

March Success Stories from the School of Informatics

Mirella Lapata has been appointed a Fellow of the Royal Society of Edinburgh. Mirella is a chair in Natural Language Processing. She's a graduate of Carnegie Mellon University (MA) and University of Edinburgh (PhD). In 2009 Lapata became the first recipient of Karen Spärk Jones Award. Her research focuses on probabilistic learning techniques for natural language understanding and generation. She is interested in the general problem of extracting semantic information from large volumes of text. Mirella is among 62 new Fellows from the worlds of business, science, creative arts, health, technology, law, public engagement and international development who were announced on 1st March. They join an existing Fellowship of over 1600 individuals who give their time and expertise for free to support the RSE in delivering its mission of 'knowledge made useful'.

Two Centres for Doctoral Training offering postgraduate programmes in Natural Language Processing and in Biomedical Artificial Intelligence have received funding from UKRI. They form part of a drive to train the next generation of experts in AI (Artificial Intelligence) and build on the UK's reputation for emerging technologies. The UKRI Centre for Doctoral Training in Natural Language Processing, will aim to support greater understanding of human language processes, to allow AI to respond to the reality of human speech. The Centre, led by Professor Mirella Lapata of the School of Informatics, will develop expertise and research related to recognising or producing speech, retrieving documents or facts and responding to commands. It will also focus on summarising articles, translating text, and simplifying texts. The UKRI Centre for Doctoral Training in Biomedical Artificial Intelligence will seek to develop techniques to extract knowledge from biomedical data sets, with potential impact for public health and the economy. The centre, led by Professor Guido Sanguinetti of the School of Informatics, aims to spearhead the development and deployment of AI techniques in the biomedical sector. It will focus on the technical, biomedical and socio-ethical aspects of biomedical AI. Both are among 16 newly created centres in the UK, across the fields of industry, healthcare and climate change. The initiative is funded by £100m investment from UK Research and Innovation with £78 million in cash or in-kind contributions from project partners and £23 million from partner universities.

Miltos Allamanis, Informatics PhD graduate from 2017 won this year's Distinguished Dissertation, a prestigious academic award, run jointly by BCS, The Chartered Institute for IT, in collaboration with the Council of Professors and Heads of Computing (CPHC). Miltos won the competition for his PhD 'Learning Natural Coding Conventions'. He was supervised by Charles Sutton and Andrew Gordon. Currently, Miltos is as a researcher at Microsoft Research in Cambridge, UK and is part of the Deep Program Understanding project. The basis of his PhD was using machine learning to create better tools for software engineers. By learning from existing code, his aim is to create useful machine learning-based software engineering tools, interfaces and insights.

ICSA researchers won three paper awards at two co-located conferences in Washington DC: the International Symposium on High-Performance Computer Architecture (HPCA) and the International Symposium on Code Generation and Optimization (CGO). Artemiy Margaritov's and Boris Grot's paper titled 'Stretch: Balancing QoS and Throughput for Colocated Server Workloads on SMT Cores'

was selected as the Best Paper at HPCA. The main problem tackled in this work is that of improving server CPU utilization by concurrently running both batch and latency-sensitive applications while guaranteeing quality-of-service for the latter. The work was done in collaboration with ARM. Rodrigo Rocha, Pavlos Petoumenos, Murray Cole, Hugh Leather along with Zheng Wang from Lancaster University won CGO best paper award for their paper "Function Merging by Sequence Alignment" co-written by. It proposes using techniques developed in bio-informatics to allow merging similar functions and reduce the size of executables. Hugh secured another win, with Mike O'Boyle and Edwin Bonilla, as their 2009 paper "Automatic Feature Generation for Machine Learning Based Optimizing Compilation" won CGO's the test of time award. The paper introduced a mechanism for generating and selecting properties that are important for deciding code optimization decisions without any expert help.

Huawei Technologies Co. Ltd. has made a £1.2 million grant award to support studentships in the areas of dialog systems and data systems. These students will be working with Prof Mirella Lapata, Dr Shay Cohen, Dr Alexandra Birch, Prof Mark Steedman, and Prof Chris Williams. Huawei is actively looking to apply AI technologies in its products. They are working widely on machine learning and AI methods, including: AI systems that can learn and reason from small data, adapt over time, collaborate with each other, and interact with humans; speech recognition and synthesis, machine translation, natural language dialog, natural language generation and question answering; computer vision, including semantic understanding and reasoning about images; decision making and reasoning, making good decisions and learning to make better decisions.

Kenneth Heafield was awarded £750000 from Horizon 2020 for his project Browser-based Multilingual Translation project. Many companies and governments need translation but cannot use Google Translate for confidential documents. Statoil, the Norwegian state oil company, leaked confidential documents because they used translate.com for translation. In collaboration with Mozilla, researchers are creating a Firefox extension that does translation directly in the browser. This preserves privacy because the text never leaves the user's machine. However, desktops are far less powerful than servers used by online translation providers, so translation needs to be made more efficient. The extension also displays to the user how confident it is that the translation is correct and assist the user in filling out web forms in another language, helping them confidently produce answers in a language they do not speak.

Christophe Dubach received £37500 studentship from Microsoft Research Ltd for his project High-Level Synthesis of Neural Networks for FPGAs with LIFT. Deep learning applications are transforming our world and are extensively used in areas such as automatic language translation, business data analytic or self-driving cars. Specialised hardware is necessary to achieve maximum efficiency where speed and energy consumption is important. However, developing such hardware is extremely challenging and require many years of engineering efforts. This project investigates techniques to make the design of such hardware completely automatic, reducing drastically the time it takes to develop future machine-learning accelerators.

Walid Magdy received £30,000 for his project Human Behavioural Analysis for Better Understanding of Online Hate Speech. It has been reported multiple times that online hate speech can lead to hate crimes in the real world. As online communities play a greater role in real life, detecting and understanding hate speech in these communities grows in importance. This research proposes applying feature analysis of a state-of-the-art hate speech classifiers to better understand: 1) hate-speech and its characteristics, 2) user behaviour and interactions that lead to online hate-speech, and 3) characteristics and background for those users who post hate-speech online. During this project researchers build new datasets that cover hate-speech towards different people groups. They explore hate-speech beyond the used language by including the characteristics of those who use it and their online interactions.

Stephen Renals and Kenneth Heafield received a £320000 supplement from U.S. Airforce for their project Systems for Cross-Language Information Processing, Translation, and Summarization. The project is building a search engine that works across languages. The search engine can also find words in audio and video. Researchers are working with languages like Somali, Swahili, and Tagalog that have large speaker communities but have only modest amounts of translations and close-captioned audio available. Since computers learn from existing translations and closed captions, the challenge is to make the most out of a small amount of data. Researchers hope to make language technologies like speech recognition and translation available for far more languages.