Machine Translation: Theory and Practice

.....

Overview

There are several thousand languages in the world and there is a wealth of knowledge and information in all these languages. To make this knowledge available to all, it needs to be translated to other languages. There is only a limited amount that can be translated by human translators. Machine translation is translation of text from one language to another by computers.

There are two main approaches to MT. The first is rule-based and is inspired from linguistic insights. The second is statistical MT, which is mainly based on ideas from probability and statistics. There can also be a hybrid approach, which combines these two approaches.

The purpose of this course will be to introduce in sufficient detail the techniques for machine translation to those who have little or no background in this area, but have some background in Computer Science or languages/Linguistics. We will try to prepare the participants completing this course to initiate or take part in building MT systems for their own (or someone else's) languages. There will be emphasis on showing MT in practice, not just in theory.

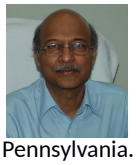
The whole course will consist of some seven lectures of around two or three hours each. Each lecture will be followed by a two hour lab session. These lectures and labs will be covered over a period of seven days. During the course of their lectures, the instructors will refer to actual working MT systems to provide context for the concepts presented in the class. Home assignments can be given if the participants desire. Through labs, the participants will be expected to get some hands on experience with building MT systems, using both the approaches mentioned above.

Modules	**A: Rule-based MT : Dec 9 – Dec 11** **B: Statistical MT : Dec 12 – Dec 15** **Number of participants for the course will be limited to fifty.**
You Should Attend If...	▪ you are an engineer or research scientist interested in language technology, and want to know more about machine translation. ▪ you are a linguist or from language studies and want to find out how your knowledge could be used for machine translation. ▪ you are a student or faculty member from an academic institution interested in natural language processing or computational linguistics and realize the importance of machine translation, but would like to learn more ▪ you want to get hands-on practice with machine translation
Fees	The participation fees for taking the course is as follows: **Participants from abroad : US $200** **Industry/ Research Organizations: Rs. 15000** **Academic Institutions: Rs. 5000** The above fee include all instructional materials, computer use for tutorials and assignments, laboratory equipment usage charges, 24 hr free internet facility. The participants will be provided with accommodation on payment basis.

The Faculty

Dr. Bruno Pouliquen is a senior software engineer with specialization in machine translation. He is working at the World Intellectual Property Organization (WIPO) in Geneva since 2009, where he is in charge of exploring and building patent machine translation software. He works now in statistical machine translation. Focusing on building automatic machine translation tools based on open source and working with 10+ languages. The software that he developed in WIPO is now installed and used in the United Nations, and in various United Nations agencies (ITU, IMO, FAO), as well as in patent office (Eurasian Patent Office, Moscow). He gives specific trainings on Machine translations to international organization officials and gives lectures to researchers on practical use of open source software.

Anil Kumar Singh is a researcher and a teacher who has been working in the area of NLP for the last thirteen years. He is working as an Assistant Professor in the department of Computer Science and Engineering at IIT (BHU), Varanasi, India. He has published on various topics in NLP and has organized a couple of research workshops and a couple of introductory workshops on NLP. He will be organizing ICON-2016 at IIT (BHU) immediately after this course. He is the creator of Sanchay, a collection of tools and APIs familiar to some researchers in India, and also some other open source software related to NLP. At present, he is also involved in building machine translation systems for Bhojpuri, Maithili and Magahi, which are all less resourced languages.



Prof. Rajeev Sangal is the Director of the IIT (BHU), Varanasi. He holds a B Tech in Electrical Engineering from IIT Kanpur, and MS and PhD in Computer and Information Science from the University of Pennsylvania. He is known for his work on computers and processing of human languages, and in particular on automatic translation among Indian languages. His research in Natural Language Processing led to the development of Computational Paninian Grammar (CPG) framework. It has been applied to parsing; semantics and morphological analysis of Indian languages etc. and forms the basis of several machine translation systems among Indian languages as well as English. He leads/mentors the multi-institutional consortium that has developed MT systems for 18 Indian language pairs. He has authored 4 books and many research papers. He is a fellow of the Indian National Academy of Engineering and of Computer Society of India. He was given Vasvik award in 2008. As the President of NLP Association, India from 2002 onwards he helped organize the community of NLP researchers and started ICON, a high-quality annual NLP conference.

Course Co-ordinator

Anil Kumar Singh
Phone: +91-8763594240
E-mail: aksingh.cse@iitbhu.ac.in

Address:
Assistant Professor
Department of Computer Science and Engg.
IIT (BHU), Varanasi-221005 (UP), India

Registration:
http://www.gian.iitkgp.ac.in/GREGN

Course Website:
http://gian-mt.sanchay.co.in