

# ADAPT Undergraduate Internship Programme 2017

## PROJECT DESCRIPTION

|   |  |
|---|--|
| <b><i>Institution/Team:</i></b>   | ADAPT/ Irish Language Technology   |
| <b><i>Project Title:</i></b>  | Data Gathering for Irish Language Applications   |
| <b><i>Suitable for students who are studying in the following areas:</i></b>                                | Ideally the student will be undertaking either machine translation studies or translation studies.   |
| <b><i>Skills needed:</i></b>  | <ul style="list-style-type: none"> <li>• Solid Irish language skills (spoken and written).</li> <li>• Strong communication skills are necessary, as is the ability to simplify technical concepts to non-technical people where necessary.</li> <li>• Confident with the ability to work well as an individual and as part of a team.</li> <li>• Computer literate.</li> <li>• Strong technical ability (programming/scripting skills preferred but <b>not required</b>. However, a comfortable level of computer literacy is a must).</li> </ul>  |
| <b><i>Project Description:</i></b>  | <p>Together with the Department of Arts, Heritage, Regional, Rural and Gaeltacht Affairs, the Irish Language Technology team at DCU are responsible for the collection of English-Irish data that will be used to improve automated translation for public administration purposes. This is a European Commission led project and all data will be delivered to the European Language Resource Coordination.<br/> <a href="http://www.lr-coordination.eu">http://www.lr-coordination.eu</a></p> <p>The type of data we focus on includes previously translated text, terminologies, glossaries and corpora. The data collection process is ongoing and we require assistance in both the gathering of data from various public body sources and in managing the data (aligning translated documents, removing unwanted characters, analysing the content etc). In most cases, the gathering of data can be carried out electronically via email or through shared repositories. However, on occasion, onsite visits are required to assist public servants in finding and packaging their data for delivery to us.</p> |
| <b><i>The Role of the student &amp; benefits gained from participation in this project:<sup>1</sup></i></b> | <p>The student will be involved in conducting data collection and analysis from a number of different public bodies across Ireland and Northern Ireland.</p> <p>Through this work the student will gain experience in coordinating data contributions from various sources, data manipulation and formats, liaising with government officials and public servants. They will learn about machine translation systems, how they work, and how translators use translation systems to enhance their work.</p> <p>This project would suit a student with EITHER a linguistic/language background or computational background. The data collected is linguistic in nature but will be processed and applied in digital applications. The student will be part of an effort that is linked across Europe and directly benefitting the European Commission.</p>  |
| <b><i>Who will be working with you?</i></b>   | The student will be working closely with Dr. Teresa Lynn, a postdoc in the group. S/he will also have the support of Dr John Judge, who is also involved with this European project. The student will participate in all our project meetings during their time with us.)  |
| <b><i>Short description of the group:</i></b>   | We currently have 3 in our group – Dr Teresa Lynn (team leader), 1 PhD student and 1 Research Assistant)   |
| <b><i>Recommended Reading Material:</i></b>   | <p>Report from 2016 Irish ELRC workshop, kicking off the data collection for Ireland<br/> <a href="http://lr-coordination.eu/sites/default/files/Irleand/ELRC-Workshop-Report_Ireland-Public.pdf">http://lr-coordination.eu/sites/default/files/Irleand/ELRC-Workshop-Report_Ireland-Public.pdf</a></p> <p>Related publication on work on English-Irish Machine translation in DCU</p>   |

<sup>1</sup> ***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.***

|   |   |  |
|---|---|--|
|   | <a href="http://www.computing.dcu.ie/~tlynn/LRLpaper.pdf">http://www.computing.dcu.ie/~tlynn/LRLpaper.pdf</a><br><a href="http://www.computing.dcu.ie/~tlynn/tapadoir-cltw.pdf">http://www.computing.dcu.ie/~tlynn/tapadoir-cltw.pdf</a>                              |  |
| <b><i>For further details on this project please contact:</i></b> | <b>Name:</b> Dr Teresa Lynn<br><b>Phone:</b> 01-7006721<br><b>E-Mail:</b> <a href="mailto:teresa.lynn@adaptcentre.ie">teresa.lynn@adaptcentre.ie</a><br><b>Website:</b> <a href="http://www.computing.dcu.ie/~tlynn.html">http://www.computing.dcu.ie/~tlynn.html</a> |  |