Title: Defining Emerging Technologies by Employing Advanced Machine Learning and Text Engineering Approaches

Proposed by: Prof Phil Shapira and Dr Abdullah Gök, Manchester Institute of Innovation Research, Alliance Manchester Business School and Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM)

Context:

Precise definition of emerging technologies is critical for a complete understanding of technology areas. Definitions of emerging technologies informs a wide variety of activities from making policy decisions on these technology areas to setting research strategy by individual organisations and scientists. Science and innovation studies has produced a stream of research on defining emerging technologies by employing bibliometric techniques. However, the current techniques have two shortcomings:

i) they mostly rely on static and limited information by employing only a set of given keywords contained in scientific publications (e.g. identifying nanotechnology outputs as publications including keywords such as "nanotechnology")
ii) they produce a binary outcome (i.e. a definitive set of publications) while in most of the cases, definitions of emerging technologies are heuristic and relative to the timing and purpose of the definition.

These shortcomings often result in definitions with considerably low levels of precision and/or recall.

Outcomes:

In this project, we propose to develop a machine learning approach to define a specific emerging technology (i.e. synthetic biology). Keyword based definitions underperform for synthetic biology more so than other emerging technology areas as most often synthetic biology papers do not use universally accepted keywords and the list of common keywords changes very frequently.

The machine learning approach we propose will go beyond simple and static keyword-based definitions by taking multiple criteria into consideration: (i) a moving list of keywords based on a core list, ii) known authors, iii) known journals, iv) known conferences, v) known references etc.) and it will create a probabilistic distribution rather than rigid binary outcomes.

Thus, the following outcomes will be achieved:

i) a new dynamic method of defining emerging technologies (methodological impact)
ii) a more precise definition for synthetic biology (impact on policy-making and research strategy)

Student will develop an algorithm together with the proposers. This algorithm will then be employed in constructing a dataset of synthetic biology publications. Knowledge of machine learning and text engineering concepts and preferably competence in Phython are the required capabilities.

Next Steps:

Together with the student and the CS partner, we will produce an 3* academic article on a new way of defining emerging technologies by employing machine learning. Potential outlets might be JASIST, Scientometrics or Informetrics.

This will also help us forge collaborative links with the CS colleagues. The successful completion of this mini-project will also provide us with the credibility for joint proposals for larger projects.