Post Title: PhD Studentship in Term Extraction for Under-resourced Languages

Anticipated start date: September 2021 / October 2021

Stipend: €18,500 per annum (non taxed) plus university fees

Closing Date: 21st May 2021

Why ADAPT?

- **Contribute** to the ADAPT research agenda that pioneers and combines research in AI driven technologies: Natural Language Processing, Video/Text/Image/Speech processing, digital engagement & HCI, semantic modeling, personalisation, privacy & data governance.
- **Work** with our interdisciplinary team of leading experts from the complementary fields of, Social Sciences, Communications, Commerce/Fintech, Ethics, Law, Health, Environment and Sustainability.
- **Leverage our success.** ADAPT’s researchers have signed 43 collaborative research projects, 52 licence agreements and oversee 16 active commercialisation funds and 52 commercialisation awards. ADAPT has won 40 competitive EU research projects and obtained €18.5 million in non-exchequer non-commercial funding. Additionally, six spinout companies have been formed. ADAPT’s researchers have produced over 1,500 journal and conference publications and nearly 100 PhD students have been trained.

As an ADAPT funded PhD researcher you will have access to a network of 85 global experts and over 250 staff as well as a wide multi-disciplinary ecosystem across 8 leading Irish universities. We can influence and inform your work, share our networks and collaborate with you to increase your impact, and accelerate your career opportunities. Specifically we offer:

1. Opportunity to build your profile at international conferences and global events.
2. A solid career pathway through formalised training & development, expert one-on-one supervision and exposure to top specialists.
3. A Fully funded, 4 year PhD postgraduate studentship which includes a stipend of (€18,500 per annum - non taxed), along with equipment, annual travel funding
4. Funding for annual student fees

Research project/Challenge

**Term Extraction for Under-resourced Languages**

Terminology is a key part of understanding texts from specific domains and analysing digital content at a large scale. Term extraction is the task of finding terms where they are used in text and this task has largely been applied only to a few majority languages such as English. For under-resourced languages, including Irish, there is comparatively little study of this task and gold standard corpora are missing or very small. The student will develop an approach that takes a transfer learning approach using the resources that are available for major languages, such as English, and applying this to under-resourced languages and scenarios. In addition to transferring annotations, the student
will investigate approaches that are based on meta-learning to create classifiers for multi-word expression which distinguish n-grams into terms and non-terms. Further this approach will develop models that can compose word embeddings of multi-word expressions and deduce whether there is a compositionality to the meaning, and hence whether the n-gram represents a term, based on our previous models for this task. This work will provide students with a deep understanding of artificial intelligence and natural language processing that will equip the students for their further career in industry or academia.

Minimum qualifications:
Candidates for this position should have a MSc or equivalent in a relevant field of study with an emphasis on areas such as machine learning, natural language processing or linguistics.

Preferred qualifications:
The candidate should have good programming skills in Python or Java (and preferably experience with Tensorflow or other deep learning tools).

Application Process
The candidate will be based in the Data Science Institute (DSI) of the National University of Ireland, Galway.

Send your application (CV and cover letter, both in PDF only) by 21st May 2021 to Dr. John P. McCrae john.mccrae@nuigalway.ie.

Diversity
ADAPT is committed to achieving better diversity and gender representation at all levels of the organisation, across leadership, academic, operations, research staff and studentship levels. ADAPT is committed to the continued development of employment policies, procedures and practices that promote gender equality. On that basis we encourage and welcome talented people from all backgrounds to join ADAPT.

About the ADAPT Centre
ADAPT is the world-leading SFI research centre for AI Driven Digital Content Technology hosted by Trinity College Dublin. ADAPT’s partner institutions include Dublin City University, University College Dublin, Technological University Dublin, Maynooth University, Munster Technological University, Athlone Institute of Technology, and the National University of Ireland Galway. ADAPT’s research vision is to pioneer new forms of proactive, scalable, and integrated AI-driven Digital Content Technology that empower individuals and society to engage in digital experiences with control, inclusion, and accountability with the long term goal of a balanced digital society by 2030. ADAPT is
pioneering new Human Centric AI techniques and technologies including personalisation, natural language processing, data analytics, intelligent machine translation human-computer interaction, as well as setting the standards for data governance, privacy and ethics for digital content.

Our Research Vision
Governments and civil society are starting to recognise the need for urgent and concerted action to address the societal impact of the accelerating pace of digital content technologies and the AI techniques that underpin them. ADAPT provides an ambitious, ground-breaking, integrated research programme that assembles three interlocking Strands that together are capable of addressing this challenge. Each of these complementary and reinforcing research Strands takes one of the different perspectives on the provision of personalised, immersive, multimodal digital engagement, i.e. the individual’s experience and control of the engagement, the algorithms underlying digital content processing, and the balanced governance by enterprise and societal stakeholders.

Digitally Enhanced Engagement Strand
From the individual perspective, research within this Strand will deliver proactive agency techniques that sense, understand and proactively serve the needs of individual users to deliver relevant, contextualised and immersive multimodal experiences which also offer them meaningful control over the machine agency delivering those experiences.

Digital Content Transformation Strand
From the algorithmic perspective, new machine learning techniques will both enable more users to engage meaningfully with the increasing volumes of content globally in a more measurably effective manner, while ensuring the widest linguistic and cultural inclusion. It will enhance effective, robust integrated machine learning algorithms needed to provide multimodal content experiences with new levels of accuracy, multilingualism and explainability.

Transparent Digital Governance Strand
From the enterprise and societal perspective, new structured knowledge frameworks and associated practices for AI data governance will be required to balance the needs and values of individuals, organisations and society when it comes to rich digital experiences. This requires the advancement of research in the areas of data ethics, data quality, data protection, data value, data integration, and multi-stakeholder governance models.