Essays on the Art and Science of Academic Journal Editorship and Publication

The Case for the Evolution of the All Sciences Journal Classification (ASJC) System

David Anthony Rew, MA MB MChir (Cambridge) FRCS (London)

Honorary Consultant Surgeon to the Faculty of Medicine, University of Southampton, UK
And to the Clinical Informatics Research Unit.

Subject Chair for Medicine to the SCOPUS Content Selection Advisory Board, Elsevier BV, The Netherlands, 2009 to the Present

Former Editor in Chief, The European Journal of Surgical Oncology, 2003-2009

A Working Paper for publication on the ePrint Server, University of Southampton

28th July 2025

Correspondence to
dr1@soton.ac.uk

Key words

Academic subject classification; All Sciences Journal Classification; Scopus; Web of Science; Medline; Ulrich's Periodicals; Article level classification, Machine Learning, Artificial Intelligence;

Essay Contents

Abstract

Background: The classification of documents, articles and journals provides structure, logic and integrity to a large and proliferating ecosystem of academic and research information. Reliable classification systems are of value to many users, including publishers, institutions and authors who are looking for the most appropriate repository for their publishable work.

The Elsevier All Sciences Journal Classification System (ASJC) is the best known exemplar of such systems. It is closely associated with the SCOPUS bibliometric system. The ASJC is more than 20 years old, and it has significant limitations in terms of granularity of subject coverage. It also now addresses much academic content outside science subjects, including Arts and Humanities content. It has particular limitations in the classification of the multi-disciplinary journals, which are now commonplace.

Methods: I report a series of supporting studies by Elsevier data scientists with machine learning (ML) and artificial intelligence (AI) software tools and SCOPUS data to investigate the optimisation the classification of journals across the ASJC. These studies included the use of an ML generated, article-based classification in a subset of journals which were classified in Medicine (All) (ASJC Code 2700) subgroup.

Results: ML based quantitative analyses confirmed the very uneven distribution of journals by subject across the ASJC. However, ML based attempts improve and simplify the process of re-classification of the ASJC using an article based methodology proved challenging.

Conclusions: The modernisation of the ASJC remains a worthwhile objective. The practicalities of securing a logical and “best fit” system commend a combined approach to optimise the use of the available ASJC codes with human subject matter experts.

This may be supplemented by a machine learning approach to parse the content of multidisciplinary journals into subsidiary categories. This will help the expert to score each journal according to the percentage of articles which it publishes in each high level subject area.

Introduction

It is natural to seek to classify complex information systems to simplify their comprehension and navigation. The corpus of human knowledge which is represented by the publication of academic papers and articles in journals, Theses; academic books, textbooks and book series; patents; conference proceedings; preprints and policy papers at Government and Institutional level from many countries and in many languages, is one such complex ecosystem.

Many researchers and academics in the STEAMM subjects (Science, Technology, Engineering, Arts, Medicine and Mathematics) continue to use conventional journals and Conference Proceedings to promote their outputs. However, many career academics and researchers now publish their outputs in other formats and non-traditional platforms , including eprint and preprint systems, social media systems and academic blogs.

~30,000 active academic journals are listed in 2025 in the SCOPUS quality assured citation system. Many other journals exist outside the quality assessment systems. Inactive and secondary journal sources may comprise a further 25,000 journals in SCOPUS.

Academic publications are living instruments in a Darwinian knowledge ecosystem. They are born and evolve. The most successful thrive over generations, but many others fade away. This vast ecosystem spans some 300,000 serial publications across all scholarly disciplines, and is recorded in Ulrich's Periodicals Directory (Proquest LLC). For the purposes of this paper, I will use the descriptor "Journals" flexibly to include other periodical publication formats.

An academic journal is defined by characteristics which include:

- The title, aims and scope of the journal;
- The subject matter, whether consistent with the Title, Aims and Scope, or not;
- The authorship of the content of the journal;
- The professional inputs into the journal, including the editor(s) and the editorial board;
- The institutions, societies and associations which the journal serves, where appropriate ;

- The publisher of the Journal, and the characteristics of that publisher, which include its ethics, its transparency, its governance arrangements and the jurisdiction in which its corporate and commercial operations are legislated;
- The history and “geography” of the journal, in terms of its origins and the target communities from which it derives content and which it serves;
- the metrics of the journal, which include measures such as the number and types of article and content per annum, the citation activity, and its performance metrics vis a vis its immediate peer publications.

The All Sciences Journal Classification System (ASJC) is generally accepted as the primary such system. It is based upon the title of each journal, along with the aims and scope. It was developed more than 20 years ago by a team at Elsevier Science. A list of all of the subjects and codes is available from the SCOPUS support centre on the Elsevier.com website.

The detailed origins, history, and maintenance methodology of the ASJC are uncertain, and seemingly lost in the corporate memory banks and oral history of Elsevier. Other bibliometric systems, including the Web of Science (WoS), use bespoke variations of this journal-based model in their content classification schema.

The ASJC predates the rapid expansion and diversification of modern academic publishing practices, which are built upon the universality of the Internet, the digitisation of content, and variations on the open access and “author pays” publication models. In the earlier era of print publication, journals were the primary product, and their titles generally reflected and determined the subject matter of their content. The title and aims of the journal were often closely aligned to the articles that they published. The journal title, on which the ASJC was developed, was generally considered to be synonymous with the designation of its content.

In the present era of digital rather than library shelf search, the individual article is now the primary product, and readers rarely see a journal in its entirety, even if it is co-published in a print version.

The Origins of this Project

This project originated in discussions which I had with colleagues in 2023 in the Elsevier SCOPUS management team about the growing non-alignment of the ASJC to the large number of journals of diverse content types which I was evaluating for SCOPUS in the fields of Medicine and the Health Sciences, and routes to possible solutions.

Specifically, there is insufficient use of the ASJC codes to address many of these journals with reasonable granularity. For example, the generic collection of Medicine journals (ASJC code 2700) contains more than 7,700 journals in SCOPUS, alongside which the Medicine (Miscellaneous) (ASJC Code 2701) alone now contains more than 2500 otherwise unclassified journals. The particular user cases that I had considered for an updates ASJC included:

- a researcher who was looking for a subject specific journal in which to publish his or her article, in for example, breast cancer, and
- a publisher of a journal who was looking for a subject-defined cohort of equivalent journals for comparative performance purposes.

The solution which I initially proposed was the development of a test system of human-machine collaboration through which I would manually review all journals in the relevant SCOPUS collections with a view:

- to correcting any glaring errors in journal coding in relation to the existing ASJC codes;

- to allocating new subsidiary subject codes to journals around common themes which were presently stacked in the Med (Misc);

- to further differentiating journals in an updated version of the ASJC according to the publishing and publisher model, which I will address in a separate essay in this series.

On further discussion, it became apparent that even with such an exercise, there would still be a large number of multidisciplinary journals which would not fit an updated version of the ASJC. Moreover, the content of non-journal sources was not addressed in the current classification, and further work was needed to determine how best to include it.

Subject Area	Subject Area Classifications
Physical Sciences	Chemical Engineering Chemistry Computer Science Earth and Planetary Sciences Energy Engineering Environmental Science Material Science Mathematics Physics and Astronomy Multidisciplinary
Health Sciences	Medicine Nursing Veterinary Dentistry Health Professions Multidisciplinary
Social Sciences	Arts and Humanities Business, Management and Accounting Decision Sciences Economics, Econometrics and Finance Psychology Social Sciences Multidisciplinary
Life Sciences	Agricultural and Biological Sciences Biochemistry, Genetics and Molecular Biology Immunology and Microbiology Neuroscience Pharmacology, Toxicology and Pharmaceutics Multidisciplinary

Figure 1. The four high level subject areas (super-groups) and the subsidiary high level subject classifications within the ASJC.

The Structure of the current All Science Journal Classification (ASJC) scheme

A Google Search of Journal Classification systems highlights the All Science Journal Classification (ASJC) scheme as the de facto primary Journal Classification system in common use. As originally created (Figure 1), the ASJC is characterised by four high level subject areas (super-groups). These are the Physical, Health, Social and Life Sciences. Each of these super-groups support a number of loosely connected subject areas.

Each subject area is allocated a four digit code. For example,

Life Sciences subjects are allocated 11xx codes,

Arts and Humanities are allocated 12xx codes

Biochemistry, Genetics and Molecular Biology are allocated 13xx codes

The Medicine (All) (2700) classification lies within the Health Sciences super-group. This allows for 99 further sub-classifications (2701 to 2799). ~50 of these numbers are as yet un-allocated, leaving scope for further sub-classification of Medicine journals.

Life Sciences and Medicine	Medicine	Emergency Medicine	2711
Life Sciences and Medicine	Medicine	Endocrine and Autonomic Systems	2807
Life Sciences and Medicine	Medicine	Endocrinology, Diabetes and Metabolism	2712
Life Sciences and Medicine	Medicine	Epidemiology	2713
Life Sciences and Medicine	Medicine	Family Practice	2714
Life Sciences and Medicine	Medicine	Gastroenterology	2715
Life Sciences and Medicine	Medicine	Genetics (clinical)	2716
Life Sciences and Medicine	Medicine	Geriatrics and Gerontology	2717
Life Sciences and Medicine	Medicine	Health Informatics	2718
Life Sciences and Medicine	Medicine	Health Policy	2719
Life Sciences and Medicine	Medicine	Hematology	2720
Life Sciences and Medicine	Medicine	Hepatology	2721
Life Sciences and Medicine	Medicine	Histology	2722
Life Sciences and Medicine	Medicine	Immunology and Allergy	2723
Life Sciences and Medicine	Medicine	Infectious Diseases	2725
Life Sciences and Medicine	Medicine	Internal Medicine	2724
Life Sciences and Medicine	Medicine	Medicine (all)	2700
Life Sciences and Medicine	Medicine	Medicine (miscellaneous)	2701
Life Sciences and Medicine	Medicine	Microbiology (medical)	2726
Life Sciences and Medicine	Medicine	Molecular Medicine	1313
Life Sciences and Medicine	Medicine	Nephrology	2727
Life Sciences and Medicine	Medicine	Neurology	2808
Life Sciences and Medicine	Medicine	Neurology (clinical)	2728
Life Sciences and Medicine	Medicine	Neuroscience (all)	2800
Life Sciences and Medicine	Medicine	Neuroscience (miscellaneous)	2801
Life Sciences and Medicine	Medicine	Obstetrics and Gynecology	2729
Life Sciences and Medicine	Medicine	Oncology	2730
Life Sciences and Medicine	Medicine	Ophthalmology	2731
Life Sciences and Medicine	Medicine	Orthopedics and Sports Medicine	2732
Life Sciences and Medicine	Medicine	Otorhinolaryngology	2733

Figure 2. Screenshot of a section the ASJC codes for Medicine, ordered alphabetically by subject matter. The screenshot is taken from the QS website, which lists all of the active codes and their subject allocations. The Medicine (All) 2700 and Medicine (Misc) 2701 codes) are highlighted in a red rectangle (see https://support.qs.com/hc/en-gb/articles/4406036892562-All-Science-Journal-Classifications-ASJC-Codes.

Note that some subjects which would logically be coded as Medicine subjects are not allocated 27xxcodes, for example Molecular Medicine (1313) and Neurology (2808), thus further highlighting inconsistencies in the original structure of the ASJC.

There is therefore considerable scope for improvement of the ASJC, to minimise overlap. For example, Environmental Sciences is classified as a Physical Science, while Agricultural and Biological Sciences are classed as Life Sciences. Many subjects in Health Sciences are classified in Life Sciences. For example, Psychiatry is listed as a subcategory of Health Sciences, while its partner subject, Psychology, is listed under Social Sciences.

Many journals are now multidisciplinary in content in order to attract the widest possible flow of papers. They may publish papers from more than one Primary Subject Field, as for example Nature, Science, PNAS, PLOS1, and Heliyon. These journals may be described as General Multidisciplinary Journals.

The rapid expansion of academic books within the SCOPUS collection and of Theses in the WoS collection and of Book Series, Patents and Conference Proceedings in both systems indicates that the classification needs of such content must also be incorporated in any globally useful future evolution of the ASJC.

The ASJC system is not comprehensive For example, Law is not afforded a high level classification, even though SCOPUS now lists ~1100 law journals.
(see www.scimagojr.com/journalrank.php?category=3308).

The Practical Uses of the ASJC

The ASJC is used particularly by Elsevier coders when a serial title is set up for Scopus coverage. It is also used in various forms by other organisations for classification purposes in the absence of a globally agreed and standardised system. For example:

- The widely used Quacquarelli Symonds (QS) ranking system evaluates institutions across five broad faculty areas and 55 subject areas, which are based upon ASJC codes.

- In the WoS, codes are assigned to journals from any of 250 categories. The WoS Subject category web page lists these 250 subject headings by alphabetical order from Acoustics to Zoology. However, but it offers no other sub-classification or granularity.

In the WoS system, a journal may have up to six assigned categories to it, including “a journal's categorisation in other bibliographic databases”. Although this is not made clear in the website description, it appears that multiple category codes may be assigned to multidisciplinary journals (Web of Science Core Collection subject categories 2023).

The Public Evidence Base for Design Problems with the ASJC

Various independent authors have observed the limitations of current journal subject classification systems. For example:

Qi Wang and Ludo Waltman (2016) studied the accuracy of the journal classification systems of Web of Science and Scopus, with particular reference to the field of Library and Information Science. They noted that some journals had weak connections with their assigned categories, while other journals were not assigned to categories with which they have strong connections.

Shir Aviv-Reuven and Ariel Rosenfeld (2024) noted unusually sized categories, high overlap and a lack of cohesion between categories in both Web of Science and Scopus systems, and that journals are often classified inconsistently. They concluded that these irregularities and discrepancies could not be easily disregarded.

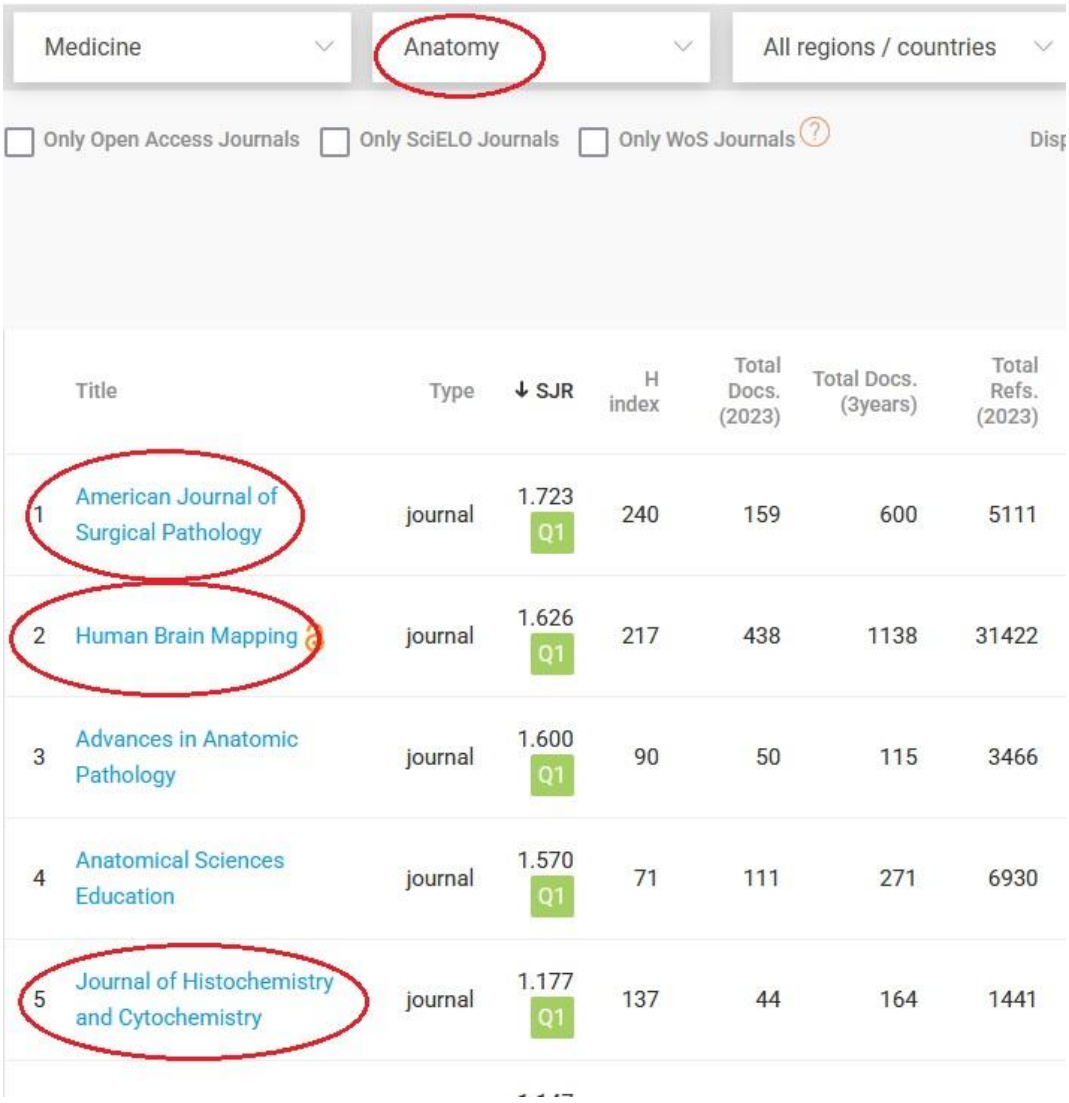
Mike Thelwall and Steven Pinfield (2024) assessed the publication practices of specialist, cross-field and general academic journals against their Scopus classifications. They compared the Scopus subject fields of journals with the fields that best fit their articles’ titles and abstracts. They also sought to distinguish between Scopus classification errors and misleading journal aims. They noted that some journals had titles and aims that do not match their contents. They concluded that such variations undermine citation-based indicators that rely on journal-level classification and may confuse authors in the search for appropriate journals in which to publish.

The issue of poor correlation of SCOPUS content with the ASJC was also forcefully highlighted in a blog post on Retraction Watch by Aleksandar Stević, under the title “Scopus is broken – just look at its literature category” (2024).

The SCIMAGO Journal Ranking Resource

The SCIMAGO journal ranking model is very useful for visualising the existing problems with journal classification. SCIMAGO is a Spanish research consortium which provides regularly updated metrics, journal rankings and data visualisations on SCOPUS data. It makes this data available in openly accessible formats on https://www.scimagojr.com/.

The Scimago website lists the journals and ranking of all journals which are published by SCOPUS and Web of Science by a number of data points and by characteristics which usefully include the inferred country of origin of a journal see: https://www.scimagojr.com/journalrank.php).

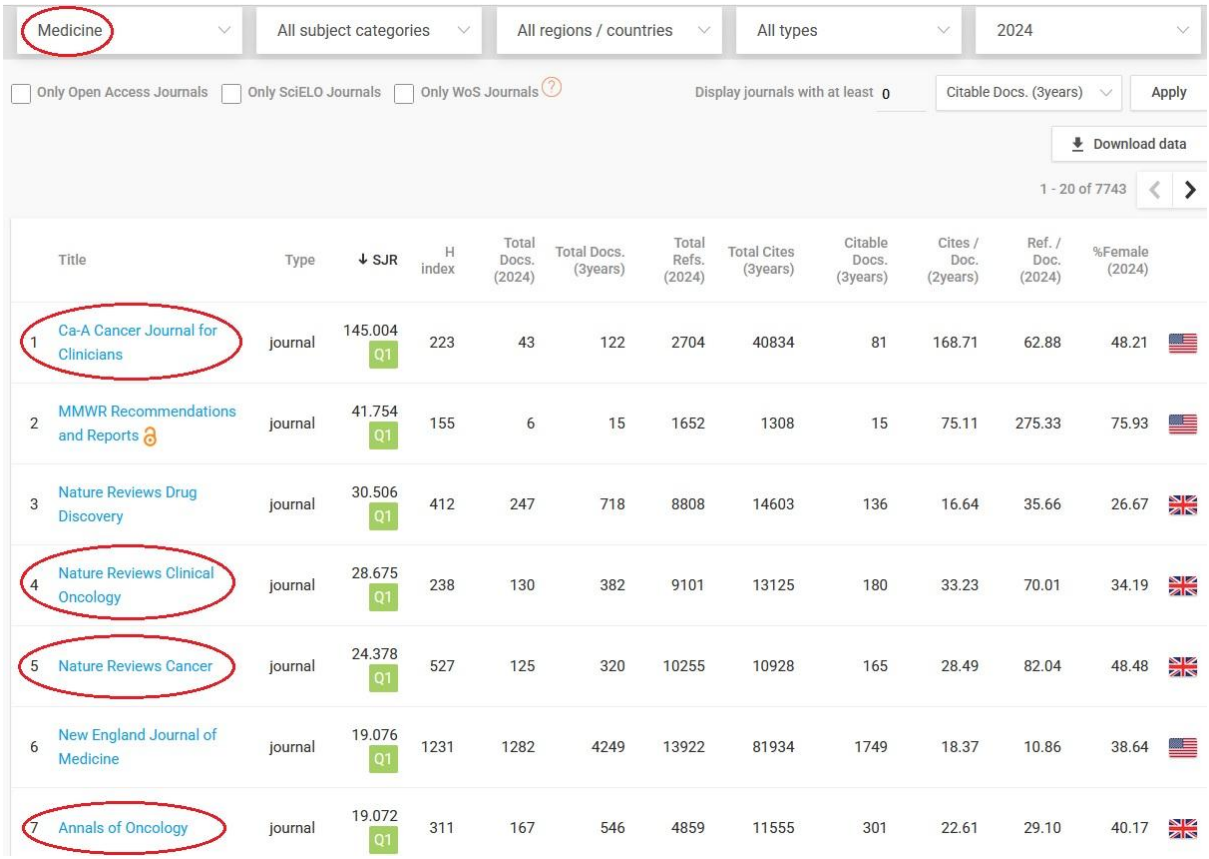


	Title	Type	SJR	H index	Total Docs. (2023)	Total Docs. (3years)	Total Refs. (2023)
1	American Journal of Surgical Pathology	journal	1.723 Q1	240	159	600	5111
2	Human Brain Mapping	journal	1.626 Q1	217	438	1138	31422
3	Advances in Anatomic Pathology	journal	1.600 Q1	90	50	115	3466
4	Anatomical Sciences Education	journal	1.570 Q1	71	111	271	6930
5	Journal of Histochemistry and Cytochemistry	journal	1.177 Q1	137	44	164	1441

Figure 3a; A screenshot of the SCIMAGO journal listings for Anatomy, (Feb 2025, see text)

The SCIMAGO listings are very helpful in the direct visualisation of apparent mismatches between the subject category and the title of a journal. Regular use of the SCIMAGO interface highlights the need for a clean up of many of the journal title allocations and reclassification within the ASJC.

This tendency to mismatches between journal titles and assigned journal categories is illustrated in Figures 3a and 3b. 3a is a contemporary screenshot of the Anatomy category. It was chosen at random from the Medicine subject fields. Journal No 1 (American Journal of Surgical Pathology) belongs in the Pathology category, as would Journal 5 (Journal of Histochemistry and Cytochemistry), while Journal 2 (Human Brain Mapping) is more appropriate to Neurosciences.



	Title	Type	SJR	H index	Total Docs. (2024)	Total Docs. (3years)	Total Refs. (2024)	Total Cites (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc. (2024)	%Female (2024)
1	Ca-A Cancer Journal for Clinicians	journal	145.004 Q1	223	43	122	2704	40834	81	168.71	62.88	48.21
2	MMWR Recommendations and Reports	journal	41.754 Q1	155	6	15	1652	1308	15	75.11	275.33	75.93
3	Nature Reviews Drug Discovery	journal	30.506 Q1	412	247	718	8808	14603	136	16.64	35.66	26.67
4	Nature Reviews Clinical Oncology	journal	28.675 Q1	238	130	382	9101	13125	180	33.23	70.01	34.19
5	Nature Reviews Cancer	journal	24.378 Q1	527	125	320	10255	10928	165	28.49	82.04	48.48
6	New England Journal of Medicine	journal	19.076 Q1	1231	1282	4249	13922	81934	1749	18.37	10.86	38.64
7	Annals of Oncology	journal	19.072 Q1	311	167	546	4859	11555	301	22.61	29.10	40.17

Figure 3b. This is a screenshot of the Medicine (All) (ASJC 2700 code) category on SCIMAGO (dated 28th July 2025). Four of the top seven journals of the 7743 titles in this category are self-evidently cancer (Onoclogy) journals, and yet they are not classified in the Oncology section.

METHODS AND FINDINGS

The observations on the deficiencies of the ASJC prompted a series of discussions with the Elsevier Classification and Analytics teams around options for the modernisation of the ASJC. In this section, I describe two experimental studies with ML/AI techniques which will inform our further discussions of how to improve the ASJC through a combination of machine learning and human-computer interaction.

Project 1. A study of the current pattern of content distribution within SCOPUS

In order to understand the current pattern of journal allocation to subject codes within the ASJC, Dr Rob Schrauwen and colleagues in the Elsevier data science and analytics team undertook a series of analyses the SCOPUS data set. Rob reports that:

"We had previously created a classifier which was designed for and trained upon a corpus of grant award notifications, which we acquire from various funding bodies. We connect these to researchers, organisations and research output using our Knowledge Graph tool. The subject classification which used the same terms as the Scopus corpus helped with this process.

The Elsevier website records that "A knowledge graph (KG) is a way of organising and connecting data to show relationships between elements. It links facts, concepts and entities to make it easier to explore complex information. KGs structure data in ways that support the further adoption of AI and ML, natural language queries, automation and scenario modelling... KGs are a powerful tool for the representation of data science. They offer a 3D visual representation of complex relationships between entities and for the navigation of large, disparate and multifaceted datasets". *(See elsevier.com/en-au/industry/knowledge-graph)*

Rob Schrauwen's report continues:

"A propriety classification model was created to assign classifications for articles. It was based on machine learning and widespread classifier techniques. There were two main reasons for an article classification system, vis:

- Use cases increasingly look at the articles themselves, and therefore classifying based on the journal's subjects does not correctly represent the subjects in the corpus of articles.

- Increasingly, Elsevier combines the Scopus corpus with non-Scopus material, sometimes with many hundreds of journals, whose subject classifications we don't have and don't need.

Academic articles differ from grant documents, but there was reasonable success in extending and tuning the algorithm for research article output. The following figures (4-8) show results of an early version of this model, but its development is incomplete.

There are inherent complexities in applying the ASJC classification to articles. Multi-disciplinarity is a feature of a journal and not of an article. Hence, not all classification codes apply to article classifications, and a separate All Sciences Article Classification "ASAC" scheme has been proposed.

The design of an algorithm to aggregate this to journal level is nontrivial. For the figures supplied, a simple aggregation was used, based on frequency. This model is not suitable for commercial production use but it provides useful insight into the accuracy of journal subject classification."

The distribution of numbers of journals in the current ASJC is shown in the histogram in Figure 4. The chart highlights the substantial disparity in numbers from one category to another, with a particular peak in the 2700 (General Medicine) Category.

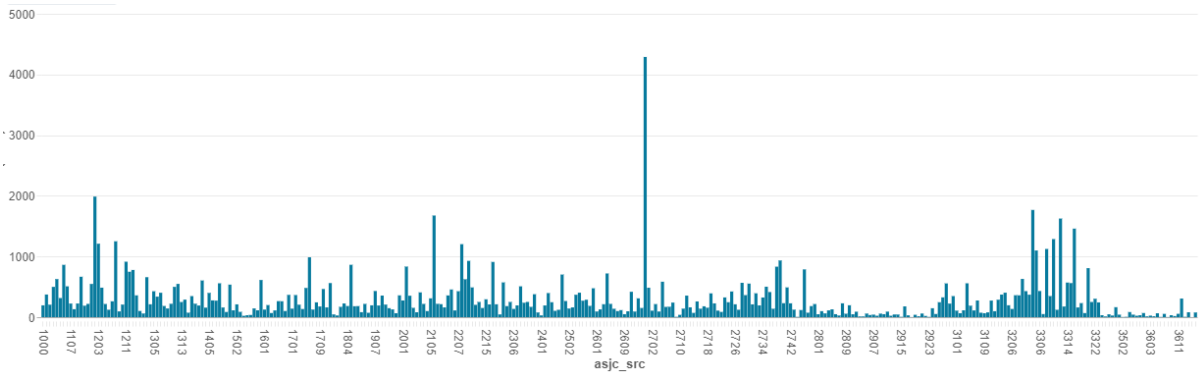


Figure 4: This displays the Number of **academic journals** per ASJC category. The X axis displays each of 334 categories in 27 broad subject areas (in thousands). The Y axis displays the number of journals in each ASJC category. (courtesy of Dr Rob Schrauwen, May 2024)

This very large peak reflects the lack of granularity in the ASJC. Therefore, a very large number of journals have been allocated the holding code for want of a more detailed Medical subject classification and journal allocation across ASJC codes 2701-2799.

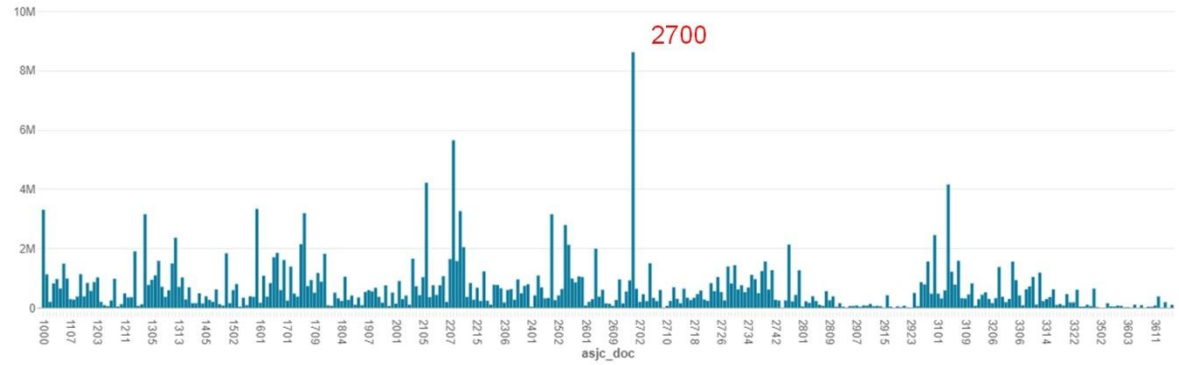


Figure 5. This figure displays the number of **documents** per ASJC code, as originally allocated to the journal. (see text). (Figure courtesy of Dr Rob Schrauwen May 2024)

Figure 5 similarly displays the number of individual documents which are listed in SCOPUS as they are associated with the ASJC codes through the journals in which they are published. Rob Schrauwen comments that:

"*We refer to the underlying document allocations as "aspirational", in that they reflect the documents which journals have published through their subject coverage aspirations, as expressed in their aims and scope. By far the largest number of documents remains associated with the 2700 General Medicine category. On this graph, the Y axis displays the numbers in millions of documents."*

In the next study with their proprietary algorithm, Rob and his team analysed every document in SCOPUS and applied a machine generated classification of each article to its most appropriate subject category. This classification reflected the "observed" or deduced topic of the article, rather than the topic which had been allocated to it simply by the classification code of the journal in which it had been published (the "aspirational" allocation).

The redistribution of articles was striking. For example, articles from journals in the Medicine (code 2700) were widely redistributed across the medical specialities for which specific codes exist. Elsewhere in the data set, many articles were re-allocated to Code 2208 (Electrical and Electronic Engineering) as shown in Figure 6.

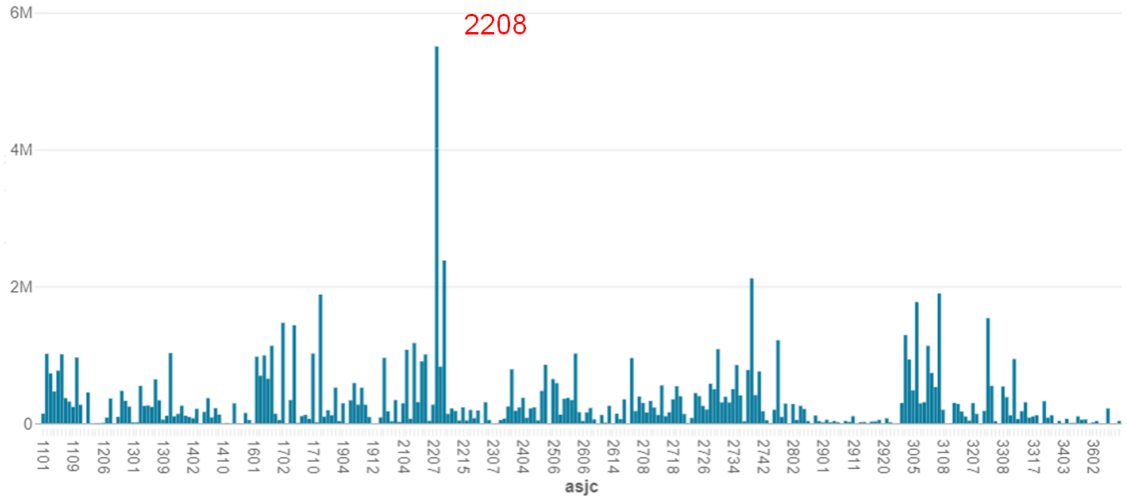


Figure 6. This figure displays the redistribution of documents in SCOPUS from existing journal subject categories to observed article subject categories after application of an experimental proprietary algorithm. The Y axis displays the number of articles. The X axis displays the ASJC subject codes. (see text) (courtesy of Dr Robert Schrauwen May 2024)

Figure 7. The Physical Sciences supergroup (courtesy of Dr Rob Schrauwen May 2024)

We can look at this data in another way, using a hierarchical tree for the subject fields in the Physical Sciences supergroup (Figure 7). Each coloured block is labelled with the relevant high level ASJC code, and the colour of the block represents the range of the number of articles that have been re-allocated to the block through the machine algorithm.

This shows the redistribution of documents SCOPUS from existing journal subject categories to observed article subject categories by ASJC code in the Physical Sciences. See text
All categories held between 10,000 and 10 million articles, with one exception. The 2101 (Engineering, Miscellaneous) code, highlighted within the red oval, was allocated fewer than 10 articles.

This figure demonstrates that the Physical Sciences subject codes in the ASJC are sufficiently granular to accommodate all of the reclassified documents, all be it that there is a thousand-fold difference between the most and least “popular” topics.

This movement of articles and documents between subject areas is also illustrated in the Heat Map in Figure 8. In this figure, the algorithmic calculation of the most appropriate (the “observed”) category for each article has been plotted against the (“aspirational”) category to which its parent journal had originally been allocated

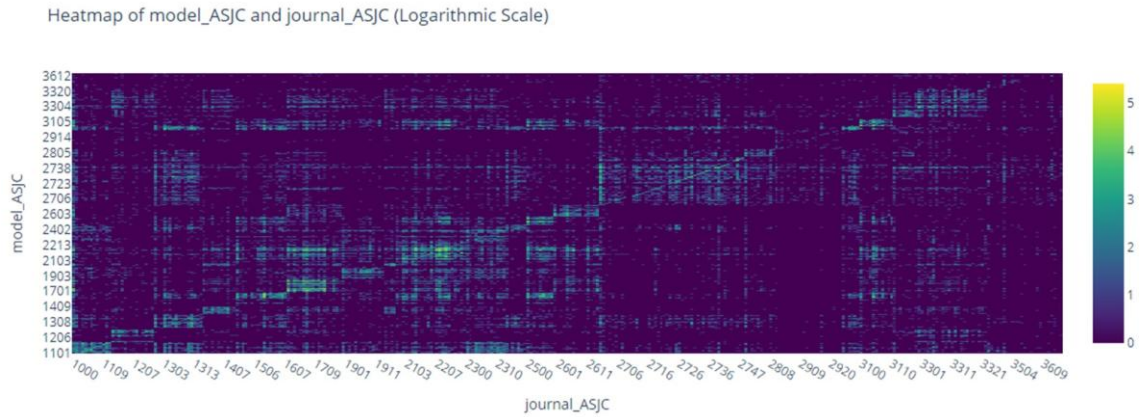


Figure 8. This is a heatmap of an article-based model of ASJC subject allocations, (Y axis), plotted against the journal-based ASJC classification in which the article was published (see text). The brightness on the colour scale reflects the number of articles which have been reallocated to each code. (Image courtesy of Dr R. Schrauwen May 2024).

This analysis highlights the challenges of how best to allocate both articles and journals to the most appropriate ASJC subject classification code. If there were complete concordance between the calculated (new model) and pre-allocated parent journal classification, there would be a linear diagonal relationship between the two groups. In practice, while this relationship is discernible for many articles, it is also clear that there is a substantial difference between the new model "best fit" allocation of the article and the old subject allocation of the journal in which each article was published.

Project 2. A detailed study of the ASJC Classification for 2700 Medicine (All) Journals

This exploratory project was undertaken by Mr Sujit Pal, Technology Research Director for Elsevier Health Markets business team and colleagues (referred to as "we") in this section) to explore the development of a Human-in-the-Loop ASJC analysis and code allocation tool. This phase of the project involved a series of data manipulations using a complex series of artificial intelligence tools.

Sujit Pal's statement of the Categorisation problem

Sujit writes that *"the test software system explores the original proposal for a human-machine teaming tool that would expedite the re-classification of the large number of journals in the Medicine (All) (ASJC 2700) category.*

Many Medical Journals are effectively uncategorised in this category. We hypothesised that a proportion of these journals may be re-categorisable into one or more of the 50+ sub-categories that already exist under ASJC Code 2700. Given the inherent ambiguity of the task, a software model that helps the human expert might be a better solution than a fully automated process.

For the purpose of this study, we also assume that the journal name is sufficient to categorize the journal accurately in many cases. In other cases, a study of the journal's website may provide clues for accurate categorization."

In the first Model, the test system is trained on the titles of journals that are already assigned to medical subcategories. It is then applied to those journals that are assigned to

the 2700 category in decreasing order of confidence. An assumption is made that those that are already categorised other than in 2700 already are correctly categorised. This may or may not be true, but is a logical starting point.

This model predicts the most likely top five subcategories to which any journal can be mapped from its name alone. It is meant to assist the human expert to decide on the ultimate subcategory (or subcategories).

The uncategorized (2700) journal titles are presented in decreasing order of confidence, so the expert sees first the examples where the model can provide the most accurate decision assistance. As the expert scrolls through the journal titles, the model will become less confident and hence less helpful.

At that point we applied the second Model (as described in the Next Steps section) on the remainder, and then the third Model when the limit of the second model is reached.

Model 1: Generating the Training Dataset

In this exercise, we extracted journals in the Medicine Category (27xx) from a proprietary dataset (ops_etl.sources_20250331) on Databricks. Databricks is a commercial system which integrates generative AI tools with a "data lakehouse" into a Data Intelligence Engine. A data lakehouse is a modern data architecture that creates a single platform from "data lakes", which are large repositories of raw data in its original form, and data warehouses, which are organised sets of structured data.

We treated journals with the ASJC code 2700 as uncategorized, and all other journals as they had originally been categorised to the specific ASJC coded subcategories.

Some Journals mapped to multiple sub-categories of Medicine were "exploded" such that each record was a (journal title, ASJC code) pair. Journals which mapped to sub-categories outside Medicine were also exploded, and records containing non-Medicine ASJC codes were ignored.

This method identified resulted 14,359 unique journal titles, each of which was mapped to one or more ASJC codes. After further processing, we ended up with 14,206 categorised (journal title, ASJC code) pairs and 4,372 uncategorised pairs which were mapped to 27xx ASJC codes.

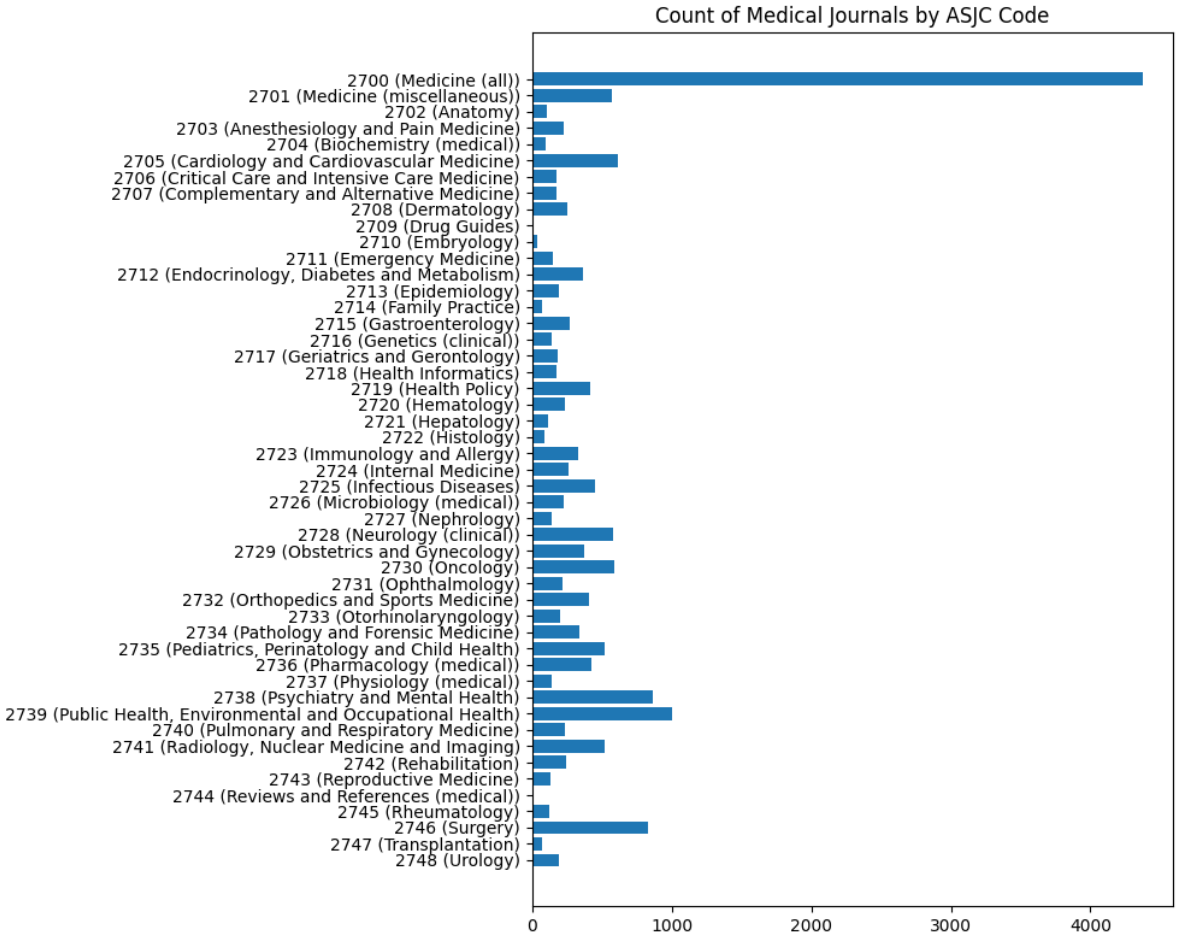


Figure 9: Journal Counts by ASJC Code for Medicine (27xx) Journals. This Figure shows the distribution of journal titles by ASJC code.

The journals that are categorized as Medicine (All) (ASJC Code 2700) were treated as uncategorised for purposes of this exercise. The other bars identified journals which were allocated to any one of the AJSC (27xx) Medicine sub-categories, which were used to train the Predictive Model.

Model 1: Training the Predictive Model

The journal names were each then converted to Dense Vectors of size (768,) using the BiomedNLP-BiomedBERT-base-uncased transformer model. This is a large neural language model which is pre-trained from scratch using abstracts from PubMed and full-text articles from PubMedCentral. Dense vectors are a form of mathematical objects that represent data in machine learning and artificial intelligence applications, where-in non-zero values populate its elements.

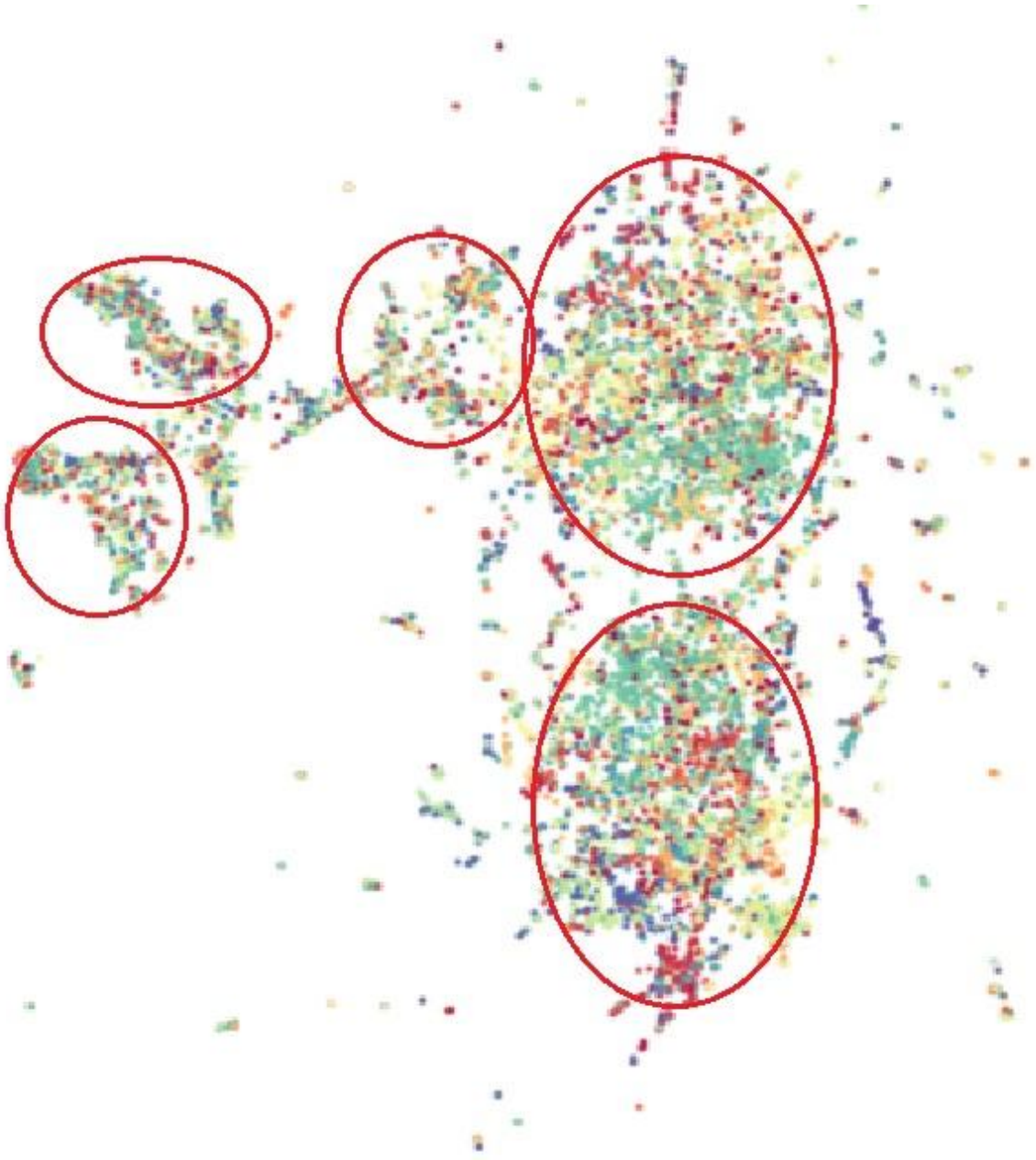


Figure 8: Projection of Vectors representing Journal Titles categorized by ASJC sub-category The manually drawn red circles highlight the clustering patterns (Courtesy of Dr Sujit Pal)

The dense vectors for the journal titles in the categorised group (27xx ASJC codes) were then used to train a **Support Vector Machine** (SVM) Model with 5 fold cross-validation. An SVM is a supervised machine learning algorithm that classifies data by finding an optimal distance between each class in an N-dimensional space. Support Vectors are data points that are closest to the decision boundary in a support vector machine.

We also projected the encodings to two dimensions using **Uniform Manifold Approximation and Projection** (UMAP) to verify that the encodings has enough discriminative power. UMAP is a nonlinear dimension reduction method that is used for visualising data and as pre-processing for further machine-learning tasks such as clustering. The visual evidence of the Vector Projection indicated that the vectors from the encoder model did indeed capture a sense of the sub-categories.

Application of the trained SVM model against the uncategorised journal titles

We then applied the trained SVM model against each of our uncategorized (ASJC code = 2700) journal titles to predict the probability of it being categorised in each of the 48 ASJC Medicine sub-categories. We used a mathematical probability model to capture the top five classes from the model.

Annotation using the LabelStudio Integration system

We integrated data for the annotation step with LabelStudio, which is a popular open-source annotation tool. In machine learning, data annotation is the process of labeling or tagging data to make it understandable for machine learning algorithms.

For Data Importation, a LabelStudio project was created manually, and the task data was uploaded as a JavaScript Object Notation (JSON) file. This contained a sequence of records, each of which consisted of the uncategorized journal title, the top five predicted classes and probabilities from the SVM model, and the prediction entropy, which quantifies the uncertainty or randomness in a model's predictions. The lower the entropy predictions, the more confident is the model of predictions.

The Next Steps in the Development of the Predictive Model

Given the experience from these studies, it seems unlikely that we would be able accurately to re- assign all titles from ASJC Code 2700, using successively more elaborate categorisers and AI methodologies, when compared with an efficient human expert with easy access to the list of journals to be assigned, and to their websites for crosschecking, and to a simple data entry system to record the decisions.

Sujit summarised this project as follows:

"My goal was to use the Support Vector Machine model to assist the human expert, in identifying the obviously easy titles and getting them out of the way. You could either choose one or more of its top- five predictions or skip it if none of these were relevant.

For the expert, the intended benefit is faster decision making (selecting from machine generated choices rather than deciding from scratch for every journal). The side effect of this is that we learn the confidence threshold at which point we can no longer rely on this model, and switch to a better model for our predictions.

My goal was to create a sequence of models which would be applied to undertake sequential data analysis to reach a decision. For example, while the SVM model just looks at the title, we might next identify the journal website from SCIMAGO or similar, then scrape the website to extract the scope of the journal.

This exercise might also help the human expert to apply one or more ASJC codes to a journal, with enhanced information.

It may be that for the most difficult cases, no computational model can deliver a good prediction. At that point, the human expert would have to use his or her judgment to apply the best fit ASJC code. The intent is simply to assist the human expert in the task to the greatest possible extent."

Discussion

Coded classification systems are well established in science and in Clinical Medicine in particular. For example, READ Codes are a coded thesaurus of clinical terms which have been in use in the UK National Health Service since 1985, and they long predate the Electronic Patient Record (EPR). The SNOMED CT code is a structured clinical vocabulary for use in an EPR.

In discussions about this paper, Professor Peter Brimblecombe of the SCOPUS Content Advisory Board commented that:

"I have always been interested in the problem of knowledge classification. I recall that the Universal Decimal Classification (UDC) is very clever in handling the categorization of knowledge. Elsevier might have looked at this when they created the original ASJC system".

Wikipedia tells us that *"the UDC was developed by the Belgian bibliographers Paul Otlet and Henri La Fontaine. In 1895, they created the Répertoire Bibliographique Universel (RBU). the UDC is a bibliographic and library classification which represents all branches of human knowledge as a coherent system in which knowledge fields are related and inter-linked. The UDC is an analytico-synthetic and faceted classification system which features a detailed vocabulary and syntax.*

These allow powerful content indexing and information retrieval in large collections. [1] *Since 1991, the UDC has been owned and managed by the UDC Consortium. This is an international non-profit association of publishers with headquarters in The Hague, in the Netherlands.*

The UDC is used in around 150,000 libraries in 130 countries and in many bibliographical services which require detailed content indexing. It is a primary classification system for information exchange and it is widely used in public, school, academic and special libraries".

The Evolution of the All (or Academic) Subjects Journal Classification Scheme

Over some 25 years, the ASJC has provided a pragmatic and useful classification tool for academic journals, which is based primarily upon the title of any journal. The original reasoning for the creation of the ASJC has been lost in the Elsevier corporate memory, but it must have been persuasive at the time to justify the investment in resources and effort to its creation and maintenance.

Journals are allocated an ASJC subject classification on accrual to the SCOPUS system by trained staff who are not subject specialists, on a "best guess" basis. Once allocated, journals are very rarely if ever reallocated to a different code, so they carry any original allocation errors indefinitely.

Moreover, there have been a number of major changes in academic publishing since the system was conceived and the 4 digit codes were allocated. New subject areas have emerged and multi-disciplinarity has become a major factor in publishing on the back of the expansion of low cost digital publishing and the commercial drivers of the open access movement.

The studies reported in this paper grew out of extensive collective reflection on the nature and purpose of the All Sciences Journal Classification scheme and its future development. At the most basic level, the title of the scheme no longer reflects;

- The scope of subject coverage of the system, which now includes (for example) many journals from the Arts and Humanities, Law, and other subjects. A change of name from the "All Sciences" to the "All (or Academic) Subjects" Journal Classification scheme would more accurately reflect its current coverage, while retaining the ASJC Branding.

- The scope of content coverage beyond conventional academic journals. This includes books, book series, theses, patents, conference proceedings and other formats. This suggests that a more appropriate title would be the Academic Subjects Publication Classification (ASPC) scheme for consistency and inclusivity of all sources of academic publications.

The past 25 years have also seen the dramatic growth in the power and scope of computing systems, and the capabilities of machine learning and artificial intelligence. This has permitted the analysis and processing of publications at the article, author and institutional level rather than merely on the title of a journal. This technological expansion therefore poses the questions as to whether the ASJC should be:

a. retained, expanded and refined all be it under a modified title;

b. replaced by a wholly new, technology driven and article based classification system; or

c. modernised as a hybrid system in which the current ASJC is retained but adapted to include elements of a new article-based classification methodology.

The more general question therefore arises as to whether or not it is necessary to the modernise the existing ASJC. We might reasonably take the view that a contemporary, reliable and workable classification system for Journals and other academic outputs may yet retain an important role for the academic publishing and bibliometric ecosystem for many reasons, which include:

- user cases for Institutional, Corporate, National and International publishers and bibliometric developers;
- the clearer definition and detection of malpractice in Academic Publishing;
- the direction of authors who are looking for a suitable home for their manuscripts;
- the support of researchers who are looking accurately and consistently to describe their academic productivity.

The Unresolved Problems with the Current Version of the ASJC

Beyond the easily adapted issue of the title of the system, the principal unresolved problems with the current version of the ASJC are as follows.

a. Code allocation errors between the ASJC subject category and the journal title

b. Underutilisation of the “spare” ASJC codes to create a more granular classification. A broadening of the subject granularity to make use of the reserved and unused 4 digit identifiers would allow a more accurate allocate on many journals which are currently allocated to large repositories such as General Medicine (2700) and Medicine

(Miscellaneous) (2701). The heavy default use of a "Miscellaneous" option in each subject area is a clear indication of the lack of granularity in the current ASJC.

c. Lack of flexibility, adaptability and dynamic monitoring to adjust the ASJC listings as the purpose and content of journals change over time

d. The need for a standard method to integrate all publication sources beyond conventional journals into an updated ASJC/ASPC (All Subjects Publications Classification) scheme

e. There is duplication and redundancy among some subject categories. For example, ASJC Code 2709 ("Drug Guides") can be subsumed as a category of Pharmacology.

f. Duplication and mis-classification across top level subject fields is also evident. For example, many of the sub-classifications in Life Sciences include clinical elements which would be more logically sub-classified in Medicine, such as fields for Psychiatry in both Medicine (Code 2738) and Neurosciences (Code 2803).

g. There is a pressing need for a solution to the challenge of Multi-disciplinarity in journals and other sources, for examples with books with chapters on multiple subjects.

The Particular Challenges of the "Multidisciplinary" Classification

When all journals have been checked and reclassified and re-allocated as necessary to appropriate subcategories within the ASJC, there will still be a large number of journals which cannot be allocated to specific subject fields, in consequence of their multidisciplinary content. These journals will remain with "Multi- disciplinary" or "not otherwise classifiable" ASJC codes.

This is not a wholly new problem of journal classification, In 1999, W. Glänzel and colleagues (Glänzel et al 1999) reported an item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. They noted a serious shortcoming of bibliometric studies in the *(Social) Science (s) Citation Index* through the lack of an universally applicable subject classification scheme.

Sub-categorisation of Multidisciplinary Journals

The generic "multidisciplinary" classification is itself in need of much more rigorous analysis and subdivision. "Multi-disciplinarity" works at a number of levels in journal classification.

A. At the highest level, the content is so broad as to defy any form of subject classification, as with current "Mega-journals" such as the journals Heliyon, PLOSOne, or F1000Research.

B. Next down are journals which provide wide subject field within any one super-group. High profile journals such as Nature, Science and Cell might be considered in such a class. In some cases, publishers of such journals have accommodated the breadth of subject content in the flagship journals by moving to greater subject specialisation in the creation of topic-specific publications, such as "Nature Genetics". Many university and institutional journals provide similarly broad content coverage on a smaller scale.

C. Next down are journals which cover a wide variety of subjects within any one major subject area. For example, within the subject field of Medicine, there are journals such as the British Medical Journal (BMJ), The Lancet; and The New England Journal of Medicine (NEJM), which provide with subject coverage within the Medicine classification. However, their coverage also overlaps with various other subject areas in the Health, Life and Social Sciences areas.

D. Next down are journals which provide broad subject coverage within a particular professional or academic field of endeavour. For example, as a surgeon, I would find it helpful to see journals of Surgery further subclassified into more detailed subject categories, such as plastic surgery, neurosurgery, colorectal surgery and hepatobiliary surgery, for which significant numbers of journals exist in each subject area.

E. At the most detailed subject-specific level, with the least multidisciplinarity, a journal will describe its intended content in its title and adhere to that intent in its actual content. Such journals are often affiliated to specialist societies, for example The European Journal of Heart Failure.

Moreover, the level of multi-disciplinarity may or may not be discernible from the journal title. It may be wholly indecipherable in a journal with a non-specific title such as Heliyon, while it may be obvious elsewhere. For example, the cross disciplinary journal Computational and Systems Oncology (E-ISSN 2689-9655) self evidently addresses both Computer Sciences (Physical Sciences) and Cancer (Medicine, Health Sciences).

"Proportionality" and Multi-Disciplinary Content Mix

A further problem with the classification of multi-disciplinary journals lies in understanding the quantitative mix of content from different subject fields.

Multi-disciplinarity may vary between a collection of articles which fall into one ASJC subject super goup (eg Life Sciences); or which fall into multiple subject areas across supergrups. The article mix across the various subject fields may also vary considerably over time within any one journal, and from one journal to another.

Machine Learning and Improving the Multidisciplinary Journal Classification

In this essay, I have sought to highlight an important classification challenge with the help of a series of technical experiments by colleagues in the Elsevier SCOPUS Analytics Team using machine learning techniques, and with particular reference to my own operational needs as the Subject Chair for Medicine for the SCOPUS Content Selection Advisory Board.

The first series of studies under the direction of Dr Rob Schrauwen helpfully validated the observations around the lack of have alignment between journal titles and their article content, using machine learning and AI techniques.

However, Sujit Pal went on to demonstrate just how complex are the mathematics, and probabilistic tools which are needed accurately to further classify individual titles into accurate ASJC codes. It is difficult to avoid the conclusion that this aspect of any reclassification project would be better undertaken primarily using the intelligence and experience of human subject experts.

Considerations of an Article Level Classification Scheme

The complexities of refining the academic subject journal classification scheme into a clear, simple and comprehensive system have encouraged the exploration of article level classifications in SCOPUS, in the Web of Science and in other bibliometric reference systems. Medline uses only an article level search system.

The evolution of Machine Learning (ML) and Artificial Intelligence (AI) systems provides an opportunity in theory for the re-classification of journal content at the article based level, using the wording of the titles, the abstracts the authorship and possibly even the full text content of every article to discern the core theme of the article and to populate the information in various ways.

A more radical option than modernisation of the ASJC might therefore be to abandon it entirely and to adopt an article based classification scheme. The questions then arise as to what would be the purpose of an article based scheme? How would it differ from the journal classification scheme? How granular would it be?

Article based searches, whether on free-to-use systems such as Google, Google Scholar, Medline/PubMed and Open Alex, or on firewall-protected systems such as SCOPUS, Web of Science and Dimensions, are long established, and it is not clear that a specific article classification scheme would have any particular use..

Ludo Waltman and Nees jan Van Eck tried unsuccessfully to classify journal content and to compare proprietary classification systems using mathematical and statistical approaches at the level of article bibliometrics (see Waltman and van Eck 2012).

Jing Zhang and colleagues from the Chinese Academy of Sciences (2016) (10) have also sought to refine journal subject classification schema using measures of Journal coupling strength, which is a bibliographic measure of how closely related two documents are based on the number of shared references, and text mining of keywords.

Prashasti Singh and colleagues of the Department of Computer Science, at Banaras Hindu University, Varanasi, India, considered that the classification of research articles into different subject areas is still an important task in bibliometric analysis and information retrieval (Singh et al 2020).

They noted that the Dimensions academic database that uses article-based classification scheme. However, they considered that such a classification was no better than journal-based subject classifications, as used in SCOPUS and the Web of Science.

Maxine Rivest and colleagues noted equivalence in the performance of a deep learning approach with graph-based bibliometric approaches, and that machine learning approaches remained as yet no better than manual classification methods (Rivest et al 2021).

The Assignment of ASJC Codes based upon Author Profiles.

A further modification of the application of ML/AI to article level content classification is to use the author profiles which are associated with each article. The analytical model assumes that the authors of any article will be most closely and consistently associated with the subject classification, thus allowing the article to be allocated to the most appropriate subject classification through author association (see Bayraktar et al 2023) . This hypothesis nevertheless requires rigorous testing, as many authors are multidisciplinary in their skills and interests.

Towards a Hybrid Article and Publication Level Classification System

No consistent journal classification has yet emerged using Machine Learning, and all such approaches have significant limitations. The trained human brain can process complex information and see solutions which may elude computer algorithms. Text analysis by machines depends upon the quality of the original writing and the clarity of communication of the titles, the abstracts and the full text content of an academic paper. The subtleties of interpretation may well be more fully understood by the human expert in a range of circumstances.

ML and AI technologies are rapidly evolving, and it may be that such a machine- directed methodology of classification will emerge which is both persuasive and widely adopted. ML and AI readily address the challenges of scale, as we move from hundreds and thousands of journals to tens of millions of individual documents. However, for the foreseeable future, it is reasonable still to focus on simpler and more obtainable objectives, which include modernisation of the existing ASJC in a human expert-led project.

Despite its deficiencies, the ASJC will not be abandoned in the near future, not least as it represents a substantial financial and technical investment which is interlinked with other operational systems. We should nevertheless consider the properties of an updated version of the ASJC which would permit reclassification of the existing corpus of academic journals, and embrace the various and many other forms of formal academic outputs.

A phased approach to the modernisation of the ASJC

A phased approach to the modernisation of the ASJC using through human-computer teaming would involve the following steps:

1. The development of a software tool which would sequentially present each journal in each ASJC code to one or more human experts in the subject area, who would have a series of options, vis:

- to accept the existing ASJC code
- to select an alternative ASJC code from the existing portfolio
- to apply a new subject name, for formal allocation of a new ASJC code when all suggestions have been polled on completion of the review process.
- on agreement of the new ASJC codes, journals would be appropriately reassigned.

This activity would correct any glaring errors in the existing coding and increase the granularity of the ASJC, while exposing more clearly the multidisciplinary journal cohorts for which further refinement of the classification was needed.

The cycle would then need to be iterated to extend the ASJC to other publication sources, including the large number of academic books, patents and conference proceedings which have been listed in the SCOPUS core collection.

Refinement of the content description of multidisciplinary journals

At this point of the process, machine analysis of the articles within multidisciplinary journals becomes important. This would permit categorisation of each multidisciplinary journal by its supergroup and subgroup coverage, and by the proportion of articles in any such group.

For example, an all subjects (supergroup) multidisciplinary journal might contain 25% each of articles from Physical, Health, Social and Life Sciences, while a single subject area multidisciplinary journal in the Life Sciences might contain 20% each of articles from Agriculture, Biochemistry, Microbiology, Neuroscience and Toxicology, and so on. Such proportions could be represented and used in many quantitative ways in reporting and analytics systems.

In Summary

In this essay, I have considered the challenges of updating then academic subject classification, vis a vis the development of an academic article-based classification system, and the current weaknesses to both approaches.

There remains a strong case for major updating of the ASJC across all subject areas, and for renaming it as the All Subjects (or Sources) (and) Journal Classification scheme. The core principles of ASJC system continue to be sound and useful in the conceptual organisation of tens of thousands of journals.

The ASJC offers the capacity for expansion and refinement of the subject codes and allocations without fundamental changes, major costs or operational disruption. There are plenty of “reserved” or unused four figure codes which could be activated to increase the usability and granularity of the system, and modernisation can be introduced incrementally.

Journals evolve and metamorphose in various ways, some of which are “legitimate” and others of which are “predatory”. The ASJC needs to be adaptable to significant changes in policy and content range in individual titles. For such reasons, any modern iteration of the ASJC will always need dynamic adaptations and updates.

Future developments with Source- and Journal based Classification systems

In daily practice, researchers will invariably seek out articles by topic and purpose rather than by prior search for the journals in which they are published, as was once necessary in the era of paper. Journals themselves are now also challenged by ePrint and preprint servers and by other "direct to public" communication systems. However, for so long as journals, conference proceedings and books remain the principal vehicles for quality assurance in academic publication, there will be a need for a reliable classification system.

It therefore seems likely and necessary that one or other version of the ASJC will need to evolve and adapt, and human expert input will be important. Ultimately, commercial considerations as to the value and importance of updating the ASJC, and the further evolution of ML and AI systems will determine the scale of any further investment and the form and rate of implementation of change.

It seems likely that eventually a collegiate multi-publisher and cross industry approach will be of particular value for the common good and for global standardisation purposes.

Acknowledgements

The opinions expressed in this essay are entirely those of the author and do not reflect or should be inferred as reflecting corporate opinion and policies of Elsevier BV. I am nevertheless grateful to Elsevier for inviting and supporting my participation in the SCOPUS journal evaluation programmes and discussions which have informed my present thinking on this subject.

I am particularly grateful to Dr Rob Schrauwen VP, Platform and Data Strategy Research Data Platform for Elsevier for support and input into this project, and for the data analysis and the figures in Section 1 of the Methods; and to Mr Sujit Pal, Technology Research Director for Elsevier Health Markets business team, for the project work described in Section 2 of the Methods.

References (in alphabetical order)

Anon: The Universal Decimal Classification Scheme
https://en.wikipedia.org/wiki/Universal_Decimal_Classification

ASJC: The Complete List of Codes (updated August 15th, 2024)
https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/

Aviv-Reuven, S., Rosenfeld, A. A logical set theory approach to journal subject classification analysis: intra-system irregularities and inter-system discrepancies in Web of Science and Scopus. *Scientometrics* **128**, 157–175 (2023). https://doi.org/10.1007/s11192-022-04576-3

Bayraktar M. Y. and Kaya M., "Author-Profile-Based Journal Recommendation for a Candidate Article: Using Hybrid Semantic Similarity and Trend Analysis," in *IEEE Access*, vol. 11, pp. 45826-45837, 2023, doi: 10.1109/

Glänzel, W., Schubert, A. & Czerwon, H.J. An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. *Scientometrics* **44**, 427–439 (1999). https://doi.org/10.1007/BF02458488

Rivest M, Vignola-Gagné E, Archambault É. Article-level classification of scientific publications: A comparison of deep learning, direct citation and bibliographic coupling. PLoS One. 2021 May 2011; 16(5): e0251493. doi: 10.1371/journal.pone.0251493. PMID: 33974653; PMCID: PMC8112690.

Singh, P, Piryani, R, Singh, VK and Pinto, D. Revisiting subject classification in academic databases: A comparison of the classification accuracy of Web of Science, Scopus & Dimensions. Journal of Intelligent & Fuzzy Systems, vol. 39, no. 2, pp. 2471-2476, 2020 DOI: 10.3233/JIFS-179906

Stević A. Scopus is broken – just look at its literature category. blog post on Retraction Watch https://retractionwatch.com/2024/07/17/scopus-is-broken-just-look-at-its-literature-category/

Thelwall, M. and Pinfield, S. (2024) *The accuracy of field classifications for journals in Scopus.* Scientometrics, 129 (2). pp. 1097-1117. ISSN 0138-9130

Waltman L and van Eck NJ: "A new methodology for constructing a publication-level classification system of science" 2012 in Arxiv and Vol 63, The Journal of the American Society for Information Science and Technology (DOI 10.1002/asi.22748)

Wang Q and Waltman L. Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus,
Journal of Informetrics, Vol 10, 2, 2016, pp 347-364, doi.org/10.1016/j.joi.2016.02.003.

Web of Science Core Collection subject categories (accessed 28th October 2023)
https://webofscience.help.clarivate.com/en-us/Content/wos-core-collection/wos-core-collection.htm?Highlight=Subject%20Categories

Zhang J, Liu X and Wu L.
The study of subject-classification based on journal coupling and expert subject-classification system. *Scientometrics*, volume 107, pages 1149–1170 (2016)