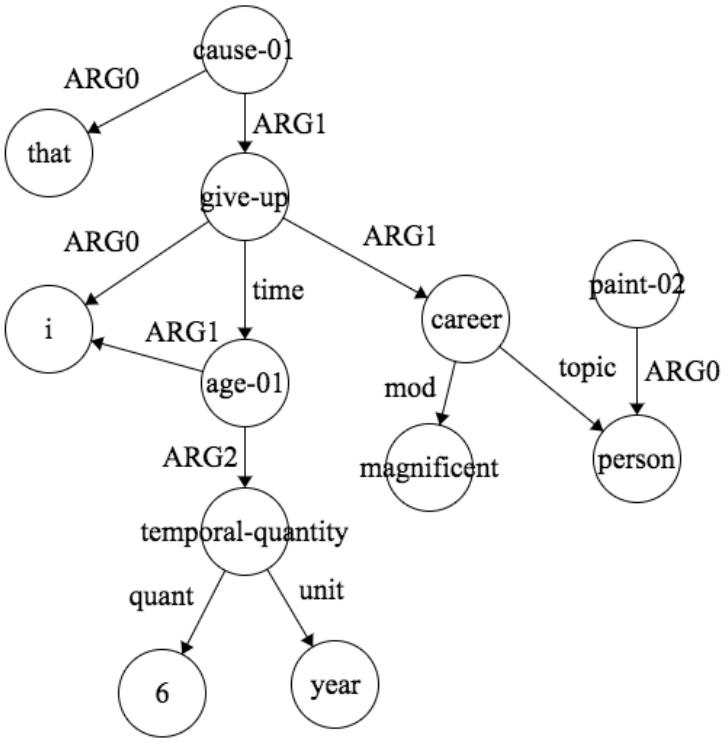


Figure 1: AMR above can be expressed variously in English as:
The boy wants the girl to believe him.
The boy wants to be believed by the girl.
The boy has a desire to be believed by the girl.
The boy's desire is for the girl to believe him.
The boy is desirous of the girl believing him.

Using Semantics

Currently one of the key challenges that we are facing while developing computational models for parsing text into AMR is identifying the right relations between concepts, especially the relations that are closer to the root of the tree. For e.g. below is a sentence and its AMR.

"That is why, at the age of six, I gave up what might have been a magnificent career as a painter."



For a computational model, it is more difficult to identify the relation "ARG1" between the concepts "give-up" and "career" than the relation "ARG0" between "give-up" and "i" because the former is a long-range dependency. Humans are faced with the same kind of challenge when trying to parse the meaning of a complex sentence, but they are good at it. Humans are able to compute the meaning of a sentence from the meaning of their parts. The way they learn to generalize is as Donald Davidson puts it:

> "The work of the theory is in relating the known truth conditions of each sentence to those aspects ("words") of the sentence that recur in other sentences, and can be assigned identical roles in other sentences."

In the field of linguistics, there has been a lot of work in developing a theory of meaning composition. According to Frege's Conjecture, semantic composition is a functional application on the parts of the sentence. In this work, I plan to borrow some of these principles and incorporate them into our computational models for improving AMR parsing.

Distant supervision and the use Pragmatics

Another key challenge in AMR parsing is parsing sentences that contain concepts or relations (between concepts) that were not seen in the training data used to train the parser. Since annotating sentences with their semantic representation is a hard task (requiring humans with linguistic expertise), we may not be able to get huge amounts of annotated data to train a parser. Even though annotating sentences is a hard task, there is a lot of structured data available out there in the form of knowledge bases like OpenCyc, DBPedia, Freebase, etc In this work I plan to use some of these to automatically get annotated sentences that we can train our AMR parser on. This technique of using external knowledge resources to get annotated data is called 'distant supervision' in computational land and has been successfully used for getting state-of-the-art results for relation extraction task. Since identifying relations is a key component in AMR parsing, this approach of distant supervision seems promising for AMR parsing as well. Another obvious justification for such an approach is the fact that humans definitely make use of their common sense knowledge a lot when trying to parse a sentence for its meaning. Linguists study this very phenomenon under the area of Pragmatics. Hence, in this work, I plan to first read through some of the existing work in Pragmatics that looks into how humans achieve this and try to apply them to AMR parsing.

References

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking.

Training Plan

Overview

The research plan that I have proposed requires me to have a good understanding of Semantic parsing in computational land and of human language learning in linguistics land. I intend to acquire these by taking a combination of relevant courses from the Computer Science and the Linguistics department. Along with these, I will require a deep understanding of the Abstract Meaning Representation paradigm some of which I hope to acquire through my summer internship this year (2015) at ISI USC working with the very creators of AMR. My graduate research and training will be conducted under the joint advising of my two advisors - Dr. Hal Daume III (from Computer Science) and Dr. Philip Resnik (from Linguistics department).

Coursework

In the first two years of my PhD program I completed eight out of the 10 course requirements of my home department (Computer Science). Out of the eight courses, five of them were cross-listed in the Linguistics department. Computational Linguistics I & II, seminar on Semantics in Computational Linguistics and Multi-lingual Natural Language Processing are some of these courses that will help lay the foundation of my training. In future, along with completing the two course requirements for my home department, I intend to take a few additional courses from the Linguistics department (like Syntax and Semantics) that will help me understand better some of the linguistic intuitions that humans use during parsing.

Science Policy Experience

Having been born and brought up in India, I have been exposed to a lot of different Indian languages. However I have seen very little computational work being done in any of them. Most of the Indian languages easily fall under the category low resource languages since digital data for them is almost non-existent. For my policy internship I would be very much interested in looking into what are the challenges that one would face if one were to start doing computational work on some of these low resource Indian languages. Some examples could be - digitization of data available in hard copy texts, putting together some of the existing research in these languages (if any), etc.

Term	Courses/other activities	CS dept requirements	LSF-NRT (cross listed in LING)
Year 1			
Fall 2013	CMSC651 -Analysis of Algorithms CMSC723 - Computational Ling I	Core Core	 In LING
Spring 2014	CMSC733 - Computational Ling II LING848 - Linguistic Prediction	Core Seminar	In LING In LING
Year 2			
Fall 2014	CMSC754 - Computational Geometry LING848 - Seminar on Semantics	Core Seminar	 In LING
Spring 2015	CMSC828I - Multi-lingual NLP CMSC878O - Sparsity and Machine Learning	Core Seminar	In LING
Summer 2015	Internship at ISI (USC) working on AMR		
Year 3			
Fall 2015	CMSC701 - Computational Genomics CMSC660 - Scientific Computing	Core Core	
Spring 2016	LING311 - Syntax I CMSC899 - Dissertation Research	 PhD req	In LING
Year 4			
Fall 2016	LING660 - Introduction to Semantics CMSC899 - Dissertation Research	 PhD req	In LING
Spring 2017	CMSC899 - Dissertation Research	PhD req	
Year 5			
Fall 2017	CMSC899 - Dissertation Research	PhD req	
Spring 2018	CMSC899 - Dissertation Research	PhD req	

UNIVERSITY OF MARYLAND
DEPARTMENT OF LINGUISTICS

1401 Marie Mount Hall
College Park, Maryland 20742
301.405.6760
resnik@umd.edu

July 6, 2015

Dear Colleagues:

This letter confirms that Sudha Rao and I have discussed her plans and that her participation in the LSF-NRT program has my support; Hal Daumé can, I'm sure, confirm that she has his support and the support of the CS Ph.D. program. I look forward to mentoring Sudha in this program, including reviewing and endorsing her training plan; playing an active role in her broad professional skills development; being an LSC member, and contributing to the success of the LSC.

Yours sincerely,

Philip Resnik, Professor
Department of Linguistics and Institute for Advanced Computer Studies
University of Maryland
College Park, Maryland 20742
resnik@umd.edu