**CNGL Postgraduate Internship Programme**

**PROJECT DESCRIPTION**

<table>
<thead>
<tr>
<th>Institution/Track:</th>
<th>CNGL – El Funded project</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Title:</strong></td>
<td>Improving domain-specific machine translation</td>
</tr>
</tbody>
</table>
| **Suitable for students who are studying in the following areas:** | Machine Translation  
Domain Adaptation |
| **Skills needed:** | Extensive familiarity of the open-source **Moses** machine translation platform  
**Java programming skills** would be an advantage  
Knowledge of **Chinese** or **Japanese** would be a huge bonus |
| **Project Description:** | We have developed state-of-the-art domain specific machine translation systems are used on a daily basis to translate patent information between English and French, German, Spanish, Portuguese, Chinese, and Japanese. Due to the variety of languages involved, the quality of the systems varies and we are constantly investigating ways to improve performance using techniques specific to each language.  
The aim of this project is to advance the investigation for three specific languages (to/from English), namely Chinese, Japanese, German. Various potential approaches include:  
German: developing improved processes for compound splitting and joining; investigating input pre-ordering techniques to improve translation fluency.  
Japanese: exploring transliteration to reduce out-of-vocabulary terms across Japanese writing systems; also investigating input pre-ordering techniques.  
Chinese: improving input segmentation and linguistic analysis to improve translation accuracy and lexical choice respectively.  
We are also interested in the adaptation of MT systems across locales via automatic post-editing. For example, producing Brazilian Portuguese output from our English-European Portuguese system.  
We would like to work with interns who have an interest or experience in any of the above areas related to machine translation, and how they can be implemented and deployed in real-world systems. |
| **The Role of the student & benefits gained from participation in this project:** | (e.g. what you foresee the student doing within the lab, what new laboratory skills/expertise they might acquire)  
- investigating methods for improving existing high-quality machine translation systems using techniques such as reordering, segmentation  
- evaluation and implementation of methods for new languages, particularly Asian languages  
- It is expected the student with gain significant experience dealing with practical machine translation applications and the associated challenges, e.g. the trade-off between quality, speed, and computational resources |
| **Who will be working with you?** | The postgraduate student will be working closely with Hala Almaghout, a postdoc in the group. S/he will also have the support of John Tinsley, the PI on the project and Fakih Karademir, our software developer. |
| **Short description of the group:** | We have 6 people in our group – Dr. John Tinsley and Dr. Páraic Sheridan who are the Pis on the project. 1 postdoc, 1 software developer, and 1 PhD student who used to work full time on the project. |

---

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.
**Reading Material:**


For how the technology is used, see www.iptranslator.com

**Other information (Funding?)**

The project in question is supported by an Enterprise Ireland Commercialisation Fund award

**For further details on this project please contact:**

<table>
<thead>
<tr>
<th>Name:</th>
<th>John Tinsley</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone:</td>
<td>(01) 700 6916</td>
</tr>
<tr>
<td>E-Mail:</td>
<td><a href="mailto:tinsley@computing.dcu.ie">tinsley@computing.dcu.ie</a></td>
</tr>
<tr>
<td>Website:</td>
<td><a href="http://www.iptranslator.com">http://www.iptranslator.com</a></td>
</tr>
</tbody>
</table>