Crowdsourcing classification and causality to power a search-and-innovation engine

Richard Absalom
Independent Researcher, Luxembourg

Abstract

Durham Zoo (DZ) is a project to create a search-and-innovation engine for science and technology. The engine has been designed to capture the knowledge of experts from different areas of expertise via the classification of the literature. The architecture combines the higher-level cognition of humans and their powers of language, abstraction, of inference and analogy, with the storage and processing power of computers. The system is adapted for searching both what already exists, and novel solutions to problems. To be built and operated by the community, the goal is to democratize innovation whilst funding societal causes such as climate-change mitigation or the search for new antibiotics. The original design, first published in Absalom and Absalom (2012), relied upon fuzzy and faceted classification. The fuzziness related to the similarity of concepts in each of the facets. A search query would be matched with the literature in multiple facets to retrieve holistically similar literature, or to suggest a solution to a problem from elsewhere in technology or the natural world. The facets used to describe a concept in science and technology included a problem and a solution. A recent reappraisal of the project design recognised the potential of causality for modelling and matching problems. This paper proposes a design compatible with the crowdsourced classification.

1.0 The original motivation for the project

The initial challenge was how to crowdsource the classification of the literature in support of patent searching. Patent offices still use classification by experts as a cornerstone to searching the prior art. For whilst information retrieval and artificial intelligence are making great strides, the human brain is still class leading at identifying and understanding concepts. It is still best able to work through imperfect language, abstraction, jargon, and terminology to distill the essence of a disclosure. The essence of analogous concepts is encoded with a same classification code. In the patent world there is the added complexity of the legal nature of a patent: things are often described in broad terms so as not to restrict the scope of protection. As an example, a magnetic disk drive may be described as a “storage device”. This can complicate search using keywords.

Rapidly increasing numbers of patent applications and a massive increase in the scientific literature resulted in a scalability problem. Could automatic classification produce the goods? Could not patent applicants and authors better classify their own disclosures?

Much patent office classification is based to a greater or lesser degree on the International Patent Classification (IPC). The IPC is a fantastic resource, the result of the considerations of experts over many years. Unfortunately, the complexity, the classification rules and the esoteric patent-speak constitute a barrier to entry as regards a crowdsourcing effort. Could we not design something simple and intuitive?

The IPC’s origins as a paper classification scheme with a hierarchical tree structure restrict its ability to evolve with technology. Digital convergence saw an increase in ‘sameness’ between computing, on the G root class, and telecommunications and television on the H root class. This was reflected by an increasing overlap between the two classes and much dual classification. Dual classification is not a problem *per se*, however creeping uncertainty and ambiguity in classes has consequences for precision and recall. The multidisciplinarity of nanotechnology, combining all manner of physical sciences, life sciences and engineering from across all the A to H root classes complicated matters further. Was the tree structure not the root of the problem?

The paper-classification origins have resulted in a limited use of faceting, perhaps with the exception of the Japanese Patent Office’s electronic implementation of the IPC. The use of “on-the-shelf-or-not” Boolean classification fails to represent the degree of sameness of different concepts. This is better done with fuzzy classification. Fuzzy mathematics can return a ranked list of hits to a search query. What of a fuzzy and faceted classification scheme?

How to manage a complex classification scheme? The IPC required experts to assemble and discuss both the “what is what” and the ‘what goes where” of new technologies. This requirement for centralised management is incompatible with a distributed crowd of individuals working independently.

2.0 The basic design for single concepts

The design process proceeded in ignorance of knowledge organisation theory and terminology. The terminology used here is neither the original terminology used, nor that adhering to a standard. However, the basic design has much in common with a faceted thesaurus and where possible consistency with the ANSI/NISO standard has been sought.

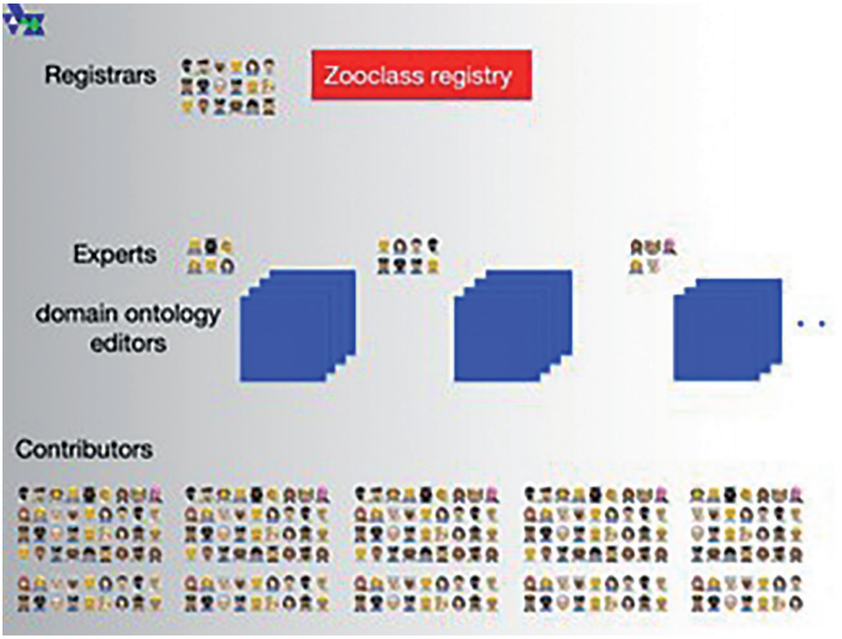

Figure 1. Three-tiered structure

The design has three actors, Registrars who manage a controlled vocabulary, Experts who use the controlled vocabulary to build domain ontologies, and a crowd of Contributors who are invited to classify and search the literature, see Figure 1.

It was decided that defining "what is what" with a controlled vocabulary was essential to the design: experts from different areas of expertise commonly interpret a same terminology differently or assign a new name to an established and accepted terminology. The controlled vocabulary terms are called Zooclasses, abbreviated to Zoocs, and define concepts.

Experts from different domains are encouraged to submit proposals for new Zoocs to a Registrar. The Zooc may be very similar to existing Zoocs, however the differences need to be clear. After a successful peer review, including feedback from the proposers of any similar Zoocs, it is accepted into the Registry. Whilst centralised management was to be avoided there appeared no other way of limiting ambiguity and overlapping classification.

The requirement for simple and intuitive navigation through the classification resulted in a graphical representation. Zoocs are not presented singly, but are displayed together with narrower, broader, and related terms on a simple ontology called a Zooc Steering Diagram (ZSD), see Figure 2a to 2d.

It is the expert or group of experts that proposed the Zooc who assume the responsibility for the development of the corresponding ZSD: picking and placing Zoocs from the Registry to create a representation of their domain of expertise. The experts decide "what goes where" from their perspective.

The crowd is invited to navigate through the library of ZSDs to find and attribute Zoocs to the literature. The ZSD for the Pipe Zooc is shown in Figure 2a. The Pipe Zooc is called the Subject of the ZSD. Beneath the Subject are narrower terms that we call Types, in our example Pipes for liquids, gases and structure. Above the Subject are terms that are similar to the Subject. Collectively they are known as Sims. Sims can be unrelated terms that are similar to the Subject in a holistic manner: for example, the Tunnel and Trough Sims. Alternatively, Sims are similar due to their being a broader term, effectively a hypernym that we abbreviate to Hype. The Pipe Sim is such an example in the Pipe for liquids ZSD in Figure 2d. Polyhierarchies require multiple ZSDs and disambiguation.

A mouse pointer hovering over a Zooc will reveal the Zooc metadata and scope notes: see Figure 2b. Clicking on a Zooc loads its ZSD via a hyperlink: see Figures 2c to 2d. The display of related terms in a simple structure, of available scope notes and hyperlinking is we believe a simple and intuitive user interface.

The Sims are placed on the ZSD as a function of their similarity. So, the lower down the ZSD, and thus closer to the Subject they are, so the more similar they are to it. Whilst not shown, the sliding scale of similarity represents the weighting of a fuzzy classification. For the Pipe ZSD, shown in Figure 2a, a Tunnel is 40% similar to a Pipe, whilst a Trough is 20% similar. The Types on the other hand are all 100% Pipe.

For the purpose of search, we can choose to select the ZSD, now called a Zooc Similarity Diagram, rather than just the Subject. In our example we can expand a single Zooc query to include Pipes, Types of Pipe and Sims of Pipes, each with their related degree of similarity. This query expansion is analogous to the semantic query expansion of Tudhope and Binding (2008).

There is also an implicit expansion in terms of classification. As an example, the classification of a disclosure with the Tunnel Zooc will classify it as the Subject of the Tunnel ZSD, but also as 40% similar to a Pipe, given that it appears on the Pipe ZSD. If attributed to a disclosure it will be classified as many times as the selected Zooc appears on a ZSD, each time receiving the degree of similarity judged by the expert or experts in the field.

The ZSD is an ontology with the unique relationship of similarity. This is a key characteristic to being able to join up independently created ZSDs into what could be called a knowledge graph. The other characteristic is the fractal-like nature of the ZSD. Each Zooc that appears on a ZSD has its own ZSD "hidden" underneath it. These ZSDs in turn have Zoocs that have

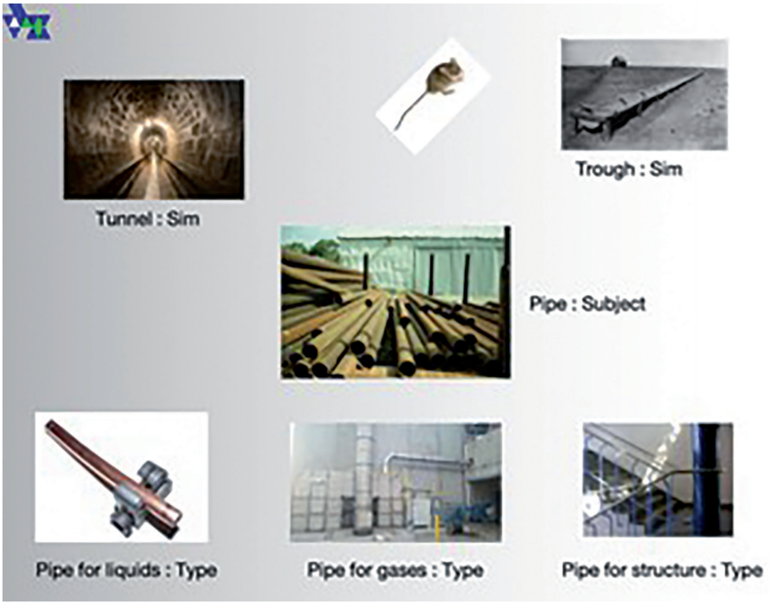

Figure 2a. Zooclass Steering Diagrams

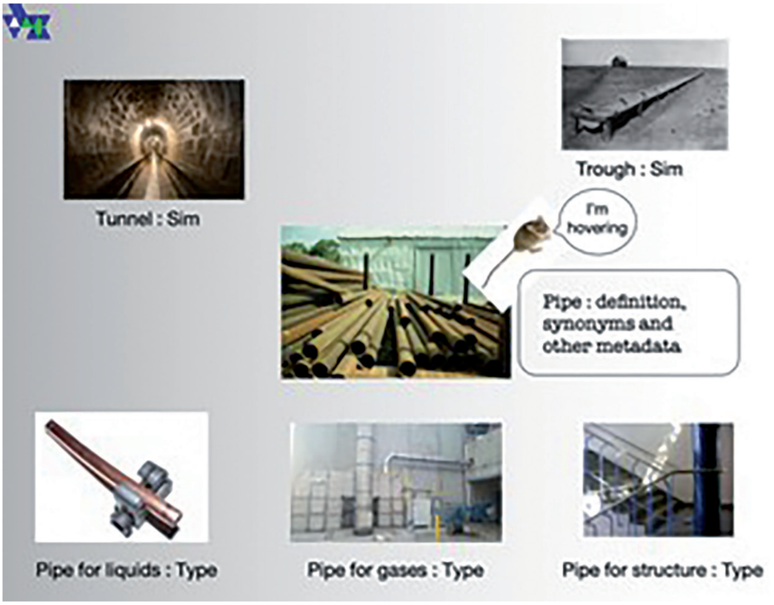

Figure 2b. Zooclass Steering Diagrams

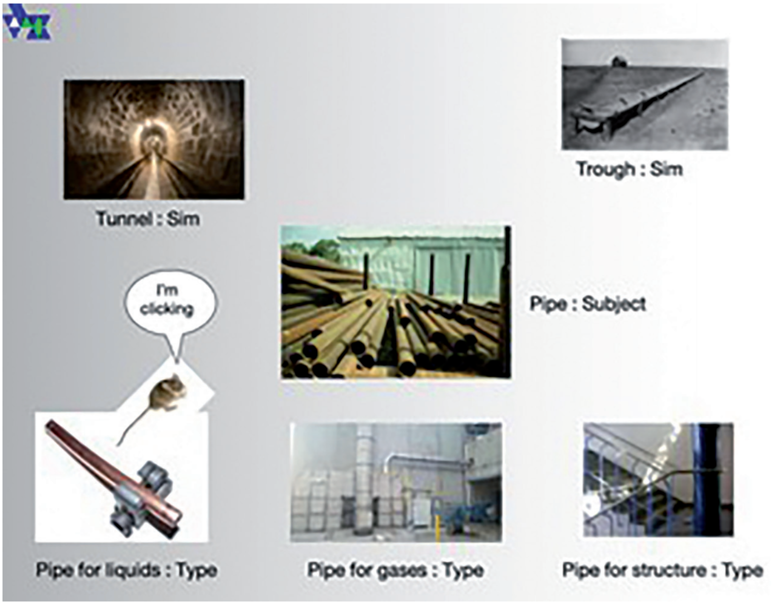

Figure 2c. Zooclass Steering Diagrams

Figure 2d. Zooclass Steering Diagrams

ZSDs and so on down the levels, zooming down to reveal more detail akin to a Mandelbrot fractal.

As an example, the Tunnel Sim that appears on Figure 2a will have all the Tunnel Types and Tunnel Sims on it as decided by the Tunnel Experts. The Pipe Experts can leverage this work during a search. The Tunnel Sims are similar to something similar from the perspective of the Pipe. The unique relationship enables the overall similarity to be calculated with a simple algorithm as published by Absalom and Absalom (2012).

The fractal representation also serves to fill in the gaps of higher-level ZSDs. For example, if the Tunnel ZSD did not have the Trough Sim on it, but it included the Pipe Sim, the Trough Sim would be picked up on the first fractal level of the Tunnel ZSD via the Pipe Sim.

Going down the levels the algorithm will discover Zoocs that it has already seen before. The algorithm takes care of eliminating these repeat appearances: the first appearance with the fewest iterations of similarity is assumed to be the most accurate. Importantly the algorithm is run from each Subject Zooc enabling a knowledge graph to be developed with the perspective of each Subject Zooc expert.

Experts could map classification completed in different classification schemes onto their ZSDs. A perfect Zooc match could be placed alongside it, a close match placed in horizontal proximity as a function of the similarity.

As science and technology evolves so new Zoocs and ZSDs can be created. The newly created ZSDs can incorporate existing Zoocs and their ZSDs. Updating of existing ZSDs is distributed amongst the different experts.

Zoocs have living, zombie or dead status. Zombie status is attributed to a Zooc that has been superseded. Zombie classes cannot be attributed but are included for search for as long as all the literature that received the previously living class has not been reclassified. As an when reclassification is completed the zombie becomes a dead class maintained for information only.

Contributors only need to classify and search with Subject Zoocs, the cognitive task of estimating similarity with other concepts having been performed by the different domain experts.

To be granted a patent an application must be both new and non-obvious in relation to the prior art. A Subject Zooc and its Types can define all the different manifestations of a concept. This is useful for searching whether something is new. Non-obviousness excludes concepts that are too similar and is well served by the Sim design.

3.0 Selecting and combining multiple facets

Although originally intended as a search engine for patent and non-patent literature prior art, it was recognised at an early stage that the prior art is also a source of inspiration for new ideas. What facets would best support a search-and-innovation engine?

Problems and their solution are used to evaluate non-obviousness during patent prosecution. And many dormant patents have been described by Hartmann (2014) as solutions looking for a problem.

Cross-industry innovation applies known solutions to analogous problems to a different application or a different field of technology. Biomimetics seeks inspiration to unsolved problems from the natural world. Fortunately, analogy is similar to the similarity of the ZSD.

Elsewhere a problem may become a solution: the scanning tunnelling microscope's problem of attracting atoms from a surface under investigation provided the solution to the picking and placing of individual atoms was disclosed by Stroscio and Celotta (2004). The peelable adhesive that eventually found good use on 3M Post-it® notes is another good example disclosed by Hiskey (2011).

And so, an initial facet design was chosen including Solutions to a Problems in a context defined by an Application, Technology and Operation. As an example, a Fedora hat takes up too much room in the wardrobe. The Fedora hat is the Application or product. The Problem is the Fedora being too voluminous. The technologist called upon to investigate the Problem is an expert in textiles Technology. The Operation relates to the storage of the Fedora. The Solution is an improved memory effect in the textile of the hat that allows it to be collapsed flat for storage and then reformed for use, and this over the lifetime of the hat. This facet structure was abbreviated as ATOPS.

Zoocs from any-and-all ATOPS facets can be included in a query. The selected Zoocs are typically expanded to the Zooc ZSDs including the Subject, the Types, and the Sims, each of the latter with their similarity values. The fractal algorithm can continue the expansion as explained earlier. The resulting similarity lists are then matched with the ATOPS of the classified literature. The magnitude of any-and-all matches of the facets are combined as vectors in different facet dimensions. Combining the vectors provides a ranked list of holistically matched literature. Whilst the fuzzy values in all-and-any of the ZSDs remain somewhat arbitrary, the imprecision across multiple facets is of less concern: literature that matches in multiple facets is expected to rise high in the ranking.

Our first use case disclosed by Absalom, Absalom, and Hartmann (2012) considered the search for a stent, an artificial tube used in medicine to keep a body tube open as our Application. The Problem was stent thrombosis, where the stent becomes blocked. The Solution was a non-smooth surface as a lining.

We wondered if a simulated sharkskin lining to the stent would prevent material sticking to it in the same way a simulated sharkskin coating prevents the fouling of ship's hulls. Not having a corpus of classified literature our considerations remained hypothetical. In terms of the innovation engine, we imagined a situation where a catheter, a medical device similar to a stent as the Application, with the similar Problem of bacterial deposition, and with the Solution of a pimpled lithographed surface could have stimulated the non-smooth sharkskin lining, had it not existed. Whilst there would be no perfect match in either of the Application or Problem facets the catheter disclosure would rank highly. The pimpled surface would likely prove food for thought and alternative non-smooth surfaces considered.

A review of a mini pilot conducted in 2014 highlighted shortcomings of the basic ATOPS structure. We considered designs incorporating additional facet complexity. However, any theoretical increase in information-retrieval power from such increased complexity need take account of the increased cognitive burden on Registrars, Experts and Contributors. Would a theoretical increase in information-retrieval power be met in practice? Whilst simplicity is the ultimate sophistication,[1] everything should be made as simple as possible, but not simpler.[2]

We have considered alternative representations to enhance the design, including the provision of meronomy We have considered if-and-how artificial intelligence could bridge the gaps in ATOPS. We have also considered how to develop the wisdom of the crowd from multiple independent classifications of a same disclosure.

Recently we reviewed the stent example and realised that the Problem of bacterial deposition is better described as a cause of stent thrombosis than being similar to stent thrombosis. The rest of this paper will present a design for modelling such causality to enhance DZ.

4.0 Incorporating causality

Causality is complex. Studied in metaphysics as part of contemporary philosophy it was used by Robb (1911) to construct notions of time and space. More generally it sits at the intellectually demanding conjuncture of philosophy, physics, and mathematics, and has occupied many brilliant minds over thousands of years.

---

[1] Attributed to Leonardo da Vinci.

[2] Attributed to Albert Einstein.

Not unsurprisingly there are different schools of causality: regularity, probabilistic, counterfactual, mechanistic, and manipulationist.[3] All the schools require study. Such an investment is incompatible with even the most erudite and committed of crowds.

We require something simple and intuitive and would trade accuracy and academic rigour to meet these requirements. But it needs to support the search-and-innovation engine.

We looked to engineering. Root cause analysis (RCA) is used in both science and engineering to model the origins of problems and help find their solution. RCA uses causal graphs, where nodes representing causes and effect are joined by arrows to model their sequence in time. From a mathematician's perspective they are a directed acyclic graph (DAG).

That time travels uniquely in one direction with a cause necessarily preceding an effect makes the structure suitable for causality. Causal graphs and DAGs exclude directed cycles, where a cycle can travel forward in time but remain in a loop. In directed cycles an effect is a function of both the cause and of history. A simple example would be a waste bin that is push-to-open and push-to-close. Ignoring directed cycles maintains simplicity with minimal negative consequences for our design. RCA suggests the use of the Ishikawa diagram, a simple DAG, to brainstorm the root causes to a problem.

5.0 The Ishikawa Cause and Effect Diagram

Commonly called a fishbone diagram, Ishikawa (1968) designed diagrams to aid investigation into the causes of an effect or problem. An example is shown in Figure 3 below.[4] The first step in completing a diagram is the brainstorming of the different categories of cause, shown as the large bones along the fish's spine. Primary causes are then identified within each category and represented as smaller bones feeding into the large bones. The process is repeated identifying secondary causes that can cause the primary causes and be represented as even finer bones. The method terminates at the identification of the root causes of the problem. A repeated asking of why causes are produced, a technique formalised in Serrat (2017) as the 5 Whys, often accompanies the process. The Ishikawa diagram is simpler and more intuitive than causal graphs.

---

[3] "Causality," Wikipedia, last edited February 1, 2022, https://en.wikipedia.org/wiki/Causality.

[4] "Ishikawa diagrams," Wikipedia, last edited December 29, 2021, https://en.wikipedia.org/wiki/Ishikawa_diagram.

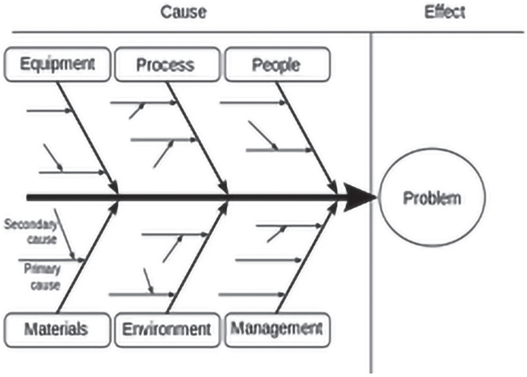

Figure 3. An Ishikawa Diagram

6.0 The Modified Ishikawa Diagram (MID)

Most patent and academic literature is not concerned with identifying and mitigating all the potential causes to a problem. Developing a single definitive diagram for a problem would likely be a major investment, even if done collaboratively. And different contexts may complicate and even compromise a resulting diagram.

It was therefore decided that the representation be unique to a particular disclosure, in a manner akin to document classification, or perhaps more akin to annotation.

Causes and effects can be desirable or nefast according to their context. Physics has principles. As an example, the Bernoulli *principle* relates an increased rate of horizontal flow with a reduction in pressure. The related Bernoulli *effect* underpins winged flight: very much a solution. It also produces the squat effect whereby the difference in speed of water passing underneath and aside a ship's hull in shallow water creates a downforce. This was a problem for the ship called the QE2 as disclosed by MAIB (1993), when it caused it to run aground.

From our perspective the causes and effects that end in a problem can all be viewed as problems. These can all be Zoocs in the Problem facet and as such the Bernoulli effect could be a Problem Zooc. In contravention with good thesaurus practice the Bernoulli effect can also be a Solution Zooc and be used to model Solutions: but this is the subject of a future paper.

A causal sequence is represented in a similar manner on the MID shown above. The end problem, shown here as PROBLEM at the base of the diagram, has a vertical timeline t=0. Direct problems A, B, C and D are placed

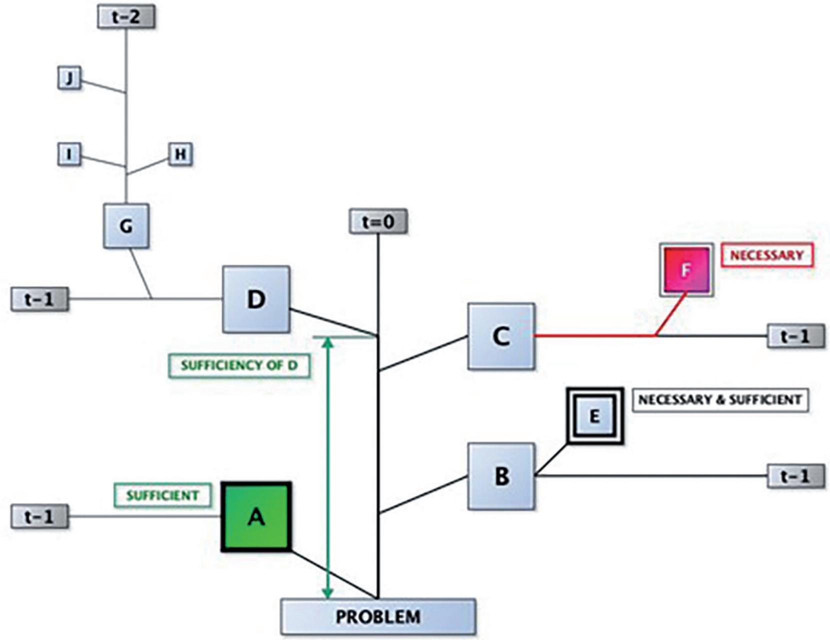

Figure 4. A Modified Ishikawa Diagram (MID)

alongside the t=0 timeline. They also appear at the base of their own t-1 timelines, indicating that they occur prior to the end problem. Intermediate problems, such as G in Figure 4, are placed alongside the t-1 timelines and have at their base a t-2 timeline. The process is repeated until all root problems are displayed, in our example problems A, E, F, H, I and J. It may well be the case that a disclosure only focuses on a single causal sequence.

In so far as they exist the problems should be represented as Problem Zoocs. If they do not exist, they should be referred to the Registrars for consideration.

Ishikawa diagrams have attracted criticism from Gregory (1992) for not including the logic of causation, and in particular the lack of provision for necessity and sufficiency conditions. We have included them on the MID.

Sufficiency is typically considered as a Boolean variable: a cause is sufficient or not sufficient to produce a subsequent cause. We interpret sufficiency as a continuous variable where the sufficiency of a problem, effectively its ability to produce the subsequent problem, is represented by the height from the base of the subsequent problem's timeline. Direct Problem A, shown in green at the base of the t=0 timeline is sufficient, meaning that every time it occurs it will produce the problem. Problem D is only sometimes sufficient, perhaps 3 times in 10, against 4 times in 10 for Problem C and 8 times in 10 for Problem B. Insufficient problems, those that cannot produce the subsequent problem, as not shown on the diagram as they do not represent a problem.

It is important to distinguish the sufficiency of a problem from the likelihood it will arise in the first place. There is no graphical representation of such likelihood on the MID. Unlike engineering or science modelling we are not seeking to quantify a problem, to determine a failure rate or calculate the risk or severity of an outcome. Existing frameworks exist to serve such purposes. Fault Tree Analysis (FTA) uses Boolean logic and statistical probabilities of component failures to model how systems fail, and to design mitigation strategies. Failure Mode and Effects Analysis (FMEA) and its derivative Failure Mode, Effects and Criticality Analysis are related and popular frameworks.

That said, a MID could eventually be annotated with a simplified likelihood of occurrence, such as FMEA's extremely unlikely, remote, occasional, reasonably possible, and frequent scale. Alternatively, a colour-coded log scale could be developed, or the length of the connection between the subsequent problem's timeline and the causal problem could be used. However, likelihood is often represented by a statistical distribution.

Quantifying sufficiency is also difficult given that there may be many variables. Braking hard in a car can result in the wheels locking if the car does not have ABS brakes. Braking harder will increase the sufficiency. The sufficiency will also be higher on a wet road than on a dry road, and even light braking may be fully sufficient on icy roads. As an alternative example imagine a glass of water on a table in a train. It could be sent flying if the train hits an obstacle on the line. The same could happen from a large jolt of the train due to a damaged track. A worn track could produce a series of jolts with the same result. And a poor-quality track could produce a vibration to do the same. Such single, multiple, and constant problems of differing magnitude are easier to define than they are to quantify. That said, combining likelihood with an empirical quantification of sufficiency could produce a simplistic ranking of how big a problem the different causal sequences are.

A necessary cause is shown in red in Problem F on Figure 4. Problem C can only happen in response to Problem F. Of note is that Problem F is alone on C's t-1 timeline. Identifying necessary causes is useful in that mitigating them removes the subsequent problem. More exactly the necessary problem needs to be prevented rather than cured. Of note is that Problem E is both necessary and sufficient.

The MID serves to link related problems for the purposes of search and innovation. Annotating the different Zooc Problems with related ATOS

Zoocs would appear a worthwhile endeavour. The basic ATOS, and especially the Operation facet, requires sub-faceting to facilitate this. Sub-facets are required to identify the different stages of a product lifecycle and well as operations *per se*. A work in progress.

The MID can supplement search with ZSDs. A search query, with all-or-any ATOPS facets could proceed using ZSDs and the fractal algorithm as per the original design, whilst at the same time searching for the query Zoocs in the MIDs, including the MID Zoocs proper and ATOPS annotations. As an example, the disclosure with the MID of Figure 4, could match the Problem Zooc of the query with Problem F, where Problem F had not been attributed as a classification in the original manner. Related solutions to F could appear in the text of the Problem F disclosure and/or be included as ATOPS annotations to the MID.

The MID can be transposed into an equation. Direct problems A, B, C and D can all produce the end problem. Described in Boolean logic they have an OR relationship that is represented by a + symbol. A causal sequence can be represented as a sequence from end problem back to the root problem separated with commas. The sequence goes back in time from left to right. We have chosen to represent sufficiency with ^ and necessary with ! symbols.

PROBLEM = {A^} + {B, E!^} + {C, F!} + {D, G, H+I+J}

We are currently studying methods to match MID equations that have been developed for different documents. This could leverage algorithms from natural language processing such as spell-checkers. However, we have started with disassembling equations into triples of causal problem, effect problem and the sequence distance between them.

7.0 Contributory causes

Contributory causality can model situations where multiple causes can together cause an effect. Mackie (1974) proposed the INUS model where contributory causes are Insufficient but Non-redundant parts of a condition, which is itself, Unnecessary but Sufficient for the occurrence of an effect.

An example in Wikipedia[5] describes a short circuit, the proximity of flammable material, and the absence of firefighters being INUS conditions for a house burning down.

We prefer that sufficiency be represented as either sometimes sufficient or fully sufficient as on the MID. Contributory causes also need be represented

---

[5] "Causality," Wikipedia, last edited February 1, 2022, https://en.wikipedia.org/wiki/Causality.

as individually insufficient. We see the commonality between contributory causation and fuzzy logic and are presently investigating methods to integrate it with MID diagrams and MID equations. Our goal is to develop an all-encompassing MID-logic.

8.0 Conclusions

We believe that the work presented here lays the foundations for further study. A crowd will eventually be necessary to classify a corpus of literature to be able to test the ideas. In the shorter term there remains much to do: with the facet complexity, with developing the wisdom of the crowd from multiple independent classifications, and with MID-logic. The use of artificial intelligence to process the causal relationships of the MIDs, grafting them together to build larger and more complex models remains a very distant goal. We remain open to ideas and collaboration.

References

Absalom, Arthur and Geoffrey Absalom. 2012. "Durham Zoo: prior art and solution search." Durham Zoo. Last accessed February 10, 2022, http://durhamzoo.org/wp-content/uploads/2012/11/Durham-Zoo1.pdf.

Absalom, Arthur, Geoffrey Absalom, and Dap Hartmann, D. 2012. "A collaborative classification-based search engine for prior art and solution search." In *2012 IEEE Sixth International Conference on Semantic Computing (ICSC 2012) 19-21 September 2012 Palermo*, *Italy*. Institute of Electrical and Electronics Engineers (IEEE), 87-92. https://doi:10.1109/ICSC.2012.30.

"Causality." Wikipedia. Last edited February 1, 2022. https://en.wikipedia.org/wiki/Causality.

Gregory, Frank. 1992. "Cause, Effect, Efficiency & Soft Systems Models. Warwick Business School Research Paper No. 42". *Journal of the Operational Research Society* 44, no. 4:333-44. https://doi.org/10.1057/jors.1993.63.

Hartmann, Dap. 2014. "Turning Technology into Business Using University Patents." *Technology Innovation Management Review* 4, no. 12:37-43. https://doi:10.22215/timreview/856.

Hiskey, Daven. 2011. "Post-it notes were invented by accident." *TodayIFoundOut*. Last accessed February 10, 2022. http://www.todayifoundout.com/index.php/2011/11/post-it-notes-were-invented-by-accident/.

"Ishikawa diagrams." Wikipedia. Last edited December 29, 2021. https://en.wikipedia.org/wiki/Ishikawa_diagram.

Ishikawa, Kaoru. 1968. *Guide to Quality Control*. Tokyo, Japan: Asian Productivity Organization.

Mackie, John Leslie. 1974. *The Cement of the Universe: A Study of Causation*. Oxford, UK: Clarendon Press.

Marine Accident Investigation Branch (MAIB). 1993. "Queen Elizabeth 2 - grounding of a passenger vessel on 7 August 1992".

Robb, Alfred A. 1911. *Optical Geometry of Motion*. Cambridge, UK: W. Heffer.

Serrat, Olivier. 2017. "The Five Whys Technique," *Knowledge Solutions*, 307-10. https://doi.org/10.1007/978-981-10-0983-9_32.

Stroscio, Joseph A. and Celotta Robert J. 2004. "Controlling the Dynamics of a Single Atom in Lateral Atom Manipulation." *Science* 306, no. 5694:242-47. doi:10.1126/science.1102370.

Tudhope, Douglas and Ceri Binding. 2008. "Faceted Thesauri." *Axiomathes* 18, no. 2:211-22. https://doi:10.1007/s10516-008-9031-6.

Licenses for the use of images (lack of place with the images themselves)

Figures 2a to 2d:	Mouse pointer: George Shuklin, CC BY-SA 1.0 <https://creativecommons.org/licenses/by-sa/1.0>, via Wikimedia Commons Tunnel: Thomas Bresson, CC BY 2.0 <https://creativecommons.org/licenses/by/2.0>, via Wikimedia Commons Pipe for liquids: Tommy Halvarsson, CC BY-SA 3.0 <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons Pipe: Paul Goyette, CC BY-SA 2.0 <https://creativecommons.org/licenses/by-sa/2.0>, via Wikimedia Commons Pipe for gases: Combustion2016, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons Pipe for structure: Holger.Ellgaard, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons Collapsible: Thiemo Schuff, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons Flexible: Paul Goyette, CC BY-SA 2.0 <https://creativecommons.org/licenses/by-sa/2.0>, via Wikimedia Commons
Figure 3:	Ishikawa fishbone-type cause-and-effect diagram: FabianLange @ de.wikipedia <https://en.wikipedia.org/wiki/GNU_Free_Documentation_License>