

CNGL Undergraduate Students as Researchers Programme 2011 PROJECT DESCRIPTION

<i>Institution/Track:</i>	Located in Dublin City University (ILT-1)	
<i>Project Title:</i>	Hierarchical Template-based EBMT	
<i>Suitable for students who are studying in the following areas:</i>	Suitable for students who are interested in Machine Translation (MT), Text Processing in general, and have a keen interest in developing large-scale systems.	
<i>Skills needed:</i>	Strong programming skills and precise Object Oriented Programming concept is desired. Must have coding fluency in Java. An in-depth knowledge of C concepts such as dynamic memory allocation, data and function pointers a plus. Sound knowledge of data structures is essential. Knowledge of Machine Translation, specifically Example-Based Machine Translation is also desirable.	
<i>Project Description:</i>	<p>The project will involve developing a prototype hierarchical template-based EBMT system, which is similar to hierarchical PB-SMT in flavour, but is essentially an EBMT system.</p> <p>Generalized templates are derived from parallel treebanks. From an aligned source-target (sub)tree pair, any combination of source nodes and the corresponding target nodes can be variabilized to obtain these translation templates. Template translation probability is calculated.</p> <p>The recombination module is a greedy algorithm that looks for maximal matching template(s), and fills up gaps recursively, and returns a ranked list of most likely translations considering template translation probabilities and a language model.</p>	
<i>The Role of the student & benefits gained from participation in this project:¹</i>	<p>The work itself would expose the student to the challenges of developing a large-scale MT system. The student will also develop deep insights into MT interacting with the big group.</p> <p>On successful completion of the project, the student will get authorship of any paper emanating from the work.</p>	
<i>Who will be working with you?</i>	The student will work closely with Sudip Kumar Naskar, a postdoc in the MT research group. He will also have the support of the whole research group. The student will be supervised by Andy Way's team of senior researchers.	
<i>Short description of the group:</i>	We have 22 researchers in our group – Prof Andy Way (team leader/PI), Prof. Josef van Genabith (Collaborative Researcher), 11 post-docs, and 9 PhD students.	
<i>Recommended Reading Material:</i>	<p>(2009) Philipp Koehn: Statistical machine translation. Cambridge University Press.</p> <p>(2003) Michael Carl and Andy Way (eds.): Recent advances in example-based machine translation. Dordrecht: Kluwer Academic Publishers.</p> <p>(2009) Cyril Goutte, Nicola Cancedda, Marc Dymetman, and George Foster (eds.): Learning machine translation. Cambridge, MIT Press.</p> <p>(2009) Yorick Wilks: Machine translation: its scope and limits. New York: Springer.</p> <p>http://www.mt-archive.info/</p>	
<i>Other information:</i>		
<i>For further details on this project please contact:</i>	Name: Phone: E-Mail: Website:	Sudip Kumar Naskar 6714 snaskar@computing.dcu.ie http://www.computing.dcu.ie/~snaskar/

¹ This is an initial description of the role of the student and it is liable to change following discussions between the investigators and the student.