Durham Zoo: Powering a Search-&-Innovation Engine with Collective Intelligence

Absalom, Richard; Hartmann, Dap

Publication date
2014

Document Version
Final published version

Published in
Social Technologies

ISSN 2029-7564 (online)
SOCIALINĖS TECHNOLOGIJOS
SOCIAL TECHNOLOGIES
2014, 4(2), p. 245–267

DURHAM ZOO: POWERING A SEARCH-&-INNOVATION ENGINE WITH COLLECTIVE INTELLIGENCE

Richard Absalom

Independent researcher, The Netherlands, durhamzoo@ymail.com

Dap Hartmann

Delft University of Technology, The Netherlands, L.Hartmann@tudelft.nl

doi:10.13165/ST-14-4-2-01

Abstract

Purpose *– Durham Zoo (hereinafter – DZ) is a project to design and operate a concept search engine for science and technology. In DZ, a concept includes a solution to a problem in a particular context.*

Design *– Concept searching is rendered complex by the fuzzy nature of a concept, the many possible implementations of the same concept, and the many more ways that the many implementations can be expressed in natural language. An additional complexity is the diversity of languages and formats, in which the concepts can be disclosed.*

Humans understand language, inference, implication and abstraction and, hence, concepts much better than computers, that in turn are much better at storing and processing vast amounts of data.

We are 7 billion on the planet and we have the Internet as the backbone for Collective Intelligence. So, our concept search engine uses humans to store concepts via a shorthand that can be stored, processed and searched by computers: so, humans IN and computers OUT.

The shorthand is classification: metadata in a structure that can define the content of a disclosure. The classification is designed to be powerful in terms of defining and

ISSN 2029-7564 (online)
https://www.mruni.eu/lt/mokslo_darbai/st/apie_leidini/
https://www.mruni.eu/en/mokslo_darbai/st/apie_leidini/index.php

searching concepts, whilst suited to a crowdsourcing effort. It is simple and intuitive to use. Most importantly, it is adapted to restrict ambiguity, which is the poison of classification, without imposing a restrictive centralised management.

In the classification scheme, each entity is shown together in a graphical representation with related entities. The entities are arranged on a sliding scale of similarity. This sliding scale is effectively fuzzy classification.

***Findings** – The authors of the paper have been developing a first classification scheme for the technology of traffic cones, this in preparation for a trial of a working system. The process has enabled the authors to further explore the practicalities of concept classification. The CmapTools knowledge modelling kit to develop the graphical representations has been used.*

***Practical implications** – Concept searching is seen as having two categories: prior art searching, which is searching for what already exists, and solution searching: a search for a novel solution to an existing problem.*

Prior art searching is not as efficient a process, as all encompassing in scope, or as accurate in result, as it could and probably should be. The prior art includes library collections, journals, conference proceedings and everything else that has been written, drawn, spoken or made public in any way. Much technical information is only published in patents. There is a good reason to improve prior art searching: research, industry, and indeed humanity faces the spectre of patent thickets: an impenetrable legal space that effectively hinders innovation rather than promotes it. Improved prior-art searching would help with the gardening and result in fewer and higher-quality patents. Poor-quality patents can reward patenting activity per se, which is not what the system was designed for. Improved prior-art searching could also result in less duplication in research, and/or lead to improved collaboration.

As regards solution search, the authors of the paper believe that much better use could be made of the existing literature to find solutions from non-obvious areas of science and technology. The so-called cross industry innovation could be joined by biomimetics, the inspiration of solutions from nature.

Crowdsourcing the concept shorthand could produce a system 'by the people, for the people', to quote Abraham Lincoln out of context. A Citizen Science and Technology initiative that developed a working search engine could generate revenue for academia. Any monies accruing could be invested in research for the common good, such as the development of climate change mitigation technologies, or the discovery of new antibiotics.

***Originality** – The authors know of no similar systems in development.*

***Keywords** – concept search, classification, collective intelligence.*

***Research type** – conceptual paper.*

Introduction

Concept searching is rendered complex by the fuzzy nature of a concept, the many possible implementations of the same concept, and the many more ways that the many implementations can be expressed in natural language. This is further compounded by concept search across different languages.

Computers are brilliant at storing and processing data. Computers can beat humans at chess, however, the game of chess is effectively a restricted set of patterns and rules that can be translated into data that a computer can then process.

Humans are comparatively brilliant at understanding and reasoning in more diverse and more abstract situations, and no more so than where the subject matter is expressed in natural language.

The symbiosis of computers and humans in support of 'solution by analogy' would be a powerful expression of collective intelligence. Durham Zoo (hereinafter – DZ) seeks to bridge the gap with the classification of concepts. The intelligence of humans is used to store concepts in classification shorthand that can then be processed as data by computers to retrieve the same or similar concept. No intelligence is required on the part of the computers.

Classification of the same classes can proceed on disclosures written in different languages, facilitating cross-language search. Classification stored in a central index can be independent of media type. Such a central index would not need contain the disclosures themselves, but simply contain their address. Subject to permission, this could support the search of copyright protected literature without the prior requirement of it being scanned and indexed by a search engine.

The major part of this paper presents the progress of a Proof of Concept of DZ. Although every effort has been made to reduce any necessary prior knowledge of the project to a minimum, a more complete understanding would benefit from a review of the existing literature.

Durham Zoo began as a project to build a concept search engine for science and technology. However, the Discussion and Suggestions sections consider a way in which Durham Zoo could contribute to the broader democratization of search, and maybe even democracy itself.

Durham Zoo Proof of Concept

Theoretical background

As a gross approximation, classification is the world of information studies, the humanities and librarianship, whereas information retrieval and data science is the world of the mathematician and the computer scientist. There would appear to be relatively little that sits in the middle. In terms of logic, library classification is based on the phrases 'is a member of the set of' and 'is a sub-class of'. Well suited to libraries, the structure is simple to understand, although the information retrieval power remains

limited. Computer science can use computer power to store and process information on a different scale altogether, a scale and complexity that would be too much for the human brain. The logic of the Resource Description Language (RDF) used in the Semantic Web includes a 'triple' that defines the relationship between two different entities, where the relationship may be anything. Very complex ontologies involving many entities and many relationships can be built using such RDF triples.

DZ seeks to operate somewhere in the middle.

Much classification is root structured, a good example being the taxonomy of living organisms, where classes are subdivided into increasingly specific subclasses. Faceted classification uses a different approach. 'The need for a faceted classification as the basis of all methods of information retrieval' was presented by Broughton (2012). Typically, one facet from each of multiple facet families is combined in a manner akin to a coordinate system. In DZ, the facets, originally conceived to define a concept in science and technology, were chosen to answer the simple questions: 'what' is the science or technology, 'where' is it applied, and 'when' in the lifecycle are we talking about. Together, these define a context in which the 'why' and 'how' relate to a problem and its solution. The what, where, when, why and how are more easily remembered as: Technology, Application, Operation, Problem and Solution, or simply TAOPS.

Traditional faceted classification is Boolean, and it relies on the mutual exclusivity of the different classes with a facet. More classes that are created in a facet increase the likelihood of the classes overlapping. An overlap introduces ambiguity, which results in the classification becoming 'noisy', analogous to the background noise in shortwave radio reception. Lofti A. Zadeh, the founder of fuzzy logic, described it elegantly: "as complexity rises, precise statements lose meaning and meaningful statements lose precision".

And so, DZ implements fuzzy classification where an entity can be something to a certain degree. This is more powerful than the 'all or nothing' Boolean simplification (Baruchelli and Succi, 2013). Unlike Boolean search, that returns a set of equally pertinent documents to a search query, a fuzzy search returns a more useful ranked list.

In DZ, each entity, called a Zootag, is shown together with similar Zootags on a graphical representation called Zootag Steering Diagrams (hereinafter – ZSDs). The hyperlinked web of ZSDs can be surfed. When a Zootag representing an entity to be search is selected, all the entities on the ZSD are selected, each weighted with their respective degree of similarity.

A search query containing a single Zootag from each of the Technology, Application, Operation, Problem and Solution facets takes all of the Zootags on the 5 corresponding ZSDs for search. The search results are ranked in order of their holistic similarity by summing the weighted 'hits' in the different facets. Computers can process the 20, 50, 100 or 1000 entities of the 5 ZSDs with ease. Search is more efficient, more powerful and less susceptible to ambiguity than traditional Boolean faceted classification.

Such a system only includes 'is similar to' to the relationships 'is a member of the set of' and 'is a sub-class of'. It is, thus, simple for human classifiers to understand.

Boolean or Fuzzy, there is not known any existing classification scheme that is suitable for use and development by the crowd. Such a system needs to be simple and intuitive enough for an occasional classifier to use. The development of the classification

scheme also needs to be suited to a distributed development. Traditionally, the 'what is what' and 'what goes where' is managed centrally in order to restrict the poison that is ambiguity.

Galaxy Zoo (galaxyzoo.org), the citizen science project to classify the galaxies in the universe, and to which Durham Zoo pays homage with its name, is coordinated centrally. The classification scheme relates to the shape and form or galaxies. Volunteer classifiers are restricted to classification proper rather than the development of a classification scheme.

Also managed centrally, the International Patent Classification (wipo.int/classifications/ipc/en/) covers all of technology and is correspondingly complex.

In DZ, the 'what is what' is managed centrally. However, the 'what goes where' is distributed across the crowd. Classifiers with expertise in different domains are able to develop to ZSDs independently: 'picking and placing' Zootags as they see fit. For although there may be common agreement across different technical fields as to 'what is what', the significance of the same thing in different situations can be very different. A 'fractal' representation of the ZSDs, together with an Occam's razor inspired algorithm is the key to making sense of the appreciation of different experts and their independently generated information. The reader is referred to the White Paper on the authors' website for more information.

Research methodology

The motivation for a Proof of Concept (hereinafter – PoC) was the belief that feedback in the early stages of any design is a good thing, and, in particular, where the efforts of volunteers are to be solicited.

The PoC would be positive if, after a crowd had classified a corpus of literature, searchers were able to retrieve the same concepts with sufficient ease, accuracy and completeness.

The 'sufficient' would initially need be no more than 'encouraging': this given that lessons would be learned on the way, and that the system would be expected to improve as it was developed, and more information was fed into it.

Additionally, any cross industry innovation would only be possible when the DZ database had information in it from different areas of science and technology. But then, maybe Wikipedia was not built in a day.

Three modules in the PoC were identified: knowledge capture, crowd processing and the search engine.

1. *Knowledge Capture*

Knowledge capture requires the development of a knowledge representation to be able to define a concept, the development of a platform to run a trial, the selection of a subject area to classify, and the classification itself.

1.1 *Knowledge representation: defining the numbers and nature of the facets*

The choice of the 5 TAOPS facets was the best guess from anecdotal evidence. Previously, it has been used to describe the workings of DZ in the context of the

occlusion of a stent, also called stent thrombosis. The example demonstrated the potential innovative power of DZ in suggesting a counter-intuitive solution, that of a non-smooth surface. The solution to the problem of stent occlusion is inspired by the solution to solving the problem of the build-up of bacteria on a catheter: a similar problem in a similar application.

The first reflection in the Proof of Concept was how the same facets would support search for the prior art, and this, in particular, with regard to patent-related matters. This could be a potential inventor or patent examiner looking to see if a concept already exists. Alternatively, a manufacturer may need to check if a proposed design infringes existing patents, or need to challenge the validity of a claim for having infringed a third-party's patent.

A patent application contains a description and claims. The claims define the legal protection and are drafted in 'legalese', a terminology designed to prevent an undue restriction in the scope of protection to a particular implementation of the concept. A concept relating essentially to a disk drive may see the device described as a 'storage device with relative movement between the recording surface and the transducer'. Such a wording would include the application of the concept to a tape drive. The abstraction of a concept in patents makes it such a good area to be explored in DZ.

For a patent application to be granted, the claims must define the subject matter that is both novel and inventive. Novel means that the same idea has never been seen before. Something that passes the novelty test must then pass the inventiveness test. A patent examiner has to decide whether what is novel is also inventive or 'non-obvious'. Is what is new a routine adaptation of what is already known, or is it something 'clever/alternative'? Non-obviousness is subjective and includes a comparison with what is already known.

The claims below relate to a fictitious patent application that relates to a painkiller and, more exactly, to an analgesic cocktail that produces a synergistic effect. The claims are more typical of a patent application than a granted patent, this given that matters of scope, clarity and legal compliance need to be addressed before a patent can be granted.

Claim 1: A method for administering a composition comprising a combination of analgesics.

Claim 2: A method according to claim 1 where the combination has a synergistic effect.

(In fact, Claim 2 would not be allowed as it says what it wants to achieve but not how it achieves it, sometimes called 'desiderata' or even 'free beer')

Claim 3: A method according to claim 1 where the composition comprises paracetamol and aspirin.

Claim 4: A method according to claim 1 where the composition is for treating migraine.

Claim 5: An analgesic comprising two active components.

Claim 6: An analgesic according to claim 5 for oral administration.

Claim 7: An analgesic according to claim 5 where the active components are paracetamol and aspirin.

Of note is that the claims relate to both a 'method' and a 'product'.

The description of a potential invention also discloses the benefits of a stimulant, such as caffeine to speed up the pain relief, but this does not appear in the original set of claims. It may be included at a later stage. There are, thus, two concepts in the application.

Patent examination may be expedient if the patent examiner finds a novelty document covering the complete subject matter of the application. If the inventor has published the idea before submitting the patent application, they may have 'shot themselves in the foot'. A search on the Internet for the inventor's name together with some of the key terms used in the application could retrieve such a document, including the caffeine stimulant.

If not the case, the search for novelty-destroying documents would be extended to include those disclosing the same thing, but described in alternative terminology: so, perhaps, combining N-acetyl-p-aminophenol, salicylic acid, and 3,7-Dihydro-1,3,7-trimethyl-1*H*-purine-2,6-dione.

Such a novelty search is effectively a search for the whole application, given that it would target all the claims in one go. Searching novelty-destroying documents is, thus, supported by the use of synonyms of the different components. DZ implements such a mapping of all synonyms to a single Zootag.

Searches using Zootags from the Technology, Application, Operation, Problem and Solution facets for the concept of the synergistic effect could include:

Technology: Analgesics
Application: Orally dispensed medicine
Operation: Drug delivery
Problem: High dosage requirement
Solution: Synergistic effect of a combination of active drugs.

For the acceleration of the relief, they could include:

Technology: Analgesics
Application: Orally dispensed medicine
Operation: Drug delivery (not to be confused with the administering of drugs)
Problem: Delay in effect
Solution: Stimulant additive.

An intelligent combination of the results from both searches should lead to the retrieval of a classified novelty-destroying document.

If the combination of paracetamol + aspirin + caffeine combination were not found, may the search for a combination of paracetamol + aspirin be the next best bet?

And then, if paracetamol + aspirin were not found, then the novelty of the much broader Claim 1 would need to be targeted. This would involve the abstraction of the subject matter in the description to include all possible combinations of analgesics, whether exhibiting a synergistic effect or not.

Maybe it is unlikely in this hypothetical case, but an existing combination of ibuprofen + morphine would be a novelty destroying for Claim 1. Thus, it would be useful to be able to search for all combinations of all analgesics with the one search statement.

To support the abstraction to all analgesics, it would be useful to be able to select all members of the set of analgesics. The structure of the ZSD facilitates this (using the Subject and Example structure explained in the White Paper on the authors' website).

If the novelty destroying documents are not found, the examiner can select the closest prior art document and develop reasoning as to why the claim is not inventive. As stated earlier, this lack of non-obviousness is subjective and includes the comparison with what is already known.

To reduce the subjectivity of the examiner, inventiveness can be judged using a Problem Solution (hereinafter – PS) methodology. PS seeks to establish whether in the light of what is known, and in the context of the application, whether the solution to the problem solved by what is missing or different in the prior art, is obvious or not. The PS reasoning can combine the teachings of more than one document.

Although the PS provides a framework for reducing the subjectivity of the examiner departing from a particular piece of prior art, the examiner still has to select this closest prior art. The closest prior art is the one from which the examiner can make the most convincing argument against non-obviousness. There may well be more than one way to 'skin the cat'.

In the present example, if the combination of paracetamol + aspirin was found in a document D1, but the combination of paracetamol + aspirin + caffeine was not disclosed in D1, D1 could be taken as the closest prior art to attack the inventiveness of such a combination. The second document could be used in the reasoning to demonstrate that in the context of the claim, solving the problem of slow pain relief with a stimulant, such as caffeine, would be obvious to the skilled person.

Alternatively, if the use of the caffeine stimulant together with paracetamol was known from a document D2, D2 could also be chosen as the closest prior art. In this case, the difference would be the synergistic effect of combining paracetamol with aspirin.

The established practice of partygoers taking a morning-after double aspirin + caffeine energy drink could be another starting point.

So, it can be seen that the prior art used for an attack on inventiveness may be many and varied. Different closest prior art documents will likely result in the definition of different solutions to different problems using the PS methodology. The use of Problem and Solution facets is, thus, as well suited to searching patent applications as it is to searching novel solutions from non-obvious areas of technology[1].

However, such anecdotal evidence supporting TAOPS is not proof enough before enlisting the efforts of the crowd. A major mistake with the facets used to define a concept could require the unwinding and rewinding of work completed by volunteers: a prospect to be avoided at all cost.

---

1 The detailed Durham Zoo workings of the above example appear in annex to this paper. The reader is invited to consult our previous publications in order to understand it fully.

A working system must be intuitive and simple to use: if it is not, the crowd will not feed it. However, if the system is not seen to work, it will not be fed either.

The number and nature of the facets used by the volunteers may not be the whole story: there is also the potential for building complexity into the system that can be hidden from users, or maybe from occasional users. In the White Paper, the authors presented complexities, such as 'crossover' between the different facets. For example, DZ could 'know' that a specific type of computer memory has a particular problem. A search including the Problem could infer the computer memory as an Application, even if not explicitly entered by the user.

Another example is a cooling system that includes piping, where the piping could produce mechanical support. So, strengthening the piping could be the Solution to the Problem of flexibility in a cooling system Application. However, in the Application that is a mechanical structure, the piping would be the Solution to the Problem of the removal of heat. Such crossover between facets enables 'behind-the-scenes' linking of such classified literature.

The number and nature of the facets necessary to define a concept is, thus, dependent on the hidden complexity.

The independence of thought, so much a part of collective intelligence, was encouraged amongst the relative few involved in the reflection on facets. However, the independence of thought resulted in such divergence, with all manner of complexities and potential consequences for a complete overhaul of the design, it was decided to simply pick a subject and see how the TAOPS fared.

1.2 *The platform*

The establishment of the definitions of the Zootags, and then picking and placing the Zootags onto the different Zootag Steering Diagrams (ZSDs), requires a platform. To enable teams of experts to contribute to the design of each ZSD, the platform requires collaborative editing functionality.

The prospect of a long and complex software development encouraged to look for existing software that could be used, or be adapted. After having explored ontology development tools, the authors looked toward mind mapping software.

The authors of this paper are very grateful for the opportunity to use the CmapTools Knowledge Management Kit developed by the Institute for Human and Machine Cognition in Florida. Nothing short of a godsend, CmapTools has everything and more.

CmapTools is ideally suited to producing the ZSDs. The server version of CmapTools facilitates the collaborative development of the ZSDs. CmapTools has provision for including text and photos with the Zootags, and the all-important hyperlinks necessary to allow surfing the web of ZSDs. The ZSDs are exportable as XML files: so, the graphical representations are converted into a form ready to be number crunched by a search engine. The system is very well documented, very stable and has a very large user community. Importantly, it also provides the necessary forum for a community to engage in collective endeavours.

1.3 The choice of technology for a trial

Early candidates considered included the 'hot topic' of 3D printing and the very ambitious but fascinating 'programmable matter'. However, it was decided that traffic cones were the better choice. Traffic cone technology includes some very established technologies, but it is increasingly varied. There is at least one traffic cone patent that deals with the swarm intelligence of a group of motor-driven cones, programmable to move from one configuration to another. There are cones that create a downforce from the wind of a passing car in a manner analogous to the downforce created by racing cars. There are cones that self assemble on being thrown out of the back of a moving police car.

Traffic cones have a slightly wacky side: for example, there is the Traffic Cone Appreciation Society (www.trafficcone.com). In Glasgow repeated efforts to remove traffic cones from the head of a statue of Wellington has stimulated reflection as to whether the cone is not part of the now iconic statue (The Targe, 2013). Traffic cones are sometimes used in art installations (TTCO, 2014). And then, there is David Morgan, who has the world's largest collection of cones, and is perhaps the cone's ultimate ambassador. David Morgan, alias 'Coneman the Barbarian' or the 'Cone Ranger', goes off in search of rare cones. He described his Holy Grail as a five-sided cone from Manchester, saying it was 'like looking for Elvis'. The Youtube documentary featuring him[2] is very good. It is hoped that traffic cones will capture the imagination of the public. Once studied, traffic cone technology cannot be forgotten: traffic cones are everywhere and provide a constant reminder.

The last factor was the large collection of publicly available literature pertaining to cones. Unlike much academic literature, patent office collections have open access: the European Patent Office's (EPO) Espacenet system being a good example. Google Patents is another useful resource. Traffic cone patents are classified in the Cooperative Patent Classification (CPC) system operated by the EPO and the United States Patent and Trademark Office (USPTO). The hierarchical scheme defines increasingly specific classes. The best class for traffic cones is E01F9/0122, described as:

"arrangement of road signs or traffic signals [of the kind] *upright bodies, e.g. marker posts or bollards* [of the kind] *free-standing, e.g. traffic cones, foldable or inflatable devices* [of the kind] *in the form of a three-dimensional body, other than of the relatively thin panel type, or designed to assume such form, e.g. cones, devices expandable or erectable to form a geometric body; Inflatable structures."*

A search using the CPC class E01F9/0122 returns over 800 results in the Espacenet system.

2. Crowd Information Processing

The crowd will input information into DZ. Hopefully, much of the information will be of excellent quality, however, there will be the full spectre of good and less good classification, and maybe a plan for malicious will be needed. Developing classifier profiles

2 *King Cone*. 2009 [interactive]. [accessed on 2014-11-11]. <www.youtube.com/watch?v=tCXvUCoN-S7w>..

is a mean of promoting the good above the less good. Amazon has a ranking system for their most appreciated reviewers and Ebay provide information about the reliability of sellers. In DZ, a similar classifier profiling mechanism would provide an indicator of the accuracy and reliability of the classified information. This would be factored into the ranking of the search results.

Taking account of the hopefully many-and-independent assessments of the same document by different experts can also be used to improve the quality of the classification information. How to collate and corroborate the information was also foreseen in the crowd information processing.

3. *Search Engine*

The search engine has to match a query with the classified literature and rank the results. The simplest manner of producing the ranked list would be the simple addition of the similarities in the different facets used in the query. A 'vector space' model, using Pythagoras to more accurately calculate the overall similarity of the multiple facets of a multi-dimensional representation, would be more complex.

The 'fractal representation' is necessary to link the ZSDs. Crossover between facets, the crowd information processing, and other bells and whistles, such as non-linear fuzzy relationships, would follow at a later stage.

The fault tolerance, security and availability of the system would also require consideration. Whilst Fuzzy SQL together with a relational database appeared as possible, a Hadoop and MapReduce implementation appeared to be a better solution: the parallelism, fault tolerance and scalability inherent in such a design, together with the Open Source Apache environment, ticking all the big boxes.

Results and findings

The Crowd Information Processing and Search Engine modules were not started. However, progress was made with Knowledge Capture.

After an initial broad survey of the traffic cone literature, several sub-areas were identified that could be developed in more detail. For example, the authors of this paper found many documents relating to means of increasing the visibility of cones, the use of cones for supports, and means to keep cones upright. Around 80 patents were identified relating in some way to a 2-part construction. A 2-part construction is often used to combine an expensive but colourful plastic cone with a heavy and cheap rubberised base. However, a 2-part construction may serve an alternate purpose: for example, the separation of the two component parts of impact, or a pick-and-mix design modularity.

The 2-part construction included many relatively old publications. It was noted that when starting with the classification of a technical field, it was judicious to start with the older literature. Following the historical development helps in understanding 'why' things are done as they are at a later stage, in addition to 'how'.

Furthermore, the Problem with a design, and the proposal of its Solution, is done at a particular moment in time with only the technologies and materials available at that the

time. The Problem with a particular design may change as technology changes. Should time be the 6th facet? The 'Timewarp' functionality evoked in the White Paper to handle this was not foreseen when starting from scratch.

Perhaps an obvious observation but the asymmetry of related entities was confirmed. As an example, a mug with a handle is suited to hold both cool water and hot soup, whilst a glass tumbler is suited to cool water but not hot soup. Thus, the degree of similarity as seen from the perspective of the mug to the tumbler, and the tumbler to the mug, would likely be different. Can this be called Zootag relativity?

In the 2-part construction, it was sometimes difficult to distinguish the Application, but since all were traffic cones, this is perhaps to be expected. The same was true to a lesser degree with Operation: all traffic cones are designed and manufactured and picked up and put down. Expertise of the subject facilitates a better identification of the distinguishing information to be classified.

Often, a single particular Problem was not evoked. Rather, the design solved a multitude of separate Problems or a combination of related Problems. Once again, expertise is of benefit in seeing the wood through the trees, especially given that patent applications are often generous in the appreciation of the Problems they solve. However, a particular design may simplify manufacture and facilitate stacking. The creation of multiple TAOPS, which the authors have also called 'quintuples' or 'quints', is a simple solution to classifying the distinguishing information. Such quints can be combined in a search query, but more reflection is required about this.

Rather than proceeding, in the order TAOPS it was considered preferable to use SPOTA: beginning with the definition of the Solution and the Problem avoided the inclusion of elements of Operation, Technology or Application in their definition. This avoids tautology in the concept definition.

In 4 man weeks, the Problem and Solution facet ZSDs were designed once for the 2-part traffic cones. The CmapTools proved itself to be an ideal tool.

The development of the Zootags and ZSDs is ongoing. The initial pass needs to be completed and then repeated by people working independently. Only then can the results be shared, compared and collectively reviewed.

Conclusions: So far to go, so little time

Not having got further with the project is a testament to the need for more of a crowd. However, the authors of this paper are encouraged by recent progress and remain optimistic for the future.

Discussion: Collective or Collected Intelligence. Is society due a raw deal?

1. What are Crowdsourced, Collective and Collaborative Intelligence?

What constitutes 'collective intelligence' is open to interpretation. Galaxy Zoo (galaxyzoo.org), the project to crowdsource the classification of the galaxies in the universe, is maybe notable in its absence from the corresponding Wikipedia entry.

Galaxy Zoo appears under Citizen Science and may be better described by 'crowdsourced intelligence', given that volunteers' intelligence is focused on predefined 'microtasks', to use terminology from Amazon's Mechanical Turk (mturk.com).

But then, independence of thought is often considered as both a prerequisite and one of the strengths of Collective Intelligence. Galaxy Zoo leverages independence of thought in the many classifications of the same image that are used to establish a more accurate final result. This is analogous to Francois Galton's observation about the accuracy achieved by averaging the many independent guesses of the weight of an ox, the story evoked in the introduction of James Surowiecki's 'The Wisdom of Crowds' (Surowiecki, 2005).

Elsewhere, independence of thought may be extended to leverage the knowledge of individuals in the crowd. A good example is the crowd searching platforms developed to retrieve prior art to be used in either the grant process, or to invalidate a granted patent. These include Peer to Patent, Ask Patents, Article One Partners, Patexia and BluePatent.

Open Innovation leverages the creative energies of the crowd, and is capable of producing a radically novel solution to a problem that skilled practitioners would likely neither consider nor conceive.

Such creative energies may well include inspiration from another source. This is the case of 'cross industry innovation', where old solutions are addressed to new problems. Maybe such innovation includes an element of 'standing on the shoulders' of the original solution creators.

Maybe 'collaborative intelligence' better defines the joint reflection of contemporaries to a certain puzzle or problem. Such collaboration is commonplace, and perhaps increasingly necessary in many scientific endeavours. The discovery of the structure of DNA is one famous example, with input from both King's College London and Cambridge University.

2. *Motivation and reward*

Motivation depends on the individual and the context. Citizen Science relies on the efforts of volunteers. The same applies to the Ask Patents crowdsearching initiative for 'people interested in improving and participating in the patent system'. Contributors to Ask Patents can earn badges for being 'especially useful', and thus, enhance their profile.

Article One Partners operate a bounty hunter model for the first to retrieve prior art leading to a successful attack on a granted patent. Open Innovation company InnoCentive (innocentive.com) offer cash prizes for a good idea.

Collaborative efforts do not always lead to collective recognition or collective reward. Rosalind Franklin's contribution to discovering the structure of DNA is one example amongst many. Whilst the Rosalind Franklin entry on Wikipedia (Wikipedia, 2014a) states that her X-ray diffraction images were shown to Watson 'without her approval or knowledge', the Wikipedia page on Maurice Wilkins (Wikipedia, 2014b), her colleague at King's College, states that Wilkins 'having checked he was free to use the photograph to confirm his earlier results showed it to Watson'. Perceptions may differ, however, it is documented that Crick attested to Franklin's contribution as having been the key (Zeller, 2003), and a matter of fact that only Crick, Watson and Wilkins became Nobel Laureates.

3. Collective and Collected Knowledge: See Terms and Conditions

Facebook encourages the sharing of information in return for an awesome user experience. However, a secret social experiment recently undertaken by Facebook has been the most recent case of users' incredulity at 'their' data being used in ways they did not fully expect. In 2012, Facebook modified users' newsfeeds to see if a mood could spread like 'emotional contagion' (Sullivan, 2014). The experiment was conducted without informing the 700,000 users concerned. At a subsequent conference in India, the COO of Facebook, Sheryl Sandberg, apologized for no more than 'poor communication'.

The *'quid pro quo'* of a 'free' user service in return for information that can then be used to 'monetize' a business has become common.

Google's Gmail links with other parts of the Google ecosystem. Eric Schmidt, presently the Executive Chairman of Google, stated: "With your permission you give us more information about you, about your friends, and we can improve the quality of our searches. We don't need you to type at all. We know where you are. We know where you've been. We can more or less know what you're thinking about." He also said: "I actually think most people don't want Google to answer their questions. They want Google to tell them what they should be doing next".

Unlike Facebook, Google's Gmail is not about sharing users' personal information with other users. Whether users' information is best described as 'personal' or 'private' or Google's may solicit moral argument, however, in law it would require the study of the terms and conditions. Ownership of the information may be of less concern to consumers than the confidentiality of the information. Google has recently argued in a court hearing (Simpson, 2014) that Gmail users, or anyone that sends emails to Gmail users, should not expect privacy. Consumer Watchdog reflected that "Google is either lying to the court or to the public" because it has repeatedly said that it respects users' privacy. Google has also shared users' information with the US government via the Prism program (Greenwald and MacAskill, 2013).

So much for user information, there are also platforms that encourage Internauts to share their knowledge. The stated mission of Quora, a question and answer website, is to 'share and grow the world's knowledge'. Researchgate, a website for scientists and researchers, and described as 'a mashup of Facebook, Twitter and LinkedIn' by the New York Times, is another example (Lin, 2012).

Sharing knowledge is often linked to the democratization of knowledge. Open Knowledge (okfn.org) describes itself as 'a worldwide non-profit network of people passionate about openness, using advocacy, technology and training to unlock information and enable people to work with it to create and share knowledge'. In relation to research, science and culture, Open Knowledge believes that 'Free access to the sum of human knowledge enables everyone to fully understand their lives and our world, to make informed choices, and to build a better future together'.

Freebase, created in 2007, encouraged users to feed a database of structured knowledge. The Wikipedia entry for Freebase (Wikipedia, 2014c) states that 'Freebase aims to create a global resource which allows people (and machines) to access common information more effectively'. Freebase was developed by Metaweb: a company bought

by Google in 2010. Google have used Freebase to feed their Knowledge Graph, a system that underpins their search engine.

Google's mission statement is 'to organize the world's information and make it universally accessible'. Whilst a demonstration of ambition when Google was founded in 1998, Google's present position, where it is much closer to becoming *the* world library, raises the question as to what degree Google is offering humanity a great service or offering themselves a great commercial opportunity.

The Google Books project began in 2004. Google announced its plans to copy millions of books from multiple libraries. By 2010, they had copied and digitized 15 million books. Google argued that where permission to preview the book on-line had been refused, no breach of copyright had taken place if users were only able to see snippets of information. Whether a snippet, such as a dictionary definition, should always be freely available is questionable.

In any case, in order to provide the snippets, the books require scanning and indexing in their entirety. Does this centralisation of knowledge pose a problem for society?

Siva Vaidhyanathan, the author of 'The Googlization of Everything (and Why We Should Worry)' (2011), is concerned enough about Google's increasing hegemony to have proposed a Global Knowledge Project to counter it. The 'must read' book evokes many things but was written by someone clearly passionate about knowledge, culture and librarianship.

However, is Google's end goal about Collective Knowledge or Collecting Intelligence?

4. *Artificial Intelligence and Collected Intelligence*

IBM's Watson is famous for having won a television quiz show against human participants. 'Smart Machines: IBM's Watson and the Era of Cognitive Computing' is both a fascinating and thought-provoking read (Kelly III and Hamm, 2013). In it, today's computers are described as 'brilliant idiots' (Kelly III and Hamm, 2013, p. 4), however, the development of artificial intelligence will facilitate the development of computers that can 'sense, learn, reason and interact naturally with people in powerful new ways' (Kelly III and Hamm, 2013, p. 3-4).

How will this affect society? Apparently, the goal is not about replacing human thinking, but 'Rather, in the era of cognitive systems, humans and machines will collaborate to produce better results'. However, one can assume that with time the balance of the contribution will shift toward the computer. The book outlines the different stages to developing artificial intelligence (hereinafter – AI). First, a machine has a 'static learning phase', where human experts feed a learning system and train it to use the information. This is followed by a 'dynamic learning phase', where the machine effectively feeds itself. The end goal is 'inductive reasoning', where computers 'don't just do what they are told; they figure out what to do on their own' (Kelly III and Hamm, 2013, p. 74). Eric Brown, who heads IBM Research's Watson team, is on record as saying 'We created an architecture of discovery' (Kelly III and Hamm, 2013, p. 32). Humans have had the monopoly on this until now.

Is this only for the far future? A demonstration of the present is IBM's contribution to the DOME project, a project to build a radio telescope. This is perhaps the bleeding

edge of 'Big Data', research given the huge amount of data that needs to be processed. Rather than assemble 'upward of 250 experts in a wide array of computer science domains in order to design the architecture of the DOME system', it is an 'Algorithms & Machines' team who are 'gathering all of the pertinent knowledge in a repository, setting the parameters for the entire computing system, and creating optimization algorithms' (Kelly III and Hamm, 2013, p. 66). IBM employs brilliant people and is investing heavily in the technology. They are advancing both technology and knowledge. Were the Luddites not proven wrong? Is AI just another 'disruptive technology' along with the likes of 3D printing?

Perhaps it would be folly, maybe even irresponsible, for humanity to ignore the potential of AI. For certain, it would be against human nature. And if, as suggested in the book, the hiring of 250 experts and bringing them together 'isn't practical' (Kelly III and Hamm, 2013, p 66), then, in fact, IBM have no option. However, as is often the case, the development of technology is leading the broader debate about the consequences of the technology.

Google is also in the business of AI: this year they purchased DeepMind Technologies, an AI company from London. This year also saw the presentation of a paper by Google at the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining: 'Data Science for Social Good'. The paper introduced Google's Knowledge Vault (Dong, *et al.*, 2014). Google has apparently gone far enough with 'static learning', in part thanks to the 1.6 million facts input via Freebase. Google's Knowledge Graph is to be superseded by dynamic learning into their new Knowledge Vault. Computers will, thus, feed the Knowledge Vault with human-generated information. That Google have chosen to call their new system Knowledge Vault is perhaps indicative of a change in policy. A vault is a place where something precious is kept safe from others. Is it an implicit declaration of the ownership of the knowledge?

And whilst relative 'snippets' of the knowledge in the Knowledge Vault may be shared to power a search engine, the ability to process the information as a whole will not be.

More than knowledge *per se*, is the Knowledge Vault not more about the development of cognitive computing and AI, which in turn is about science and technology, problem solving, innovation, engineering, making drones and driverless cars and everything else that Google wants to design and build and sell?

Has there been a shift from the democratization of knowledge to the leverage of collected intelligence for commercial advantage in design and manufacture? And is the Knowledge Vault really 'Data Science for Social Good'?

The collection of 'the world's knowledge' into a machine big and smart enough to process it would represent a massive advantage when designing and building consumer goods. Even if knowledge is 'open' and democratized, few are in possession of either a server farm or the AI to process the information. And if Google is alone in having the world library, then Google will be alone in being able to process the information in its entirety.

Google has 'convinced hundreds of suppliers of electronic scholarly resources to open their indexes up to Google's spiders so that articles could be scanned, copied,

and included in Google's index' (Vaidhyanathan, 2011, p. 192). This knowledge is the result of human intelligence and human endeavour. Are the authors maybe unwittingly feeding Google's AI?

If scientists publish the many parts of an almost complete puzzle relating to a new antibiotic and Google puts in the final piece, should and would the scientists who contributed benefit in any way other than as simple consumers? Will academia more often find itself the Rosalind Franklin than the Crick or Watson? With AI and the world's knowledge at their disposal cannot Google simultaneously stand on the shoulders of all the giants and all the not so giant that ever lived?

Is the society being duped? Or has society been duped already?

Finally, will the patent system become defunct? An invention requires the assessment of non-obviousness. This is presently done by the fictitious 'skilled person', but what about the 'skilled computer'?

5. *Big is beautiful?*

In terms of the law, Eric Schmidt has said Google's policy was to "to get right up to the creepy line and not cross it." As demonstrated by Google Books, there is an argument as to where the line actually is. Google's lobbying power gives it influence as to where the line gets drawn. Google is lobbying the same government that acquires data through Google for the Prism surveillance program.

Many companies and organisations lobby. Google also has a big PR machine. Many articles in the press focus on the many good things that Google is doing. A recent article on the BBC website (Stewart, 2014) has presented Project Wing, a project 'being developed at Google X, the company's clandestine tech research arm'. In it, Google are cited as saying that its 'long-term goal was to develop drones that could be used for disaster relief by delivering aid to isolated areas'. The article evokes at relative length the delivery of medicines, batteries, humanitarian aid and even defibrillators to heart attack suspects. As a useful spin-off, however, 'Eventually Google said it could use unmanned flying vehicles to deliver shopping items to consumers at home'. Whether an element of naïve reporting or Google advertorial, it is regrettable that no mention is made of the defibrillator delivering Defikopter (Nateog, 2013) or the Ambulance Drone from Urban Aeronautics (Russell, 2014). Details of the Ambulance Drone developed by TU Delft in the Netherlands appeared after the BBC article was published (Webredactie Communication, 2014).

Suggestions: Democratising search?

1. *Copyright*

In 1620, the English philosopher Francis Bacon wrote that the printing press had 'changed the whole face and state of the world' (Wikipedia, 2014d). Legislation to prevent these new machines from copying potentially treasonable texts was introduced under Charles II in 1662 (Wikipedia, 2014e). It was only in 1710 that the Statute of Anne passed more socially minded legislation that put power in the hands of the

authors (Wikipedia, 2014f). In many languages, copyright has gladly been translated as an 'author's right'.

That an unauthorized public rendition of 'Happy Birthday to You' may be illegal is perhaps not widely known. It is also perhaps evident that copyright law has not functioned as intended by the creators of the legislation. However, a quirk of the law falls into insignificance in the face of copyright abuse so widespread that many creators of content are resigned to find alternate commercial models to pay the rent.

There would appear two issues: the control of the distribution of content, and the generation of associated revenue; and the two are linked. It may be that a creator of information desires unrestricted and free distribution of his work. It may be that a creator desires tighter control of the distribution in order to be able to secure revenue.

Viacom sued Youtube for copyright infringement. Considerations by the judge who ruled in favour of Youtube included whether 'Youtube had knowledge or awareness of any specific infringements' and whether YouTube had the 'right and ability to control' infringing activity.

Again, there appear two issues: the technology and the law, and again they are linked. If Shazam, the mobile phone app, can recognize a song in a few seconds, then a technological solution could presumably be developed to decide whether information is copyright protected or not. Smart computing could be used for checking copyright, and ideally before a copyright protected disclosure is copied or crawled. So much for technology: what of the law? As regards the 'right (and ability) to control', should not the question be whether there is the responsibility to control?

Should not the creators of information have a greater say than the distributors of information? The liability of distributors was restricted with the Digital Millenium Copyright Act. Do the present failings in the law+technology not result in much copyright protected information being shared without benefit for the creators?

With AI, the creative content of a research paper can be captured and copied across multiple artificially intelligent machines. Ridiculous as it may seem, but just as consumers pay software licenses according to the number of copies they use, could not a researcher claim a license for each AI machine that uses his work?

It is true that a researcher could protect and profit from an *inventive* concept with a patent, but Rosalind Franklin's landmark photo 51 that was the key to discovering the dual helix would not fall under that category.

Will not AI change the 'whole face and state of the world'?

2. *Mundanuem 2.0*

Maybe the 'trending' commercial ethos of 'do it first, PR apology later'; 'do it first, terms and conditions'; 'do it first, we can go through the courts and in any case the potential sanctions are no deterrent'; together with the exploitation of 'perfectly legal tax avoidance' cannot be in society's interest. And yes, maybe it is capitalism, but is it the capitalism that society wants?

Maybe academia could develop a search engine for the benefit of society rather than shareholder profit for a relative few. Academia has the skills and expertise.

Perhaps there is a need to create a contemporary version of the Mundaneum, often referred to as the 'Paper Internet'.

In 1895, two Belgians, Paul Otlet and Henri La Fontaine, began a system to classify the world's knowledge. Over 12 million index cards were created that could be used to retrieve information from a wide range of publications. Queries could be sent by telegraph or by post. Truly revolutionary, their system even included a precursor to hypertext (Rayward, 1994).

The project led to the Universal Decimal Classification, still in use in libraries today. However, the Mundaneum was re-housed and the initiative lost momentum during the 2nd World War. What is left of the Mundaneum is presently housed in Mons in Belgium, and since 2012 it has been receiving support from Google.

Perhaps DZ could help renew the efforts of Otlet and La Fontaine? The ZSDs used to classify the literature would be necessarily public: both classifiers and searchers need to be able to surf them. However, classification information attributed to the literature could be kept in a 'People's Vault' at the discretion of the individual classifier.

Whilst the information that feeds the search engine would be hidden, the search engine could be made publicly available. Such a system would require measures to prevent the People's Vault from being reverse-engineered.

If enough people were minded to make a contribution, and the search engine became sufficiently powerful, it could generate revenue via sponsored ads. Any monies earned could help pay for the system, be returned to the creators of information and/or the classifiers, or fund research into societal needs, such as climate change mitigation technologies.

For sure, more action on population control and climate change is needed. This year's 'must have' is a planet that sustains life, a planet with the atmosphere we can breathe, and a temperature we can tolerate. Our neighbour Venus is over 400 degrees centigrade, thus, uncomfortably warm, whilst Uranus, Neptune and Pluto are below minus 200 degrees centigrade, which is more than a tad chilly. Contrary to copyright law, the laws of physics are universally obeyed and we ignore them at our peril.

3. *What about social engagement?*

The knowledge representation underpinning the search-and-innovation engine is believed to be both simple and intuitive, whilst potentially very powerful. So whilst DZ was designed with science and technology in mind, the same platform could be used to solicit the input of knowledge and/or preference regarding a societal issue that can be modelled according to multiple criteria.

In the search-and-innovation engine, DZ collates and corroborates different classifications of the same document to improve the quality of the information.

The authors of this paper have retrieved what is effectively fuzzy and faceted classification for decision-making using 'ABC analysis', a scheme originally created for inventory control. It has been adapted to decision making in the design of a network of dismantling centres for end-of-life vehicles (Tadić, *et al.*, 2012). Whilst this may deal with matters of fact, could the collation and corroboration of preferences be used to render a decision more democratic?

In DZ, the collective decision on proposed policymaking could be extended to the suggestion and development of social policy if editing rights were granted to the

ZSDs. The authors of this paper believe that the simple and intuitive nature of inputting information into DZ, together with the facility with which the criteria can then be processed, is worthy of investigation.

As yet, there is little literature relating to a DZ-like system that crowdsources opinion and/or preference. However, there is help at hand. The Internet, Politics, and Policy (ipp.oii.ox.ac.uk) academic conference at the Oxford Internet Institute entitled 'Crowdsourcing for Politics and Policy' looks to be a good starting point.

Whilst the consensus with regard to many societal concerns may be difficult to achieve, a suitably adapted DZ could help the democratic process. It could even help redraw the creepy line.

Acknowledgement

The authors of this paper are grateful to Arthur Absalom and Arthaud Adriaenssens for their input on the facet design and classification of the cones, Alberto J. Cañas and IHMC for CmapTools, Mantas Bileišis for suggesting the policy development application, and Aelita Skaržauskienė, Monika Mačiulienė, Rūta Tamošiūnaitė and Marius Kalinauskas for their much appreciated and continued support.

References

Baruchelli, B., and Succi, G. *A Fuzzy Approach to Faceted Classification and Retrieval of Reusable Software Components*. DISA, Università di Trento, Via Inama 5, I-38100 Trento, Italia, 2013.

Broughton, V. The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval. *Aslib Proceedings*. 2006, 58(1/2): 49-72.

Dong, X. L., Murphy, K., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., and Zhang, W. Knowledge Vault: A Web-scale Approach to Probabilistic Knowledge Fusion. 2014.

Grenwald, G., and MacAskill, E. NSA Prism Program Taps in to User Data of Apple, Google and Others. *The Guardian*. 7 June 2013 [interactive]. [accessed on 2014-11-11].<http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>.

Kelly III, J. E., and Hamm, S. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. Columbia University Press, 2013, p. 160.

Lin, Th. Cracking Open the Scientific Process. *The New York Times*. 2012. [accessed on 2012-08-09].

Nateog. *Health from Above: A Drone to Deliver Defibrillators to Heart Attack Victims*. at The Verge on August 24, 2013 [interactive]. [accessed on 2014-11-14]. <www.theverge.com/2013/8/24/4654514/definetz-height-tech-defibrillator-carrying-drone-in-germany>.

Rayward,W. B. Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext. *Journal of the American Society for Information Science*. 1994, 45 [interactive]. [accessed on 2014-11-14]. <people.lis.illinois.edu/~wrayward/otlet/xanadu.htm>.

Russell, K. An Israeli Company Is Testing Flying Ambulance Drones. *Business Insider*. January 30, 2014 [interactive]. [accessed on 2014-11-14]. <www.businessinsider.com/israel-testing-flying-ambulance-drones-2014-1#ixzz3J9KN455K>.

Simpson, J. M. Google Tells Court You Cannot Expect Privacy When Sending Messages to Gmail -- People Who Care About Privacy Should Not Use Service, Consumer Watchdog Says. *Consumer Watchdog*. December 8, 2013 [interactive]. [accessed on 2014-11-14]. <http://www.consumerwatchdog.org/newsrelease/google-tells-court-you-cannot-expect-privacy-when-sending-messages-gmail-people-who-care>.

Stewart, J. Google Tests Drone Deliveries in Project Wing Trials. *BBC World Service Radio*. Los Angeles, 28 August 2014 [interactive]. [accessed on 2014-11-14]. <http://www.bbc.com/news/technology-28964260>.

Sullivan, G. Sheryl Sandberg not Sorry for Facebook Mood Manipulation Study. *The Washington Post*. 2014 [interactive]. [accessed on 2014-11-11]. <http://www.washingtonpost.com/news/morning-mix/wp/2014/07/03/sheryl-sandberg-not-sorry-for-facebook-mood-manipulation-study>.

Surowiecki, J. *The Wisdom of Crowds*. Random House LLC, 2005.

Tadić, D., Arsovski1, S., and Stefanović, M. *Designing of Network for the Dismantling Centers of the End-of-Life Vehicles by Fuzzy ABC Method*. 6th International Quality Conference, June 8th, 2012.Center for Quality, Faculty of Engineering, University of Kragujevac.

The Targe. *Glasgow Rallies to Save Wellington Cone*. Publication date 12 November 2013 [interactive]. [accessed on 2014-11-11]. <http://thetarge.co.uk/article/current-affairs/0137/glasgow-rallies-to-save-wellington-cone>.

TTCO. *The Traffic Cone Occasional: A Multimedia Group Exhibition*. 2014 [interactive]. [accessed on 2014-11-11]. <www.mobius.org/events/traffic-cone-occasional>.

Vaidhyanathan, S. *The Googlization of Everything (and Why We Should Worry)*. University of California Press, 2011, p. 280.

Webredactie Communication. *TU Delft's Ambulance Drone Drastically Increases Chances of Survival of Cardiac Arrest Patients*. 27 October 2014 [interactive]. [accessed on 2014-11-14]. <www.tudelft.nl/en/current/latest-news/article/detail/ambulance-drone-tu-delft-vergroot-overlevingskans-bij-hartstilstand-drastisch>.

Wikipedia. *Rosalind Franklin*. 2014a [interactive]. [accessed on 2014-11-11]. <en.wikipedia.org/wiki/Rosalind_Franklin>.

Wikipedia. *Maurice Wilkins*. 2014b [interactive]. [accessed on 2014-11-11]. <en.wikipedia.org/wiki/Maurice_Wilkins>.

Wikipedia. *Article on Freebase*. Article modification from 14 November 2014. 2014c [interactive]. [accessed on 2014-11-14]. <en.wikipedia.org/wiki/Freebase>.

Wikipedia. *Printing Press*. 2014d [interactive]. [accessed on 2014-11-14].<en.wikipedia.org/wiki/Printing press>.

Wikipedia. *Licensing of the Press Act 1662*. 2014e [interactive]. [accessed on 2014-11-14]. <en.wikipedia.org/wiki/Licensing_of_the_Press_Act_1662>.

Wikipedia. *Statute of Anne*. 2014f [interactive]. [accessed on 2014-11-14]. <en.wikipedia.org/wiki/Statute_of_Anne>.

Zeller, D. Crick's 31 December 1961 letter to Jacques Monod was discovered in the Archives of the Pasteur Institute by Doris Zeller, then reprinted in Nature 425: 15. 4 September 2003. Bibcode:2003 Natur.425...15Z. doi:10.1038/425015b.

Annex: Durham Zoo handling of the analgesic search.

The example is hypothetical and comes with apologies to the pharmacy profession.

Remember that the Zootags are held in a central repository, and picked and placed onto Zootag Steering Diagrams (ZSDs) by experts in the particular fields. The ZSDs contain the Subject, Examples that are members of the set of the Subject, and Inferrands that are things that are similar to the Subject to a greater or lesser degree.

Aspirin, Paracetamol, Ibuprofen, Caffeine, etc. would be Zootags in the controlled vocabulary with controlled meaning, as would Analgesic and Stimulant. The texts accompanying the Zootags would contain the synonyms of the Zootags, including trade names. This handles the complexity of synonyms and polysemes in the literature.

Each of the Zootags possesses their own ZSD. Each ZSD is required to have a minimum of one Inferrand. This Inferrand is necessary to link the ZSD with all of the other ZSDs via the fractal representation.

The Analgesic ZSD would have Aspirin, Paracetamol, Ibuprofen, Morphine and all other analgesics as Examples of the Analgesic Zootag. This is an effective way of logically OR'ing all the different types of analgesics.

The Analgesic ZSD would also include Inferrands, such as Cannabis, that have analgesic properties (if it were not considered by the experts as an Example), and Liquor that would be even higher up the ZSD.

The fuzzy reasoning is performed using the 5-way faceted classification.

In reality, there would likely be a Zootag and related ZSD for analgesic combinations, such as the use of multiple and different analgesics, to produce a synergistic effect. Depending on the definition of 'analgesic combinations', this could also include the combination of analgesics with atypical/adjuvant analgesics, and analgesics with stimulants as Examples or Inferrands.

Authors' application would hopefully be classified with the following, or very similar, Zootags:

For the concept of the synergistic effect:

Technology: Analgesics
Application: Orally dispensed medicine
Operation: Drug delivery
Problem: High dosage requirement
Solution: Synergistic effect of a combination of active drugs.

For the acceleration of the relief:

Technology: Analgesics
Application: Orally dispensed medicine
Operation: Drug delivery (not to be confused with the administering of drugs)
Problem: Delay in effect
Solution: Stimulant additive.

Combining the two TAOPS queries in combination would likely retrieve the 'total novelty' document, if such a document existed (note there appears to be nothing that links the synergy of the paracetamol + aspirin combination with the effect of the stimulant).

Focusing uniquely on the synergistic effect of a combination of active drugs, only the second query would be input.

It is true that the 'total novelty' document could be handled by faceted Boolean classification. However, the DZ framework would have made sense of the synonyms and polysemes, and getting everyone to speak the same language.

Searching with a Zootag also results in a search of all the Zootags on the particular ZSD, which is an efficient way of searching. This inclusion of similar entities in the different facets can pick up different but related prior art.

As an example, if a prior art document that was exactly the same as the application, but for the fact it was an anaesthetic, as opposed to an analgesic, may also be of interest?

Would the 'Boolean' searcher look in anaesthetics for prior art? Given that the anaesthetic Zootag would be on the analgesic ZSD as an Inferrand, such literature would be automatically taken into account. The ranking would be dependent on the classified literature, and it could be the closest prior art.

Non-orally dispensed medicines would also be included and searched automatically. So, a Brompton Cocktail, a liquid mixture of Morphine with Cocaine, may also rank highly.

The ranking would also depend on the existing Zootags. The fuzzy reasoning would also retrieve and rank a combination of analgesics with the combination of an analgesic with an atypical/adjuvant analgesic, the latter of which would likely contain a synergistic effect.

Importantly, DZ also provides a degree of fault tolerance in the classification process. A document that, in fact, relates to the same subject matter but that is described in slightly different terms may be classified differently. The combination of the facets into an overall ranking should see this variance 'come out in the wash'.

For example, the problems evoked in documents in the literature may not highlight the exact same problem. In this example, the problem of a high dosage of a particular drug could be expressed by the side effects linked to a large dose of a particular drug, or the limit imposed by the maximum dosage of a particular drug. The Zootags that represent these problems would be on the High Dosage Requirement ZSD and as such they would be included.

Problems can create problems. A second problem that is the unique consequence of the first problem may be considered to be an Example of the first.

The strength of DZ is the efficiency of the search process, and the accuracy and scope of the result when searching for both novelty-destroying documents as well as documents for use in anon-obviousness reasoning.