A better way to classify science

The past 20 years have seen genomics reshape the life sciences, nanotechnology drive materials research and chemistry, and climate change figure large in the Earth sciences. Megaprojects such as the BRAIN initiative in the United States and the European Human Brain Project may remake neuroscience and bioinformatics. And yet, to give one example, the most recent directive from the US Office of Management and Budget detailing Fields of Science dates back to 1978.

The names we give to disciplines matter, because the understanding and assessment of research—and ultimately policy and funding decisions—are rooted in this initial description. The Fields of Science classification is used to develop reports and statistics to review and evaluate science investments. Research taxonomies should help us to establish a clear link between these investments, outputs such as publications and patents, and outcomes such as economic growth, employment and an improved environment.

Using obsolete classifications, statistics produce only snapshots of a complex and multidimensional science value chain—and are weighted towards the input, or funding, end of the chain. It is also uncertain whether different reporters are classifying their activities in consistent and reliable ways. Inputs, outputs and outcomes are often accompanied by substantive descriptions and abundant textual data, but these data are typically unstructured and messy.

The task is to clean, process and ultimately classify this information in a useful and replicable way, so that policymakers can make better links between funding decisions and outcomes. The aim is to produce markers of activity all the way through the science value chain and map connections between all aspects of the science production cycle.

Text mining and machine learning offer powerful tools for this task. Instead of trying to fit what researchers do into predefined categories, they allow us to map the terrain of research, revealing boundaries, regions of overlap and a richness of detail. We are part of a team exploring their potential, testing a range of machine classification techniques on a grant from the US National Science Foundation.

Compared with manual methods, any computational technique needs to be: at least as accurate as human classification; consistent across time; cost effective; reliable across different science corpora; and replicable. In the first phase, we are using corpora such as dissertations and project grants, because these have manually curated classifications for comparison. Identifying disciplines—the most accessible comparison set—is the first milestone.

As well as reproducing human categories, we expect to add depth to existing classifications. We need to be able to describe different aspects of research, such as methodologies, technologies, problems and socioeconomic objectives served, not just disciplines. With this, we expect to reveal hitherto unseen interactions: between disciplines; between disciplines and technologies; and between disciplines, technologies and socioeconomic outcomes.

Two computer methods that are able to provide consistent and reliable disciplinary classifications are topic modelling, which classifies texts based on sets of co-occurring words, and wikilabelling, which mines encyclopaedic knowledge from Wikipedia pages and other repositories.

These approaches analyse document characteristics and apply external knowledge graphs, which trace the relationships between concepts in different classifications, to define the degree to which a document relates to a given research area. For example, an award abstract on genetics might be classified as biology, biochemistry, medical sciences or synthetic biology, depending on the context and content of the document. Topic modelling and wikilabelling, curated by expert users, are able to discover these multidisciplinary tags.

ONCE MACHINE METHODS HAVE proved reliable, the same model can be applied to other documents, including publications, patents and policy notes. Analysts and policymakers will be able to interrogate the literary deposits of scientific activity in great depth and from different perspectives. They will be able to test hypotheses about productive combinations of disciplines, problem sets and technologies, and track outcomes.

The goals of an improved classification system are similar to those of the Research Excellence Framework. The case studies and peer control used by the REF are valid measures, but they provide depth rather than breadth. Robust computational techniques for machine classification, combined with statistical analysis, can augment and test the insights produced by human evaluators, and help reduce the uncertainties in the science funding cycle.

Something to add? Email comment@ResearchResearch.com

> 'Machine methods will be able to test hypotheses about productive combinations of disciplines, and track outcomes.'

Evgeny Klochikhin is a researcher at the American Institutes for Research in Washington DC. Patrick Lambe is a founder and partner at Straits Knowledge, Singapore.