

The analysis and presentation of patents to support engineering design

This paper explores the role of patents in engineering design, and how the extraction and presentation of patent data could be improved for designers. We propose the use of crowdsourcing as a means to post tasks online for a crowd of people to participate and complete. The issues of assessment, searching, clustering and knowledge transfer are evaluated with respect to the literature. Opportunities for potential crowd intervention are then discussed, before the presentation of two initial studies. These related to the categorization and interpretation of patents respectively using an online platform. The initial results establish basic crowd capabilities in understanding patent text and interpreting patent drawings. This has shown that reasonable results can be achieved if tasks of appropriate duration and complexity are set, and if test questions are incorporated to ensure a basic level of understanding exists in the workers.

Overview

Although there are millions of online patent records instantly available, their volume and format of presentation combine to make interpretation of their contents and assessment a laborious process. Designers therefore need new tools to allow them to quickly and accurately understand the relevant patents in the context of a new design or innovation. We propose the utilisation of crowdsourcing to cut through the patent jungle and deliver a concise summary of the relevant Intellectual Property (IP) and its applications in an area of interest. Key components in crowdsourcing workflows are repetition (i.e. multiple, parallel tasks to generate sets of answers), peer review, iteration, and the linkage of payment to quality assessments. These characteristics can be used to locate relevant patent records, summarize their contents and collaboratively construct infographics that show the relative strength of clustering around topics. In other words, our research focuses on

the use of the crowd to provide the designer with two specific areas of functionality: patent usage assessment and patent landscape visualisation. Specifically, we aim to generate a new, visual form of patent map or gallery that incorporates measures of patent commercialization activity and technical metadata through the crowdsourcing approach that can be utilized by engineering designers during conceptualization and embodiment design. This paper reports on initial foundation steps vital to achieve this aim by developing a crowd capable of solving the upcoming high-reasoning patent analysis tasks.

Patents in engineering design

Despite the significant costs involved, patents have become a dominant facet of innovation with over 50 million being recorded in the European IPO database alone. International corporate strategies have been built around IP portfolios [1], and digital infrastructures have facilitated a significant industry in patent analytics and landscaping. The volume of patents filed has led to fresh questions on its efficacy with issues including patent “thicketing” to stifle competitors, trolling by companies who add nothing to our creative pool, protracted legal disputes between multi-national companies, and the awarding of patents to seemingly common sense designs cited as evidence of its problems. While new approaches such as open innovation, and changes to patent laws have been proposed, the patent system is a deeply entrenched and internationally recognized facet of business life.

For engineering designers, patents can support different modes of working moving through the product development process. Fig. 1 uses Pugh’s [2] Total Design model to illustrate the kinds of activity commonly undertaken by engineering designers and how these could be supported. While there are a number of tools such as landscaping and TRIZ, patents remain an under-utilized resource at a practical level for engineering design particularly through the conceptualisation and detailed design phases [3]. Researchers have employed a range of approaches to attempt to remedy this. Chan et al [4] show that providing designers with “far-field and less common” patents can have a positive effect on creative idea generation. Chang et al [5] have used keywords in conjunction with the established TRIZ inventive principles to provide “design-arounds” in the latter phases of conceptual design. And the preparation of patents has also been used as an educational approach to “directed conceptual design” [6]. To determine how patents can best be applied in these different design contexts, it is necessary to understand various approaches to patent analysis.

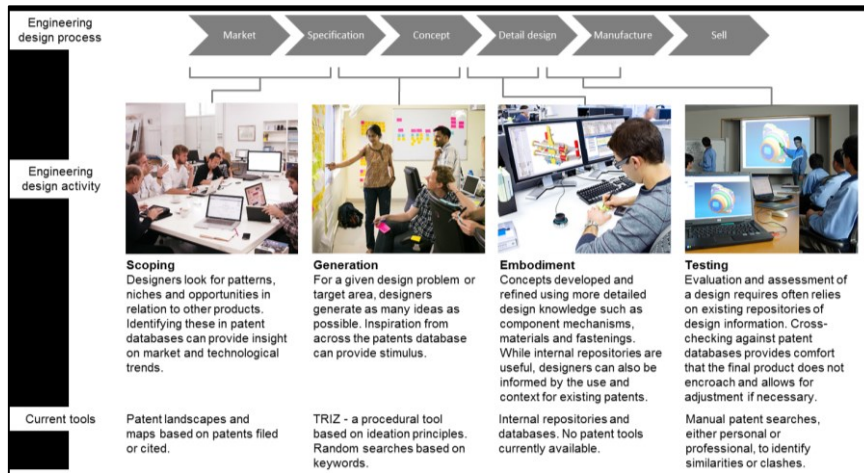


Fig. 1 Use of patents at different stages of the product development process

The challenge of analyzing patents

As an enormous publicly available source of data, patents have attracted a great deal of attention from researchers attempting to use them to identify patterns of innovation, technological trends, and creative thinking. While much academic research has been in the realm of economics, engineering design is well-placed to utilize patent information given the fact that many contain novel design solutions. Indeed, although there are specialist patents such as pharmaceuticals and plants that serve niche industries, the majority relate to novel technologies that can be applied in a range of settings. For example, it is impossible to patent an umbrella as a stand-alone concept, but novel materials, opening mechanisms, handle designs and so on can be patented. And while these may form part of an umbrella patent, they could equally be applied in other areas. The challenges in extracting and understanding the different kinds of creative thinking bound up in the formal patent document make it problematic for use in practical engineering design settings. We have therefore undertaken a review of the use of patents in engineering design, along with applications of the crowdsourcing approach.

Patent Assessment

Patent assessment can be grouped into two stages: pre- and post-grant. Before the patent is granted, novelty and usefulness are the two important parameters. Whereas after the patent is granted, patents are evaluated based on the possibility of commercialization.

Pre-grant

Many methods have been proposed in the engineering design literature to assess novelty [7-11], but only few for usefulness. For example, Sarkar and Chakrabarti [12] proposed a method using the SAPPhIRE constructs (Action, State change, Physical Phenomena, Physical effect, Organ, Parts and Inputs) for assessing the degree of novelty, and measured usefulness by multiplication of a set of parameters: level of importance, rate of popularity of usage, rate of use, frequency of usage, and duration of use, and combine them to get an overall score. However, this usefulness definition is largely varied with other definitions such as “functional and operable” [13], “manufacturable” [14], “achievable” [15], “practically useful” (US Court Proceedings), “specific, substantial (i.e. real world use) and credible (i.e. believable to a person of ordinary skill in the art)” [16], and “capable of industrial application” [17]. Among these various definitions, USPTO usefulness assessment by “specific, substantial and credible” and “capable of industrial application” [18] are accepted comprehensively. However, Nicol and Nielsen had argued that worthwhile or commercially practical is not a criterion to assess usefulness. In contrary, other view such as manufacturability scope including “economic value to the country” do exist [19]. It has been argued that a patent should include a complete specification fully describe the use of the invention and how it can be achieved [14]. However, whether this procedure is commonly applied in patent applications is questionable. Observing 36% of the European patents didn’t commercialize [20] represents that “novelty” is a primary factor assessed in-detail for giving a patent grant, whereas usefulness and obviousness are just assessed with an overall observation, reflecting The Nuffield Council observations that the guidelines to assess utility is too low leading to mention only ‘theoretically or speculative possible’ purpose [21]. However, taking steps to incorporate specific, substantial and credible utilities will increase time and cost for examination potentially delaying patent publication and which may eventually reduce patent applications [22]. Potentials of using the crowdsourcing approach for assessing both novelty and usefulness during the design process should be explored as a cost effective practices.

Post-grant

After a patent has been granted, accessing patent quality and its applications for converting patents into product development has huge potential industrial merits. However, there are no acceptable and common measures for this purpose. Trappey et al. [23] proposed a methodology to shorten the time required to determine and rank the quality of patents with respect to their

potential values in the IP rights marketplace. The methodology involves extracting relevant patent quality performance indicators, identifying the key impact factors using principal component analysis, and using a back-propagation neural network (BPN) trained model to predict the quality of patents and forecast the IP market potential. Although the paper claims that through historical patents that the proposed methodology produced 85% accuracy in automatic pre-evaluating patents which have commercialization value, the major issue is in identification of quality indicators. This paper used the following indicators: Application length (patent application and issue date), Number of international patent classification (IPC), Number of US patent classification (UPC), Forward citations, Foreign citations, Backward citations, Number of claims, Independent claims, Patent family (i.e., a set of patents in various countries taken to protect a single invention), Technology cycle time (i.e. the median age of the cited patents), Science linkage (i.e. the average number of references which are cited from scientific papers), and the length of specification.

Other indicators used to value patents are related to investment, maintenance, and litigation (e.g. patent trade and patent assignment). The following parameters are used to assess countries progress on technology development: Pending duration (the time duration of the ultimately successful patents that have been in the application grant process), Originality index (measures the extent to which the patent is based on broad technological roots, because the patent is more likely to synthesize knowledge across a wide variety of disciplines), and Technology dependence (measures the proportion of self-citations) [24]. Although these meta-data based indicators help to identify potential value of patents, the real indicators should emerge from semantic analysis of patent claims. This technical analysis of in-depth patent claims could be explored through the crowdsourcing approach, considering limitations of computer algorithms. Issues related to algorithmic-based text analysis are discussed in the next section.

There is only a single case study identified to illustrate challenges in commercializing innovative products. Roy [25] published case studies of Strida, Dyson, and Sanders UK companies in commercializing innovative products. The case studies discussed issues related to sales, risks, transferring licensing rights, and radicalness involved in invention, patenting and to commercialize innovative products. More empirical studies are required to understand how patents are applied successfully during new product development.

Patent Searches

The literature related to patent searches is predominately focused on computer-based retrieval techniques rather than understanding real needs of designers in the design process. This section reviews a range of empirical studies, and summarizes computer-based retrieval techniques applied to engineering design.

Empirical studies

Heisig et al. [26] reported the UK survey results of knowledge and information requirements of managers and engineers in design and service. The result observed that the “patent” category was noted only once within knowledge and information needs mentioned from 129 respondents. They also noted that only one other empirical study [27] reported about patent information needs. Patent infringement checks help to avoid reinventing the wheel, prevent costly litigation, and could potentially draw inspiration from existing designs. Koh [28] discussed challenges involved in answering when (before, during, and after the design) and how IP infringement checks should be conducted during the engineering design process. He reiterated that searching patents before the design stages is difficult with design problem information alone (i.e. does not really know what to look for), and search scope is much wider. Although search relevance could be increased for IP checks during the design stages, it can be costly to conduct frequent and thorough IP checks, especially involving a professional patent searcher due to decipher patent documents strategically written with intention to hide the scope of protection [29]. Koh highlighted that research is required to identify impact of patent infringement checks on design creativity, and to develop an affordable and effective means of retrieving relevant IP information from the sheer volume of documents during the engineering design process. In line with discussion mentioned for novelty, possibility of frequent and thorough IP checks should be explored through the low cost crowdsourcing approach.

Patent search techniques

Patent searches are useful for extracting useful information, and examining patents for possible infringements. Patent searches are classified into meta-data searchers (i.e. prolific inventors, cited patents, UPC classes etc.) and whole patent text searches. Keyword searches are commonly used to find relevant patent documents. However keyword searches are not adequate to find appropriate documents due to inaccurate usage of terminologies, synonyms, polysemy, pronouns, multiple attributions, varying detail levels of patent descriptions, and homographs present in patents.

For whole patent search, patent parsing is common. Structure (i.e. syntax) and dependency (i.e. word-to-word relations) parsing are the principal approaches in many search concepts, namely: two-level parser [30], function-behavior-state (FBS) information extraction [31], knowledge-based natural language analysis approach [32], concept-based patent search [33], transform queries to Subject–Action–Object-structures [34], and conceptual graph extraction [35]. Wang et al. [30] argued that the proposed Independent Claim Segment Dependency Syntax approach had improved efficiency (i.e. less computer memory and parsing time), as well as identifying some of the challenges in patent parsing due to peculiarities of claim syntax (such as claim template, post attribute past participle, parenthetical sentence, complex noun phrase as sentence, recursion, and coordination).

More coverage on Patentability search during writing a new patent application, Validity/Invalidity search (i.e. to defend a patent application or to litigate a competitor's patent), Infringement search (i.e. freedom to operate search before launching a product on the market), Technology survey, and Portfolio survey could be read from a review article from Bonino et al. [36]. There is no established standardized evaluation method to compare these proposed approaches. Since the best proposed approach accuracy rate is only 68%, there is still large research scope available in patent searches. Prospective hybrid approaches to blend the crowdsourcing approach with computer algorithms need to be explored, and hypothesis “crowd inputs increase algorithmic outcomes efficiency” should be tested.

Patent clustering

Patent clustering is a process of grouping and representing related patents graphically to support and enhance many patent-related applications, such as patent valuation, technology relatedness and competitor analysis, patent strategy development, and technology management. Kitamura et al. [37] sketched possibility of using patent map of functional decomposition to be used for the design review, to indicate applicable ways to achieve a function, and patent application. They reported a patent survey for handling semiconductor wafers which illustrated the differences in working principles and features of patents in each level of function decomposition. They argued that through a function decomposed patent map, the patent application was completed within a week as compared to original 3-4 weeks, and the patent claims were increased, in some cases doubled. However, the proposed software required a user to describe function decomposition trees on a graphical user-interface. Since this is time consuming procedure for designers, an alternative approach (possibly through crowdsourcing) is required to extract functional information from patents, and categorize to create a useful sub-

functional tree. Fu et al. [38] demonstrated through experiments that computationally-generated structure (i.e. using Latent Semantic Analysis (LSA) and hierarchical Bayesian algorithm) is sensible in clustering of patents and organization of clusters (i.e. functional similarity) compared to experts. However, it is clear that there is not necessarily one best way to structure patent clusters, and although there are many patent landscaping commercial software available in the market (such as AcclaimIPTM, Patent iNSIGHT ProTM, ThemeScapeTM) usefulness of those in engineering design is not explored adequately in the literature.

Technology forecasting

Usefulness of patent clustering is largely explored in the technology forecasting domain rather than in other applications. Trappey et al. [39] proposed a methodology which combines patent content clustering and technology life cycle forecasting to find a niche space of RFID technology development in China. Through this approach, they categorized RFID patents into the saturation stage, the mature stage, and the early growth stage. They argued that these classifications help businesses to find good development potential domains. The approach used by them for forecasting is questionable because they used cumulative patent applications (i.e. patent application volume) for forecasting future RFID technology development trends. This approach could overlook many potential gaps for further development. The ontology-based patent clustering approaches are proposed in the literature, and demonstrated for the strategic prediction of development trends and knowledge flows [40, 41].

Kim and Jun [42] analyzed all Apple Inc. patents till now using the graphical causal inference method and the semiparametric Gaussian copula regression model to show the technological trends, and relations between Apple's technologies. The paper produced many associations between Apple's keywords. Whether an expert could generate useful knowledge (i.e. to find vacant technology areas) from the generated technology path diagram is questionable, because the diagram links few terminologies without any semantics (i.e. meanings) between these relationships.

Jin et al. [43] used quality function deployment (QFD) matrices (i.e. technology-product (T-P) QFD and product-market (P-M) along with a semi-automatic technique for extracting keywords from text data and analyzing the similarities to generate the technology-driven roadmap (TRM). The authors argued that the map could be useful to identify profitable markets and promising product concepts based on technology information. The limitations mentioned for this approach are: cannot be applied to all new technology, as some technology does not have similar technology; help of experts

is needed in the process to screen core keywords among text-mining results; and presently the TRM is drawn manually from the generated links.

Patent knowledge transfer

As mentioned before, knowledge transfer is an important step in the search process. Design-by-analogy and TRIZ are the commonly used modes to apply identified appropriate search patents in the engineering design process. The next two sections review the research conducted in these two approaches.

Design-by-analogy

Busby and Lloyd [44] found that solution search activity provided more innovative influences than conservative ones. They have noted that patent-avoiding behavior (i.e. costly infringement) and high margin on product associated consumables motivated repeated searches throughout the design process, and helped cases of significant innovative designs. Design-by-analogy, serendipity, forced analogy, relational words and random input are some of the methods commonly used for solution search and transfer activities [45].

Kurtoglu et al. [46] evaluated a computational method which extracts product design knowledge at two levels of abstraction (the functional level and the component level), and creates procedural rules that depict the mapping between these two levels. Vandevenne et al. [47] proposed a scalable search for systematic biologically inspired design (SEABIRD). SEABIRD represented product and biological elements extracted from patent and biological databases. They demonstrated that the product aspects identified candidate products for design by analogy, and increase the variety and novelty of ideas [48].

Similar to functional hierarchical trees generated by Cascini and Zini [49] to search for patent similarity, Fu et al. [50] presented the results of testing a method for extracting functional analogies based on functional vector space representation [51] from patent databases to assist designers in systematically seeking and identifying analogies. The work extracted vocabulary of functions from a patent database, building on the hierarchical structure of a functional basis [52]. The results demonstrated that the approach produced significantly improved the novelty of solutions generated, but no significant change in the total quantity of solutions generated.

TRIZ

TRIZ stands for the Russian acronym of ‘Teoriya Resheniya Izobretatelskikh Zadatch’ meaning Theory of Inventive Problem Solving (TIPS).

Altshuller developed TRIZ by studying over 1.5 million patents and noticed certain patterns in the evolution of technical systems [53]. A set of universal principles for problem-solving were subsequently identified by observing recurring engineering conflicts and their solutions [54]. The research based on TRIZ is categorized as knowledge-based system to support TRIZ and application of TRIZ in different domains. The four main process steps of TRIZ (problem definition, problem classification and tool selection, solution generation and evaluation) need many support tools to apply generated principles effectively. Although the merits of TRIZ are widely cited, demerits can be less novelty when compared to the intuitive brainstorming method [55], and the difficulty in finding ‘out of the box’ solutions [56].

Although TRIZ emerged from mechanical design, it has been applied to apply in many other domains. Lee et al. [57] refined a structural service design stages based on the TRIZ and the service blueprint approach. The limitations mentioned are support is needed to choose proper service parameters which influences the setting of the contradiction matrix to find the right inventive principles of solutions, and also it is time-consuming task to identify the root cause of problems in the problem definition stage. Kobayashi et al. [58] proposed a product eco-design methodology which integrates TRIZ for idea generation. The noted problems include no principle of invention for some matrix elements of the contradiction matrix of TRIZ, and missing elements to represent environmental characteristics in the contradiction matrix.

Harnessing the crowd

Crowdsourcing (aka micro-outsourcing) is an approach where smaller tasks are posted online for a crowd of people to participate and complete. The term crowdsourcing was coined by Jeff Howe in 2006 as ‘the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call’ [59]. These activities are executed by people via virtual tools who can access tasks, execute them, upload the results and receive various forms of payment using any web browser. This is a labour market open 24/7, with a diverse workforce available to perform tasks quickly and cheaply. The distributed network of human workers provide on-line, “black-box”, reasoning capabilities that can exceed the capabilities of current AI technologies (i.e. genetic algorithms, neural-nets, case-based reasoning) in terms of flexibility and scope.

Although many automatic approaches are proposed for patent searching, still human intervention is required for verification, validation, and providing proper judgements for the outcomes generated between the search procedures. These requirements could be crowdsourced to provide proper feedback to automatic approaches to improve the accuracy rate, also reducing time to be spent by designers in refining the search process. Furthermore, the crowdsourcing approach could be utilized for retrieving appropriate graphical content from patents, where automatic approaches are difficult to apply. In terms of idea generation, proposed computational techniques for design-by-analogy proved that the platforms help designers generate novel solutions. However significant improvements are needed in classification and filtering, and extracting and visualizing complete FBS linkages effectively for proper analogy transformation. This could be supported by the crowdsourcing approach.

Applications of crowdsourcing in design research are developing slowly. Table 1 summarizes the few research work reporting the crowdsourcing approach in engineering design. These reported works could be well summarized within the three research directions proposed by Maher [60]: technology development, creative design processes, and evaluating creativity. However, the focus is predominately on the crowd evaluation process. More applications for using the crowdsourcing approach in the design process need to be identified and demonstrated.

Table 1 Summary of research work reporting the crowdsourcing approach in engineering design

Authors	Aim/Objectives	Used design tasks	Important findings
Wu et al. [61]	Propose ‘Crowdsourced Design Evaluation Criteria’ (cDEC) to support generation and evaluation of crowd-enabled design activities.	Design living room layouts	Effective evaluation of design quality is a key component to leverage virtual workforce’s creative activities, and vital to iterative optimization processes
Yu and Nickerson [62]	Support crowd creativity through an iterative process of design, evaluation and combination.	Chair design	Iterative process of design, evaluation and combination leads crowd to inherit and modify presented features enabling to generate creative product.

Burnap et al. [63]	Propose a simulation-based crowd consensus model to identify experts in crowd for design evaluation.	Bracket strength evaluation	The model predicts experts only if the assumptions made are correct (e.g. only experts have consistent Evaluations).
Maher et al [60]	Support collective design by understanding representation, communication and motivation.	Google Image Labeller, Threadless	Proposed schema for evaluating crowdsourcing platform for representation, communication and motivation.
Sun et al. [64]	Integrate crowd's sketching processes via collaborative crowdsourcing design.	Communal facilities for elderly people's recreation	Crowd rely heavily on the idea tree for inspiration, and best ideas appear around ends of the idea tree.
Luther et al. [65]	Help designers to receive design critiques from non-expert crowd workers.	Poster designs	Aggregated crowd critique approaches expert critique, improved design process, and assists to change designs.
Grace et al. [66]	A process model for crowdsourcing experience design for volunteer online communities.	Citizen science project	Increases motivation of crowd and creativity of the design.
Bao et al. [67]	Compare evaluation methods to increase crowdsourcing system's effectiveness.	Solutions to the oil spill problem in the Mexico Gulf	Argued that evaluation methods should be assigned in relation to the distribution of quality present at each stage of crowdsourcing.
Bayus [68]	Characterizing individual's ideation efforts in crowd.	Dell's <i>IdeaStorm</i> community	Ideators struggle to repeat their success due to fixation.
Poetz and Schreier [69]	Compare novelty, customer benefit, and feasibility of ideas generated between experts and crowd users.	Baby products	Crowd user ideas score higher in novelty and customer benefit, but lower in feasibility.
Vattam and Goel [70]	To catalogue and annotate research articles using the SBF-based approach.	Biological sources	Created Biologue (a social citation cataloguing system) to gather,

			organize, share, and annotate scholarly articles.
--	--	--	---

Crowd capability

The approach of using crowd in analyzing patents and providing valuable information to designers depends on the crowd having basic abilities to understand and process patent information. As patent information is rich in both textual and graphical content, the crowd should possess abilities to process both these content types. To assess the abilities of crowd, crowdsourcing tasks should be framed in a way that are easier to understand and can be answered quickly (e.g. maximum 15 minutes). Considering the requirements of patent understanding and crowdsourcing task design, two initial experiments were conducted within a crowdsourcing platform [71]. This platform was chosen because reportedly the platform has 30 million crowd workers, easy user interface to design crowdsourcing tasks and monitoring tasks completion, and provides access to register jobs from the UK. The first experiment reviews how well the crowd categorized patent written content, and the second how well they interpreted drawing information.

Crowd experiment 1: patent textual information categorization

The description of patents is mostly written with complex technical terminologies. Understanding patent information can be challenging for crowd who predominantly undertake simpler jobs such as text transcription, data cleaning, opinion survey, and image recognition. To know whether crowd has basic ability to read and understand patent text, a data categorization task was designed. The task aims to test classification of function, behavior and structure (FBS) information present in patent abstract. To enable easier understanding for crowd, FBS was presented as ‘What the object does’, ‘How the object works’, and ‘How the object is made’. FBS structure was chosen because of its comprehensiveness, lightweight, and easy to structure crowd task with few instructions. Fig. 2 shows the FBS categorization task posted on Crowdfunder. The task asked crowd to read a patent sentence and categorize it among the given three options. To give more context to the given sentence, patent title and complete patent abstract in which the sentence appears was also given. An example for each category was provided for better understanding the task. For the initial trail, we posted seventeen such tasks, and requested twenty responses for each task. In overall, 340 judgements were received and analyzed. To avoid random clicking, three

test questions were included within this task. The test question highlights the correct response with justification when a crowd member provided a wrong answer, and helped to improve the quality of responses from crowd.

Patent sentence:
 The present invention provides an improved magnetic levitation system for the stable or rigid levitation of a body.

Patent title:
 Levitation device

Patent abstract:
 The present invention provides an improved magnetic levitation system for the stable or rigid levitation of a body. The object to be levitated is maintained in an equilibrium position above a flat guideway or a plurality of continuous guideways. The rigidity of the levitation is at least in part achieved by the orientation of the axis through the poles of the magnet in a direction perpendicular to the direction of relative motion between the magnet and the conductor.

Your task is to read the given sentence (in BIG letters) and to choose the best option which the given sentence describes.

- What the object does
- How the object works
- How the object is made

Fig. 2 FBS patent information categorization task

All 340 responses for this task were received within 1h 10min. In total, 76 crowd workers participated from 29 countries. On average, each worker gave 4 responses (maximum: 12 responses/worker and minimum: 2 responses/worker). **Maximum agreement percentage for each of 17 posted questions (in brackets show the maximum chosen category in FBS)**

Fig. 3 illustrates maximum agreement percentage for each of seventeen posted questions. The average maximum agreement for the posted questions is 55.29%. The maximum agreement ranges from 40% to 70%. Analyzing further with the chosen categories reveals that except for 2 questions, the maximum chosen category for all other 15 questions were answered correctly. In those 2 questions also, the sentence posted contains elements of both function and behavior, so received highest percentage for both these categories. Therefore, the initial test shows that at least 55% of the crowd participated in this task possess abilities to understand and choose the correct FBS category for the given patent sentence. Although 45% of the participated crowd lacks patent understanding, the maximum agreement percentage for each question helps to identify the correct FBS categorization.

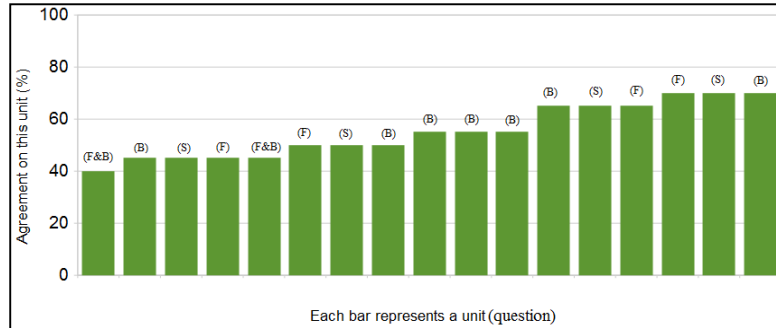


Fig. 3 Maximum agreement percentage for each of 17 posted questions (in brackets show the maximum chosen category in FBS)

Crowd experiment 2: understanding patent drawing information

Patent drawings usually contain numerical annotation for parts, where these numbers are subsequently referred in patent claims and descriptions. These numerical annotations were used to test crowd abilities in understanding patent drawing information. Fig. 4 illustrates the task posted, whereby a drawing from a patent was provided, and crowd asked to describe a particular number in the drawing. The process to go about answering this task was presented as follows: Download the given patent document; Try to understand the patent by reading the abstract; Search the number asked in the question; and Read and understand all paragraphs describing about the enquired number. The initial test was conducted to enquire about 10 numbers on six drawings from a single patent (US 9148077 B2).

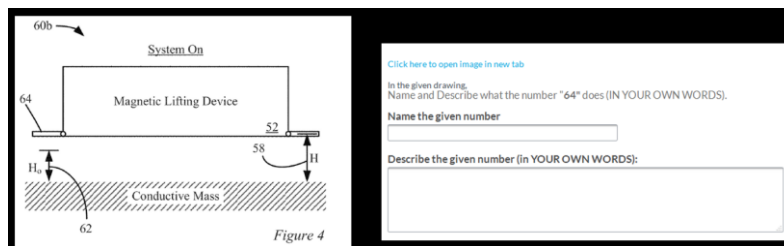


Fig. 4 Task for understanding patent drawing information

Twenty responses for each number were requested and received for this task. One hundred and ninety five responses for this task were received within 2h 40 min. The rest five responses received in 8h 40min. In total, 30 crowd workers were participated from 17 countries. On average, each worker categorized 7 responses (maximum: 10 responses/worker and minimum: 1 response/worker). The results from analyzing description for each

part number reveal that only 24% of the answers received were acceptable (i.e. provided correct description about the number asked). For example, the expected answer for number ‘64’ in Fig. 4 is “*used to give a stand-off height to the magnetic lifting device*”. It was surmised that a major reason for these low percentages compared to the patent textual categorization task was due to not including test questions within this exercise. Although adding test questions are general practice in crowdsourcing, the challenge in this experiment is how to add test questions for qualitative answers (i.e. textual description of a number in patent drawing). We were able to identify a mechanism by asking a crowd to first name the number before adding the description (shown in Fig. 4). By following this approach we could successfully add test questions for qualitative questions. It was therefore decided to develop a preliminary task that required workers to provide the exact name (given in the patent document) of a given annotation number. Only workers who could answer at least 50% of test questions correctly were allowed to progress to undertake the descriptive task. This resulted in a much better percentage of acceptable responses (66%). Fig. 5 summarizes key variables observed with and without test questions in understanding patent drawing information task.

Table 2 Key variables observed between with and without test questions in understanding patent drawing information task

	Without test questions	With test questions
Total task completion time	8h 40min	19h 22min
Number of workers participated (from countries)	30 (17)	53 (27)
Number of workers failed in Quiz	N/A	25
Total number of responses	200	305
Average responses from a worker	7	6
Average percentage of acceptable responses	24%	66%
Accepted responses percentage range	15-50%	44-83%

Comparing the results with and without the test questions reveals that “test questions mode” is preferred to get high number of responses and percentage of acceptable responses (provided correct description about the number asked). However, the total task completion time is almost double compared to “without test questions mode”. These initial tasks emphasize

the importance of test questions to group suitable crowd who are capable to undertake patent related tasks.

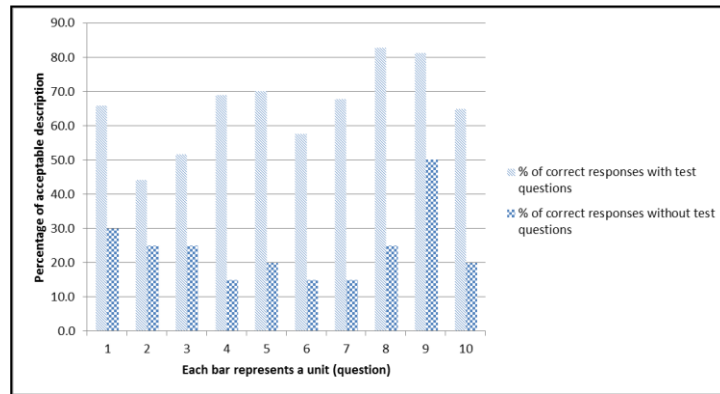


Fig. 5 Acceptable drawing part description percentage for each question

The observations made from these initial two trails are: a large crowd of people from many countries is available on demand to undertake posted tasks; the completion time for the posted tasks is quicker (most responses received in less an hour), although it is dependent on the inclusion of test questions; judgement for categorizing patent textual information works best with aggregating all responses rather than relying on individual responses; clear task instructions, test questions, and payment for each task could play major role in getting acceptable responses; best worker among crowd of people should be chosen with initial tests, th en nurtured and developed further to potentially apply them to higher reasoning patent analyzing tasks.

Discussion

The presented paper provides a broader view of patent applications in engineering design, and potential scope where the crowdsourcing approach could be applied. This paper reports an important initial step through which further work could be developed to achieve the proposed aim: to generate a new, visual form of patent map that incorporates measures of patent commercialization activity and technical metadata through the crowdsourcing approach that can be utilized by engineering designers during conceptualization and embodiment design.

These initial crowdsourcing experiments aid to study two important points: (i) selecting and nurturing a crowd that is suitable for patent analysis work, and (ii) how to study the responses obtained from the crowd. The

study observes that about 55% of the participated crowd have good reasoning ability for FBS categorization. This result is surprising providing the crowd predominantly undertake simpler jobs such as text transcription and data cleaning. The successful crowd from this task could be selected and trained for further advanced patent analysis tasks. Since there is uncertainty to trust individual crowd responses, the study finds that cumulative percentage response (in this case, 20 responses for each question) aids to identify correct FBS categorization for each question. Therefore we intend to use cumulative aggregation of crowd responses for any patent related tasks to get correct responses.

There is no acceptable/target percentage set for these crowdsourcing tasks, considering there are 30 million available crowd workers. The percentage (e.g. 66% correct in patent drawing information task) signifies how many people from the participated crowd have shown interest and possesses minimum ability to be selected for further skills improvement. The observed correct answer percentage variation in both the initial patent tasks is expected, considering complexity of patent information and varying crowd skills level. For instance, some numbers in patent drawing are described in a straight-forward fashion, whereas others are hidden within the text. This makes patent text analysis complex for humans and also computer algorithms. However, more systematic study should be undertaken to understand patterns in this percentage variation. It is difficult to answer the trust level established with crowd from these initial experiments. However, with proper crowd nurtured for patent analysis we are able to place increased trust in responses from them.

Apart from issues related to getting acceptable responses quickly, the following limitations are noted from our initial discussion with industrial partners for using the crowdsourcing approach: (i) confidentiality issues, (ii) trust in the crowd's responses, (iii) payment issues, and (iv) ethical and legal issues. Approaches to address these issues and increasing percentage of correct responses for patent related tasks are currently explored in this research.

Conclusions and future work

This paper has illustrated the relevance of patents for engineering design. While the analysis of patent statistics such as classification, nationality and use of citations is well established, making robust interpretations based on these is problematic. Qualitative interpretations of the content and nature of patents is potentially more useful to engineering designers in terms of identifying active areas of design, examples of patents in use and

understanding the characteristics of design problems. These cannot, however, be easily accomplished by computer algorithms and their magnitude is overwhelming for individuals. If crowdsourcing proves a cheap, scalable way of interpreting patents and applying appropriate taxonomic engineering information it could fundamentally alter the early phases of engineering design. The crowd can be harnessed to improve the data and information presented to designers during the key activities identified in scoping (opportunities), generation (inspiration), embodiment (context) and testing (checking).

Our initial crowdsourcing experimentation has been to establish basic crowd capabilities in understanding patent text and interpreting patent drawings. This has shown that reasonable results can be achieved if tasks are set appropriately, particularly in terms of duration and complexity, and if test questions are incorporated to ensure a basic level of understanding exists in the workers. Our planned future work will involve expanding on these initial experiments to design and test crowdsourcing workflows optimised to support working in the four design phases described. It is anticipated this will focus on analysing quality, classification and composition and content to present patent information in a way that is readily understandable and usable by engineering designers.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) grant number EP/N005880/1.

References

1. Brown, T., *Change By Design*. 2009, New York, NY: HarperCollins.
2. Pugh, S., *Total Design*. 1991, Reading, UK: Addison-Wesley.
3. Kanda, A.S., *An Investigative Study Of Patents From An Engineering Design Perspective*. Master Thesis, Clemson University, 2008.
4. Chan, J., et al., *On the effective use of design-by-analogy: The influences of analogical distance and commonness of analogous designs on ideation performance*, in *International Conference On Engineering Design, ICED11*. 2011: Copenhagen, Denmark.
5. Chang, H.-T., C.-Y. Chang, and Y.-P. Yang. *Combining surveying patent information, reappearing problem and discovering breakthrough for design-around*. in *DS 75-9: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, Vol. 9: Design Methods and Tools, Seoul, Korea, 19-22.08. 2013*. 2013.
6. Lloveras, J. *A process of conceptual engineering design for new patentable products*. in *DS 68-8: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 8: Design Education, Lyngby/Copenhagen, Denmark, 15.-19.08. 2011*. 2011.

7. Chakrabarti, A., Khadilkar, P. , *A measure for assessing product novelty*. Proceedings of the International conference on Engineering Design, Stockholm, 2003.
8. Saunders, R., *Curious design agents and artificial creativity a synthetic approach to the study of creative behavior*, in *Ph.D. thesis, Department of Architectural and Design Science, Faculty of Architecture, University of Sydney*. 2002.
9. Shah, J.J. and N. Vargas-Hernandez, *Metrics for measuring ideation effectiveness*. Design Studies, 2003. **24**(2): p. 111-134.
10. Lopez-Mesa, B.a.V., R. . *Novelty metrics in engineering design experiments*. in *Proceedings of the International Design Conference Design, Dubronik, Croatia*. 2006.
11. Felk, Y., et al. *Designing patent portfolio for disruptive innovation—a new methodology based on CK theory*. in *DS 68-2: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 2: Design Theory and Research Methodology, Lyngby/Copenhagen, Denmark, 15.-19.08. 2011*. 2011.
12. Sarkar, P. and A. Chakrabarti, *Assessing design creativity*. Design Studies, 2011. **32**(4): p. 348-383.
13. Kang, I.-S., et al., *Cluster-based patent retrieval*. Information Processing & Management, 2007. **43**(5): p. 1173-1182.
14. Commission, A.L.R., <http://www.alrc.gov.au/publications/6-patentability-genetic-materials-and-technologies/usefulness>. 2015.
15. Nicol, D. and J.L. Nielsen, *Patents and Medical Biotechnology: An Empirical Analysis of Issues Facing the Australian Industry*. SSRN Electronic Journal.
16. USPTO. *Guidelines for Examination of Applications for Compliance with the Utility Requirement [R-11.2013]*. 2013.
17. European Patent Convention, (*entered into force on 7 October 1977*), art 52(1). 1977.
18. Intellectual Property Office, *The Patent Guide: A handbook for analysing and interpreting patent data*. 2015.
19. (1959), N.R.D.C.v.C.o.P., *102 CLR 252, 275*. 1959.
20. Gambardella, A., D. Harhoff, and B. Verspagen, *The value of European patents*. European Management Review, 2008. **5**(2): p. 69-84.
21. Bioethics, N.C.o., *The Ethics of Patenting DNA*, in *31*. 2002.
22. Co, F.B.R., *Submission P84, 16 April 2004; Centre for Law and Genetics, Submission P104*. 2004.
23. Trappey, A.J.C., et al., *A patent quality analysis for innovative technology and product development*. Advanced Engineering Informatics, 2012. **26**(1): p. 26-34.
24. Chen, S.-H., M.-H. Huang, and D.-Z. Chen, *Exploring technology evolution and transition characteristics of leading countries: A case of fuel cell field*. Advanced Engineering Informatics, 2013. **27**(3): p. 366-377.
25. Roy, R., *Case studies of creativity in innovative product development*. Design Studies, 1993. **14**(4): p. 423-443.
26. Heisig, P., et al., *Exploring knowledge and information needs in engineering from the past and for the future – results from a survey*. Design Studies, 2010. **31**(5): p. 499-532.

27. Rodgers, P.A. and P.J. Clarkson, *An Investigation and Review of the Knowledge Needs of Designers in SMEs*. The Design Journal, 1998. **1**(3): p. 16-29.
28. Koh, E.C.Y., *Engineering design and intellectual property: where do they meet?* Res Eng Design, 2013. **24**(4): p. 325-329.
29. Hitchcock, D., *Patent searching made easy*, ed. R. 3rd edn. Lulu Press, NC. 2005.
30. Wang, J., W.F. Lu, and H.T. Loh, *A two-level parser for patent claim parsing*. Advanced Engineering Informatics, 2015. **29**(3): p. 431-439.
31. Fantoni, G., et al., *Automatic extraction of function-behaviour-state information from patents*. Advanced Engineering Informatics, 2013. **27**(3): p. 317-334.
32. Sheremetyeva, S., *Natural language analysis of patent claims*, in *Proceedings of the ACL-2003 workshop on Patent corpus processing -*. 2003, Association for Computational Linguistics (ACL).
33. Montecchi, T., D. Russo, and Y. Liu, *Searching in Cooperative Patent Classification: Comparison between keyword and concept-based search*. Advanced Engineering Informatics, 2013. **27**(3): p. 335-345.
34. Jeong, J.-S., C. Park, and K.-H. Yoo, *Hand Gesture User Interface for Transforming Objects in 3D Virtual Space*, in *Multimedia, Computer Graphics and Broadcasting*, T.-h. Kim, et al., Editors. 2012, Springer Berlin Heidelberg. p. 172-178.
35. Yang, S.-Y. and V.-W. Soo, *Extract conceptual graphs from plain texts in patent claims*. Engineering Applications of Artificial Intelligence, 2012. **25**(4): p. 874-887.
36. Bonino, D., A. Ciaramella, and F. Corno, *Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics*. World Patent Information, 2010. **32**(1): p. 30-38.
37. Kitamura, Y., et al., *Deployment of an ontological framework of functional design knowledge*. Advanced Engineering Informatics, 2004. **18**(2): p. 115-127.
38. Fu, K., et al., *Expert representation of design repository space: A comparison to and validation of algorithmic output*. Design Studies, 2013. **34**(6): p. 729-762.
39. Trappey, C.V., et al., *Using patent data for technology forecasting: China RFID patent analysis*. Advanced Engineering Informatics, 2011. **25**(1): p. 53-64.
40. Trappey, C.V., et al., *A knowledge centric methodology for dental implant technology assessment using ontology based patent analysis and clinical meta-analysis*. Advanced Engineering Informatics, 2014. **28**(2): p. 153-165.
41. Bermudez-Edo, M., et al., *Analyzing a firm's international portfolio of technological knowledge: A declarative ontology-based OWL approach for patent documents*. Advanced Engineering Informatics, 2013. **27**(3): p. 358-365.
42. Kim, J.-M. and S. Jun, *Graphical causal inference and copula regression model for apple keywords by text mining*. Advanced Engineering Informatics, 2015. **29**(4): p. 918-929.
43. Jin, G., Y. Jeong, and B. Yoon, *Technology-driven roadmaps for identifying new product/market opportunities: Use of text mining and*

- quality function deployment*. *Advanced Engineering Informatics*, 2015. **29**(1): p. 126-138.
44. Busby, J.A. and P.A. Lloyd, *Influences on Solution Search Processes in Design Organisations*. *Research in Engineering Design*, 1999. **11**(3): p. 158-171.
 45. Zusman, A.a.Z., B., *Overview of creative methods*. *The TRIZ Journal* [online], 1999.
 46. Kurtoglu, T., M.I. Campbell, and J.S. Linsey, *An experimental study on the effects of a computational design tool on concept generation*. *Design Studies*, 2009. **30**(6): p. 676-703.
 47. Vandevenne, D., et al., *SEABIRD: Scalable search for systematic biologically inspired design*. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 2015: p. 1-18.
 48. Verhaegen, P.-A., et al., *Effectiveness of the PAnDA ideation tool*. *Procedia Engineering*, 2011. **9**: p. 63-76.
 49. Cascini, G. and M. Zini, *Measuring patent similarity by comparing inventions functional trees*, in *The International Federation for Information Processing*. 2008, Springer Science + Business Media. p. 31-42.
 50. Fu, K., et al., *Design-by-analogy: experimental evaluation of a functional analogy search methodology for concept generation improvement*. *Res Eng Design*, 2014. **26**(1): p. 77-95.
 51. Murphy, J., et al., *Function Based Design-by-Analogy: A Functional Vector Approach to Analogical Search*. *Journal of Mechanical Design*, 2014. **136**(10): p. 101102.
 52. Otto K, W.K., *Product design techniques in reverse engineering and new product development*, ed. N.J. Upper Saddle River. Prentice Hall. 2001.
 53. Altshuller, G., H. Altov, and L. Shulyak, *TRIZ, the theory of inventive problem solving*. 1994, Worcester, MA: Technical Innovation Center Inc.
 54. Altshuller, G., *The innovation algorithm: TRIZ, systematic innovation and technical creativity*. Worcester, MA: Technical Innovation Center., 1999.
 55. Chulvi, V., et al., *Influence of the type of idea-generation method on the creativity of solutions*. *Research in Engineering Design*, 2012: p. 1-9.
 56. Reich, Y., et al., *A theoretical analysis of creativity methods in engineering design: casting and improving ASIT within C-K theory*. *Journal of Engineering Design*, 2012. **23**(2): p. 137-158.
 57. Lee, C.-H., Y.-H. Wang, and A.J.C. Trappey, *Service design for intelligent parking based on theory of inventive problem solving and service blueprint*. *Advanced Engineering Informatics*, 2015. **29**(3): p. 295-306.
 58. Kobayashi, H., *A systematic approach to eco-innovative product design based on life cycle planning*. *Advanced Engineering Informatics*, 2006. **20**(2): p. 113-125.
 59. Howe, J., *The rise of crowdsourcing*. *Wired magazine*, 2006. **14**(6): p. 1-4.
 60. Maher, L.M., *Design Creativity Research: From the Individual to the Crowd*, in *Design Creativity 2010*, T. Taura and Y. Nagai, Editors. 2011, Springer London: London. p. 41-47.
 61. Wu, H., J. Corney, and M. Grant, *An evaluation methodology for crowdsourced design*. *Advanced Engineering Informatics*, 2015. **29**(4): p. 775-786.
 62. Yu, L. and J.V. Nickerson, *Cooks or cobblers?: crowd creativity through combination*, in *Proceedings of the SIGCHI Conference on Human*

- Factors in Computing Systems*. 2011, ACM: Vancouver, BC, Canada. p. 1393-1402.
63. Burnap, A., et al., *When Crowdsourcing Fails: A Study of Expertise on Crowdsourced Design Evaluation*. *Journal of Mechanical Design*, 2015. **137**(3): p. 031101-031101.
 64. Sun, L., et al., *Collaborative sketching in crowdsourcing design: a new method for idea generation*. *International Journal of Technology and Design Education*, 2014. **25**(3): p. 409-427.
 65. Luther, K., et al., *Structuring, Aggregating, and Evaluating Crowdsourced Design Critique*, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 2015, ACM: Vancouver, BC, Canada. p. 473-485.
 66. Grace, K., et al., *A Process Model for Crowdsourcing Design: A Case Study in Citizen Science*, in *Design Computing and Cognition '14*, S.J. Gero and S. Hanna, Editors. 2015, Springer International Publishing: Cham. p. 245-262.
 67. Bao, J., Y. Sakamoto, and J.V. Nickerson. *Evaluating Design Solutions Using Crowds*. in *Seventeenth Americas Conference on Information Systems*. 2011. Detroit, USA.
 68. Bayus, B.L., *Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community*. *Management Science*, 2013. **59**(1): p. 226-244.
 69. Poetz, M.K. and M. Schreier, *The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?* *Journal of Product Innovation Management*, 2012. **29**(2): p. 245-256.
 70. Vattam, S.S. and A.K. Goel, *Semantically annotating research articles for interdisciplinary design*, in *Proceedings of the sixth international conference on Knowledge capture - K-CAP '11*. 2011, Association for Computing Machinery (ACM).
 71. Crowdfunder. 2015 [22nd October]; Available from: <https://make.crowdfunder.com>.