A knowledge-based AI for arthropod vectored diseases

This project is one of a number that are in competition for funding from the South West Biosciences Doctoral Training Partnership (SWBio DTP) which is a BBSRC-funded PhD training programme in the biosciences, delivered by a consortium comprising the Universities of Bath, Bristol, Cardiff and Exeter, along with the Rothamsted Research Institute. The partnership has a strong track record in advancing knowledge through high quality research and teaching, in collaboration with industry and government.

Studentships are available for entry in September/October 2019.

All SWBio DTP projects will be supervised by an interdisciplinary team of academic staff and follow a structured 4-year PhD model, combining traditional project-focussed studies with a taught first year which includes directed rotation projects.

Lead supervisor:	Prof Matthew Wills, Department of Biology & Biochemistry University of Bath, email m.a.wills@bath.ac.uk
Co-supervisor:	Dr Nicholas Priest (University of Bath)
Non-academic CASE Supervisor:	Dr Peter Murray-Rist (ContentMine Ltd)

Project description

Obtaining answers to many of the most pressing biological and evolutionary questions will only be possible by integrating big data. How are species related, where are they distributed, what are their physical and behavioural attributes and how is their conservation status changing? The goal of this project is to build a semantic geopolitical map of arthropod-vectored disease – a knowledge-based AI to read the scientific literature and help us (everyone) understand it.

The answers to many global problems may already be hidden in the scientific literature! The Liberian Ministry of Health told the NY Times that a 1980 paper (still hidden in obscurity) predicted the Ebola outbreak. In a recent report on AI, Wendy Hall emphasized the need for mining the literature for knowledge. For example, the Zika virus is spread by *Aedes aegypti*, so in order to understand its epidemiology, we need articles relating mosquitos and W Africa. Mosquito control is probably critical, so we need papers about policy and economics. We also know that insecticide resistance is a growing problem and that many insecticides have neurotoxicity. We need a knowledge base which is universal, accessible to everyone and of modern design, using well-established AI, ML, and NLP methods. The knowledge base is Wikidata; the semantic content of Wikipedia articles ("infoboxes") with about 10 million instances. These have been enhanced by adding formal data from Genbank, PubChem, PDB, NCBI Taxdump, IUCN, etc. and scientific bibliography (50 million items with UniqueIDs and 300 million RDF triples).

ContentMine and the Wikimedia foundation have been working to create WikiFactMine - a set of semantic dictionaries that cover the whole of science, medicines, economics, geopolitics, etc. Using existing Open software the student+AI will read 20 million Open papers and (supervised learning) index the text and the figures/tables against species (insect and mammal), plants, pesticides, viruses, countries, geolocation, genes, proteins. Then s/he will use unsupervised learning (e.g. Open Knowledge Maps) to categorise the knowledge.

ContentMine supports machine-assisted human analysis by involving citizens (through Wikimedia and Open Knowledge International). We intend to capture their interest and expertise by gamifying the annotation of distribution maps. Applicants must have obtained, or be about to obtain, a First or Upper Second Class UK Honours degree (or the equivalent qualifications gained outside the UK) in an appropriate area of science or technology. In addition, due to the strong mathematical component of the taught course in the first year and the quantitative emphasis, a minimum of a grade B in A-level Maths (or an equivalent experience) is required.

Funding

Studentships provide funding for a stipend at the standard UKRI rate (currently £14,777 per annum, 2018/19 rate), research and training costs and UK/EU tuition fees for 4 years.

UK and EU applicants who have been residing in the UK since September 2016 will be eligible for a full award; a limited number of studentships may be available to EU applicants who do not meet the residency requirement. Applicants who are classed as Overseas for tuition fee purposes are not eligible for funding.

Applications

Applications should be submitted on the University of Bath's online application form for a PhD in Biosciences. Please ensure that you quote the supervisor's name and project title in the 'Your research interests' section. You may apply for more than one project if you wish but you should submit a separate personal statement relevant to each one.

The deadline for the receipt of applications is Monday 3 December 2018.