**PhD Project Proposal**

School of Electronics, Electrical Engineering and Computer Science

|  |
| --- |
| **Proposed Project Title:** Understanding the brain and its diseases: building an ontology for the clinical neurosciences |
| **Principal Supervisor:** DrBarry Devereux  **Second Supervisor:** Prof Neil Robertson  **Third Supervisor:** Dr. Guillermo Lopez Campos (Institute for Experimental Medicine) |
| **Project Description:**  The human brain is the most complex system known to science, and understanding how the brain works, and the diseases which affect it, is an extremely challenging and rapidly developing area of biomedical research. The knowledge that is important to clinical neuroscience researchers involves a network of relationships and effects involving a multitude of inter-connected domains – genetics and ontogenetics, neuroanatomy and the functional role of different brain regions and systems, lifestyle factors and exposure over the lifespan to various environmental hazards, the aethology and pathology of neurological and age-related diseases, and the efficaciousness of different kinds of clinical intervention and cognitive rehabilitation. To understand and organize information about the complex relationships between these processes, medical practitioners and researchers need to be able to access the cumulative knowledge of the field in an efficient and reliable way. This is becoming an increasingly challenging task, given the increasing volume and complexity of the research literature (Akil, Martone, & Essen, 2011; Balan, Gerits, & Vanduffel, 2014a; Fleuren & Alkema, 2015; Gardner et al., 2008). How can we represent and access scientific knowledge in an efficient manner?  The goal of this project is to develop statistical and natural language processing (NLP) solutions to this critical problem. Although there is considerable research effort on developing information extraction methods for biomedical texts, the field of clinical neuroscience poses particular challenges, as clinically relevant information spans a wide range of domains, from chemical compounds to high-level human behaviour.  The Ph.D. researcher will develop models and approaches to text-mining a large corpus of biomedical literature relating to clinical neuroscience, and develop NLP-based methods to discover relationships and associations between these different domains. Using state-of-the-art approaches to representing the semantic content of words and sentences (Conneau, Kiela, Schwenk, Barrault, & Bordes, 2017; Devereux, Pilkington, Poibeau, & Korhonen, 2009; Hill, Cho, Jean, & Bengio, 2017; Kiros et al., 2015), the project will aim to create an ontology of a wide variety of neuroscience-related concepts.  Finally, we will use these algorithms and techniques to develop tools that allow clinical neuroscience researchers to visualise, search, and extract relevant information from a large corpus of biomedical publications relating to clinical neuroscience.  Key objectives will include:  • Develop techniques for modelling the semantic content of, and extracting information from, biomedical text (i.e. identification of relevant semantic concepts and their statistical relationship to textual terms in the corpus).  • Use these techniques to identify key trends in the clinical neuroscience literature and build a representation of clinical neuroscience knowledge that describes the statistical relationships between different clinically-relevant concepts and domains of knowledge (neuroanatomy, genetics, environmental exposures, pathologies, treatments, etc.).  • Develop information retrieval, search, and visualization tools for clinical neuroscience practitioners. |
| **Contact details**  Supervisor Name: Barry Devereux Tel: +44 (0)28 9097  QUB Address: ECIT Email: b.devereux@qub.ac.uk |