AUTOMATED FUNCTIONAL ANALYSIS OF PATENTS FOR PRODUCING DESIGN INSIGHT

Jiang, Pingfei;
Atherton, Mark;
Sorce, Salvatore

Department of Mechanical and Aerospace Engineering, Brunel University London

ABSTRACT

Patent analysis is a popular topic of research. However, designers do not engage with patents in the early design stage, as patents are time-consuming to read and understand due to their intricate structure and the legal terminologies used. Manually produced graphical representations of patent working principles for improving designers' awareness of prior art have been demonstrated in previous research. In this paper, an automated approach is presented, utilising Natural Language Processing (NLP) techniques to identify the invention working principle from the patent independent claims and produce a visualisation. The outcomes of this automated approach are compared with previous manually produced examples. The results indicate over 40% match between the automatic and manual approach, which is a good basis for further development. The comparison suggests that the automated approach works well for features and relationships that are expressed explicitly and consistently but begin to lose accuracy when applied to complex sentences. The comparison also suggests that the accuracy of the proposed automated approach can be improved by using a trained part-of-speech (POS) tagger, improved parsing grammar and an ontology.

Keywords: Functional modelling, Early design phases, Semantic data processing, Design informatics, Visualisation

Contact:
Jiang, Pingfei
Brunel University London
Mechanical and Aerospace Engineering
United Kingdom
pingfei.jiang@brunel.ac.uk

1 INTRODUCTION

The number of worldwide patent applications grows around 8% per year, with China, US and Japan taking the lead (WIPO, 2019). The increasing number of patents and the consequent increasing difficulty in patent analysis contribute to the growing number of IP disputes, highlighting the need for improved awareness of prior art during the design process (Sorce *et al.*, 2018). Studying patents should help designers to both learn whether their design/invention infringes existing unexpired patents (Ulrich and Eppinger, 2016), and also significantly influence their innovation (Koh, 2020). Studying design concepts and their physical realisations (Li *et al.*, 2012) makes patents a promising source of design inspiration to enhance ideation (Wodehouse *et al.*, 2017, 2018). However, designers do not often engage with patents, especially in the early stages of design, as analysis of patents requires considerable effort and expertise to extract the technical information buried within the legal terminologies used (Kim *et al.*, 2008; Tseng *et al.*, 2007).

In previous work, the authors developed a graphical representation of invention working principles built upon Functional Analysis Diagram (FAD) (Aurisicchio *et al.*, 2013), intended to represent a patented invention more succinctly for designers (Atherton *et al.*, 2017). The information required for these graphical representations is identified and extracted manually from a patent, which is time-consuming and not practical for analysing a large number of patents. Therefore, this approach has been automated and is presented in this paper. This automated approach, developed in Python, utilises Natural Language Processing (NLP) techniques to identify and extract working principles from patent independent claims. A graph plotting feature directly produces a graphical representation of the working principle based on patent claims. The effectiveness of this automated approach is evaluated by comparing the results with those produced manually.

2 BACKGROUND

2.1 Patent analysis to support engineering design

As a large and freely accessible database, patents are often used to support engineering design. Based on the information contained within a patent, analysis of patents can be categorised into metadata-based and content-based, with four common analysis applications (different shapes used for easy identification), presented in Figure 1.
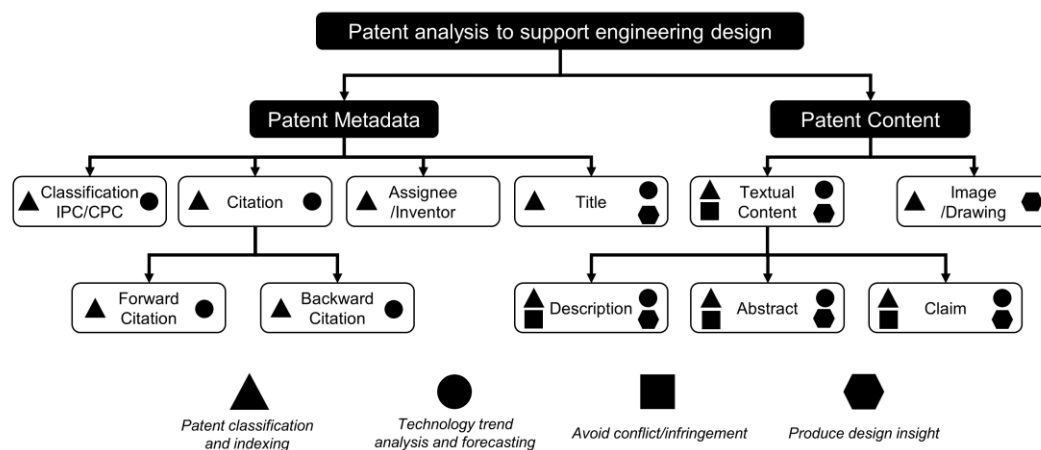

Figure 1. Four types of patent analysis applications to support engineering design

Analysis of patent metadata mainly contributes to business and strategic operations such as classification, technology trend and competitor analysis. For example, a hybrid patent classification approach is proposed by Liu and Shih (2011) that uses patent metadata to predict patent classes and is demonstrated to be more accurate than conventional approaches. Trappey *et al.* (2012) developed a patent analysis methodology to rank patent quality, using indicators and component analysis based on patent metadata. Patent metadata contains little information concerning the invention working principle itself so are rarely used in performing design analysis. Whereas, patent content contains detailed information regarding the invention revealed and so is broadly used to support design activities. Contents of a patent document differ from country to country, but the most common sections include abstract,

description, claims and drawings (images). Patent drawings provide visualisations of the design which often improves a designer's understanding of the invention but drawings alone cannot effectively capture the technical features or working principles of a design (Li *et al.*, 2014).

The patent claims are arguably the most important amongst all the textual sections as they fully define the scope of an invention. They are also the key legal instrument for determining infringement in a judicial case (Koster, 2015). Patent claims can be classified into independent claims and dependent claims. Independent claims are self-contained, describing the invention in its broadest scope. Dependent claims refer to further detail of features described in the independent or dependent claims they are referring to. The first independent claim defines the essential technical features that distinguish the invention from what is already known in the field. The patent abstract is often drafted last to offer a concise summary of the disclosed invention. From an engineering design perspective, the patent abstract can be seen as the most informative part of a patent document (Liu and Shih, 2011). Unlike the patent claims section, the abstract is expressed more closely to natural language therefore should be easier to understand by designers. Both patent independent claims and the abstract can offer valuable information covering all the four types of analysis depicted in Figure 1, with the independent claims used for the example in this paper.

2.2 Functional analysis to support engineering design

Functional analysis explores interrelationships between technical features within a system and is considered to be fundamental to engineering design (Pahl *et al.*, 2007). Graphical representations of functional analysis have the benefit of providing designers with an intuitive understanding of the system (Aurisicchio *et al.*, 2013). The Functional Analysis Diagram (FAD) for example, first disclosed by Invention Machine Corp (Devoino *et al.*, 2005), has been developed into Kinetic Functional Analysis Diagram (KFAD) (Lee *et al.*, 2013) by adding representations of movement transitions to assist mechanical design; and combined with Value Engineering and Failure Mode and Effect Analysis (FMEA) to improve product reliability and cost (Michalakoudis *et al.*, 2016).

FAD has been applied in the patent analysis (Atherton *et al.*, 2017), being further developed (e.g. into FAD+) to incorporate invention hierarchy relationships that use a domain-specific ontology to enable conflict analysis between emerging design and prior art (Jiang *et al.*, 2017, 2018). However, this process requires input from the designers, which makes it time-consuming to apply in practical situations. Cascini and Russo (2007) developed an automated functional diagram approach that produces graphical representations from patent claims. However, the method relies on the univocally defined numbers for invention feature identification and hence does not apply to patent claims or abstract that contain no numbering. Fantoni *et al.* (2013) developed an automated approach for producing functional diagrams from the patent detailed description. The approach requires statistically processed domain-dependent knowledge. This explains the use of the detailed description section of a patent as it is normally the lengthiest section and therefore can provide statistically significant results compared to the use of abstract and claims. Also, using a detailed description, the amount of information analysed and presented can be overwhelming hence against the overall purpose of saving time.

The automated functional analysis approach presented in this paper addresses the issues mentioned above by applying a series of NLP techniques to identifying invention features and their relationships from patent independent claims without the need of using numbering and lengthy detail description. Its main purpose is to provide quick design insight through a simple graphical representation of invention working principles.

3   METHODOLOGY

Figure 2 presents an overview of the methodology developed by the authors for performing automated patent functional analysis. NLP techniques are core to the approach and an established, ready-to-use, Python NLP Toolkit, NLTK (Bird *et al.*, 2009) is applied to obtain quick results. NLTK contains all the NLP packages necessary for performing the patent independent claim analysis, starting with word tokenisation. Unlike a typical paragraph composed of multiple sentences, a patent independent claim is one single sentence, hence word tokenisation is applied directly. Part-of-speech (POS) tagging is then applied to assign labels to each tokenised word. The next step is to lemmatise noun phrases (e.g. POS tags starts with N) only to eliminate duplicate expressions such as *'battery'* and *'batteries'*. This is accomplished by first converting NLTK tags to WordNet tags and then applying the WordNet

Lemmatiser. The reason for lemmatising noun phrases only while keeping other phrases unchanged is to maintain the accuracy of parsing performed later on. Before parsing, stopwords are removed from the text by referring to a customised list (see Table 1). The use of a customised list is because stopwords that are built in the NLTK corpus contain too many meaningful preposition phrases such as *'in'*, *'of'*, *'from'*, and *'to'*, making it inaccurate for analysing patents. The next step is to perform parsing using NLTK Regular Expressions to identify different types of chunks including invention features and relationships.
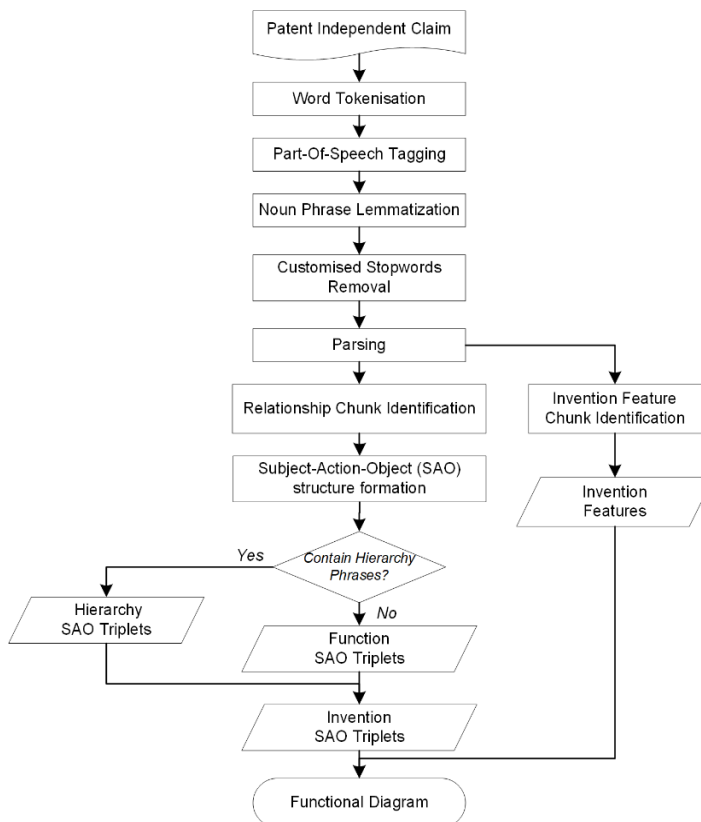
Figure 2. Methodology for automated functional analysis

Table 1. Example list of customised stopwords

a	an	said	this	that	these	those	the	which	is	are	being

The parsed results, in the form of an NLTK tree, are then post-processed for recognition of invention features, plus hierarchical and functional relationships. Subject-Action-Object (SAO) structure is used to transform the parsed results into a cluster of triplets centred around each relationship (Action) chunk identified. A recognised relationship chunk is in some cases a verb phrase such as *'for generating'*, or an adjective phrase such as *'smaller than'*. For each relationship chunk, its Subject is considered to be the next invention feature chunk on its left and its Object is considered to be the next invention feature chunk on its right. When navigating the resultant NLTK tree, if the chuck next to a relationship (Action) is not an invention feature chunk then the algorithm will go further left until an invention feature chunk can be identified. The relationship chunks for all formed SAO triplets will go through a comparison with a list of hierarchy phrases (see Table 2), to distinguish hierarchical relationships from functional interactions. Matched SAO triplets will be removed from the cluster and stored separately, grouped as hierarchy SAO triplets. The remainder will be considered as function SAO triplets. Finally, the invention SAO triplets will be plotted using Graphviz (Bank, 2020) with hierarchy SAO and function SAO plotted differently to provide visualisation of an invention working principle.

Table 2. Example list of hierarchy phrases

comprise	comprises	comprising	consist	consists	consisting	consist of
consists of	compose	composes	composing	composed of	has	have
having	include	includes	including	contain	contains	containing

4 RESULTS

In this section, the proposed automated functional analysis approach for patent claims is applied to three patents whose manual functional analysis have already been demonstrated in previous work (Atherton *et al.*, 2017; Jiang *et al.*, 2018), which are:

- US2615610 - Container with reclosure and shield device
- US3334775 - Gated can lid
- EP2219961 - Resealable beverage can end and methods relating to same

Detailed comparison of the three patents can be found in Table 3 to Table 5, corresponding to invention features, invention hierarchy SAO and function SAO triplets, respectively. Manual outcomes are used as a reference, shown in blue. In Table 3 both green and orange represent automated outcomes, in which matched results indicate features that have been identified successfully compared to manual outcomes whereas unmatched results indicate features that do not belong. For Table 4 and 5 matched results are highlighted by coloured lines. Features with a bracket, e.g. aperture (20) are the exact phrases used in patents.

Table 3. Comparison of the invention features by manual (blue) and automated (green + orange) approaches.

	Manual	Automated	
		Matched	**Unmatched**
US2615610	Container Metallic wall section Shield Reclosure element Score line Opening Depressible area Dispensing opening Raw cutting edge	container metallic wall section shield reclosure element in shield for metallic wall section thereof score line opening depressible area dispensing opening for container raw cutting edge of metal	material of shield projecting over score line opening opposite breaking of score line opening in container
US3334775	Gated can lid Unitary structure Flat sheet metal can lid member Aperture Gate panel Outward underfold Spacer strip Inward underfold Score cut Aperture edge Spacer strip outer edge Gate panel edge	unitary structure edge of aperture as 180-degree outward underfold 180-degree inward underfold gate panel flat sheet metal aperture in surface of member spacer strip	gated member gate panel underneath aperture surface of lid metal sheet outer edge of spacer strip as 180-degree inward underfold arrangement narrow spacer strip at underside of lid metal sheet portion constituting
EP2219961	Can end combination Metal can end Center panel Resealable closure Peripheral wall Upper surface Lower surface Aperture Base plate Top plate Flange seal Bore seal	combination resealable closure (30) center panel (16) upper surface aperture (20) closure (30) top plate (34) peripheral wall (12) lower surface base plate (32) bore seal	center panel (16) about aperture (20) intermediate position in closure (30) proximate aperture (20) fully open position in aperture (20) intermediate position resealed position metal base plate at first location sealed position seal position translatable fully open position engagement with center panel (16) from intermediate position

Table 4. Comparison of invention hierarchy SAO triplets

	Manual	Automated
US2615610	Container — having — Metallic wall section Container — , and — Shield Metallic wall section — in....thereof — Score line Shield — having — Opening Reclosure element — in — Shield	container — having — metallic wall section thereof score line metallic wall section — having — opening opposite
US3334775	Gated can lid — comprising — Unitary structure	gated — comprising — unitary structure
EP2219961	Can end combination — comprising — Metal can end Can end combination — comprising — Resealable closure Metal can end — comprising — Peripheral wall Metal can end — comprising — Center panel Center panel — including — Upper surface Center panel — including — Lower surface Center panel — including — Aperture Resealable closure — comprising — Base plate Resealable closure — comprising — Top plate	combination — comprising — metal resealable closure (30) — comprising — peripheral wall (12) center panel (16) — including — upper surface closure (30) — comprising — base plate (32) closure (30) — having — sealed position

Table 5. Comparison of identified invention function SAO triplets

	Manual	Automated
US2615610	Score line — setting off — Depressible area Depressible area — produce — Dispensing opening Shield — dispose over — Metallic wall section Opening — opposite to — Depressible area Opening — smaller than — Depressible area Shield — projecting over — Score line Shield — shielding — Raw cutting edge Score line — result — Raw cutting edge Reclosure element — reclosing — Dispensing opening	metallic wall section thereof score line — setting — depressible area depressible area — to produce — dispensing opening for container shield — disposed over — metallic wall section material of shield projecting over score line — for shielding — raw cutting edge of metal raw cutting edge of metal — resulting — breaking of score line depressible area — to produce — opening depressible area — dispensing — opening reclosure element in shield for — reclosing — opening in container reclosure element in shield for — dispensing — opening in container opening opposite — smaller than — depressible area
US3334775	Flat sheet metal can lid member — form — Unitary structure Flat sheet metal can lid member — provide — Aperture Aperture — in the surface of — Flat sheet metal can lid member Flat sheet metal can lid member — provide — Gate panel Gate panel — underneath — Aperture Gate panel — close — Aperture Flat sheet metal can lid member — underfolded about — Aperture edge Flat sheet metal can lid member — as — Outward underfold Outward underfold — form — Spacer strip Spacer strip — at the underside of — Flat sheet metal can lid member Spacer strip — about — Aperture edge Flat sheet metal can lid member — infolded about — Spacer strip outer edge Flat sheet metal can lid member — as — Inward underfold Inward underfold — merge into — Flat sheet metal can lid member Flat sheet metal can lid member — constituting — Gate panel Score cut — about — Gate panel edge Score cut — adjacent to — Inward underfold Score cut — permit — Gate panel Gate panel — severed from — Spacer strip	unitary structure — formed — flat sheet metal unitary structure — by drawing — flat sheet metal unitary structure — shaping — flat sheet metal member — to provide — aperture in surface of member gate panel underneath aperture — to normally close — arrangement surface of lid — underfolded about — edge of aperture as 180-degree outward underfold edge of aperture as 180-degree outward underfold — to form — narrow spacer strip at underside of lid metal sheet — then infolded about — outer edge of spacer strip as 180-degree inward underfold outer edge of spacer strip as 180-degree inward underfold — to merge into — metal sheet portion constituting 180-degree inward underfold — adapted — gate panel 180-degree inward underfold — to permit — gate panel gate panel — to be — spacer strip gate panel — severed from — spacer strip
EP2219961	Resealable closure — coupled to — Metal can end Aperture — formed there-through — Center panel Top plate — coupled to — Base plate Base plate — contact — Center panel Top plate — contact — Center panel Resealable closure — proximate — Aperture Resealable closure — expose — Aperture Base plate — downwardly moveable relative to — Top plate Base plate — translatable together — Top plate Base plate — engage with — Center panel Base plate — form — Flange seal Center panel — form — Flange seal Base plate — form — Bore seal Center panel — form — Bore seal	resealable closure (30) — coupled — peripheral wall (12) upper surface — opposing — lower surface aperture (20) — formed therethrough — closure (30) top plate (34) — coupled — base plate at first location center panel (16) about aperture (20) — to form — seal intermediate position in closure (30) proximate aperture (20) — not sealed — fully open position in aperture (20) fully open position in aperture (20) — exposed — aperture (20) fully open position in aperture (20) — to enable — aperture (20) fully open position in aperture (20) — pouring — aperture (20) top plate (34) — moved from — position top plate (34) — sealed — position intermediate position — characterized by — base plate (32) resealed position — forming — bore seal translatable — together relative to — intermediate position translatable — together relative to — fully open position base plate (32) — moveable into — engagement with center panel (16) from intermediate position

From Table 3 it can be seen that the automated approach results in some 'noisy' invention features identified, highlighted in orange. Several 'noisy' features identified overlap with matched ones highlighted in green, e.g. *'opening in container'* compared to *'opening'* and *'container'* for patent US2615610. Looking at the matched invention features (green) between manual and automated outcomes, a reasonable degree of coverage is observed, indicating that the automated approach can identify most features as a designer would do. When considering the proportion of matched results amongst all features (green + orange) identified automatically, the effectiveness needs improvement. Worse results are observed in Table 4 and Table 5, especially for EP2219961 in which only 3

invention hierarchy SAO triplets and 4 function SAO triplets were found to be matching with manual outcomes. The effectiveness of the automated approach is calculated by multiplying the ratios between the number of matched results and the number of automatic results, and the ratio between the number of matched results and the number of manual results, expressed in Equation 1.

$$Effectiveness\ (\%) = \frac{No.\ of\ matched\ results}{No.\ of\ automatic\ results} \times \frac{No.\ of\ matched\ results}{No.\ of\ manual\ results} \qquad (1)$$

The effectiveness of the automated approach for analysing the three example patents is shown in Table 6. An average is taken to provide an overview of the effectiveness of the approach proposed. From the results, the automated approach is overall approx. 43% effective in capturing the working principles of inventions. From the scattered results it can be seen that the effectiveness of the approach is dominated by the patents themselves, e.g. how working principles are described differently.

Table 6. Summary of manual-automated comparison of patent claims functional analysis.

	No. of manual results	No. of automatic results	No. of matched results	Effectiveness	Average Effectiveness
	Invention features				
US2615610	9	13	9	69%	
US3334775	12	16	7	26%	46%
EP2219961	12	24	11	42%	
	Invention hierarchy SAO triplets				
US2615610	5	2	1	10%	
US3334775	1	1	1	100%	43%
EP2219961	9	5	3	20%	
	Invention function SAO triplets				
US2615610	9	10	7	54%	
US3334775	19	13	12	58%	40%
EP2219961	14	16	4	7%	

Figure 3 presents an example comparison between the functional diagram generated manually (Figure 3a) and automatically (Figure 3b) for patent EP221996. The auto-generated functional diagram has many isolated SAO triples while manually generated functional diagram features are all interconnected. The layout of the auto-generated functional diagram is configured without manual intervention.
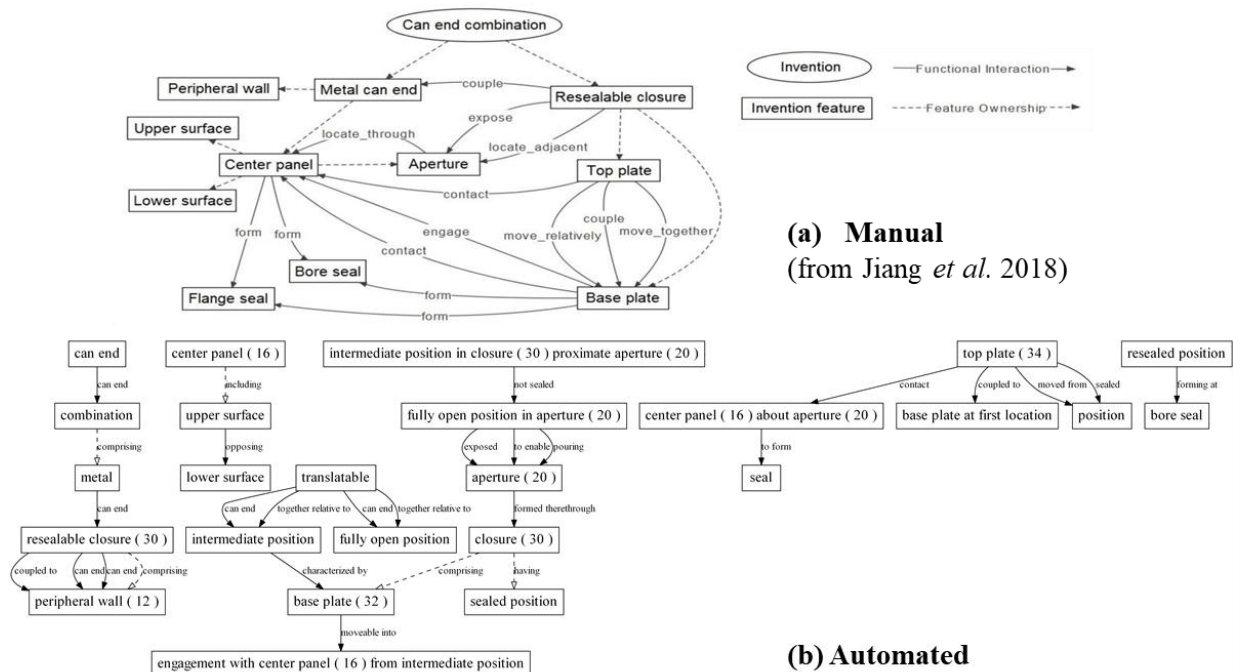
Figure 3. Functional diagrams for EP221996, produced (a) manually and (b) automatically

5 DISCUSSION

Compared to the manual approach which requires roughly 10 minutes (Jiang *et al.*, 2018), the automated approach only takes 3 seconds to produce the diagram and associated outcomes, suggesting that designers can save time, especially when analysis of a large number of patents is required. Comparison results between the automated approach and the manual approach for three previously processed patents indicate reasonable effectiveness therefore this can be seen as a good basis for further development. Main causes for the ineffectiveness of the automated approach are:

1. **The same feature with varied expressions is treated as a distinct invention feature.** For example, *'opening'*, *'opening in container'* and *'dispensing opening for container'* from patent US2615610 are regarded as three features but in fact, they refer to the same thing. This leads to the generation of 'noisy' data that would be avoided by a designer.

2. **Inaccurate POS tags assigned**. The default NLTK POS tagger used in some cases cannot assign correct POS tags which leads to inaccurate identification of both invention features and relationships. This is mainly due to the use of homonyms. For example, *'can end combination'* from patent EP2219961 should be one single noun phrase but the NLTK POS tagger sees *'can'* as a modal phrase and *'end'* as a verb. *'material of shield projecting over score line'* from patent US2615610 is describing a relationship of *'projecting over'* from feature *'material of shield'* to *'score line'*. But the NLTK POS tagger labelled *'projecting'* as a noun. This not only creates a 'noisy' invention feature but also fails to identify a function SAO triplet. Again, the likelihood of having such problems will be reduced if a designer is involved.

3. **Incapacity of dealing with inferred relationships in complex sentences.** The approach performs well when analysing explicit relationships such as *'a can end combination comprising a metal can end (10),'* or *'... having an opening opposite to and smaller than said depressible area.'* When it comes to inferred relationships involving multiple features, the accuracy of the approach reduces. For example, sentence *'a can end combination comprising a metal can end (10) and resealable closure (30) coupled to the can end,'* should be interpreted as *'can end combination'* comprises two features, *'metal can end (10)'* and *'resealable closure (30)'*. Only the relationship between *'can end combination'* and *'metal can end'* was picked up by the approach. When a designer is involved this issue would be avoided.

It can be seen that at the current stage, the automated approach cannot replace manual analysis yet but the following solutions can be applied to address the issues stated above and improve the effectiveness of the automated approach:

1. **Use an ontology to create the conceptualisation of features.** An ontology is a formal representation of concepts, data and relationships. It can be applied to create a conceptualised representation of the same feature expressed in various ways, which has already been applied in the patent analysis (Jiang et al., 2018).

2. **Develop and apply more specific POS taggers.** The default NLTK POS tagger is used in this study to enable rapid analysis of patents with a compromised accuracy. Therefore, a more specific, customised POS tagger, trained from patent analysis outcomes would improve the accuracy of POS taggers.

3. **Develop and apply patent text pre-processing techniques.** Pre-processing of patent claims or other textual sections can be carried out to first identify any inferred relationships and then transform them into simplified and more explicitly expressed text descriptions. This should enable the approach to identify invention SAO triplets more accurately.

To investigate the potential effect of implementing the above suggestions, the independent claim for patent US2615610 is re-written manually, with a more consistent naming of invention features and a simpler sentence structure (issue 3). Consistent naming of features was used across the modified claim (issue 1). The same POS tagging technique was applied. The results obtained by applying the automated functional analysis approach can be found in Table 7. Significant improvement of analysis outcomes is achieved with just two invention hierarchy SAO triplets unmatched and all function SAO triplets matched. This results in a 90% effectiveness for invention features, 90% effectiveness function SAO triplets and 60% effectiveness for invention hierarchy SAO triplets. This demonstrates that if the original claims can be transformed into descriptions that are similar to the manual re-write, by implementing the three suggestions above, outcomes of the automated analysis will be almost identical to manual analysis. Therefore, future research activities will be carried out focusing on this matter.

Table 7. Comparison between manual analysis and automated analysis of the modified claim

Original claim	Modified claim (manually re-written)
A container having in a metallic wall section thereof a score line setting off a depressible area bendable inwardly to produce a dispensing opening for the container, a shield disposed over said metallic wall section and having an opening opposite to and smaller than said depressible area, the material of said shield projecting over said score line for shielding the raw cutting edge of the metal resulting from the breaking of said score line when said depressible area is bent inwardly to produce said dispensing opening, and a reclosure element in said shield for reclosing the dispensing opening in said container.	A container having a metallic wall section. The metallic wall section has a score line. The score line setting off a depressible area. The depressible area bendable inwardly to produce a dispensing-opening. A shield disposed over said metallic wall section. The shield having an opening located opposite to said depressible area. The opening is smaller than said depressible area. The said shield is projecting over said score line. The shield for shielding the raw cutting edge of metal. The raw cutting edge of metal resulting from the breaking of said score line. Said depressible area is bent inwardly to produce said dispensing-opening. A reclosure element in shield. The reclosure element for reclosing the dispensing-opening. The dispensing-opening in said container.

Invention Hierarchy SAO triplets (left)

Container	having	Metallic wall section
Container	, and	Shield
Metallic wall section	in….thereof	Score line
Shield	having	Opening
Reclosure element	in	Shield

Invention Function SAO triplets (left)

Score line	setting off	Depressible area
Depressible area	produce	Dispensing opening
Shield	dispose over	Metallic wall section
Opening	opposite to	Depressible area
Opening	smaller than	Depressible area
Shield	projecting over	Score line
Shield	shielding	Raw cutting edge
Score line	result	Raw cutting edge
Reclosure element	reclosing	Dispensing opening

Invention Hierarchy SAO triplets (right)

container	having	metallic wall section
container	having	shield
shield	having	opening

Invention Function SAO triplets (right)

score line	setting off	depressible area
depressible area	to produce	dispensing-opening
shield	disposed over	metallic wall section
opening	located opposite to	depressible area
shield	projecting over	score line
shield	for shielding	raw cutting edge of metal
raw cutting edge of metal	resulting from	breaking of score line
depressible area	to produce	dispensing-opening
reclosure element	for reclosing	dispensing-opening
opening	smaller than	depressible area

6 CONCLUSION

Designers often do not engage with patents well enough in the early design process despite the potential reward patent analysis can bring, for example, enhanced ideation, infringement analysis and patentability check. This is mainly due to the intricate structure and legal terminologies used by patent documents, making them time-consuming to read and understand. Functional analysis has proven to be effective in analysing and representing patented invention working principles by graphically showing the relationships between invention features. In this paper, an automated functional analysis approach for patents is presented and evaluated against previous manually produced examples. The results indicate over 40% effectiveness of the automated approach providing a good basis for further development, especially considering that only standard Natural Language Processing (NLP) toolkits were used. However, the approach produces an almost equal amount of 'noisy' data to the matched results, which is mainly due to the use of homonyms, the same feature with various expressions and inferred relationships that normally can easily be understood by human interpretation. Automatically generated functional diagrams are also achieved to provide design insight. The accuracy of the proposed approach is expected to be improved by using a trained part-of-speech (POS) tagger, improved parsing grammar and an ontology.

REFERENCES

Atherton, M., Jiang, P., Harrison, D. and Malizia, A. (2017), "Design for invention: annotation of functional geometry interaction for representing novel working principles", Research in Engineering Design, Springer London, Vol. 29 No. 2, pp. 245–262.

Aurisicchio, M., Bracewell, R. and Armstrong, G. (2013), "The function analysis diagram: Intended benefits and coexistence with other functional models", Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM, Vol. 27 No. 3, pp. 249–257.

Bank, S. (2020), "Graphviz 0.15", available at: https://github.com/xflr6/graphviz (accessed 16 November 2020).

Bird, S., Klein, E. and Loper, E. (2009), Natural Language Processing with Python, edited by Steele, J., 1st ed., O'Reilly Media Inc.

Cascini, G. and Russo, D. (2007), "Computer-aided analysis of patents and search for TRIZ contradictions", International Journal of Product Development, Vol. 4 No. 1–2, pp. 52–67.

Devoino, I.G., Koshevoy, O.E., Litvin, S.S. and Tsourikov, V. (2005), "Computer based system for imaging and analysing an engineering object system and indicating values of specific design changes", Europe.

Fantoni, G., Apreda, R., Dell'Orletta, F. and Monge, M. (2013), "Automatic extraction of function-behaviour-state information from patents", Advanced Engineering Informatics, Vol. 27 No. 3, pp. 317–334.

Jiang, P., Atherton, M., Harrison, D. and Malizia, A. (2017), "Framework of mechanical design knowledge representations for avoiding patent infringement", Proceedings of the International Conference on Engineering Design, ICED, Vol. 6 No. DS87-6, pp. 81–90.

Jiang, P., Atherton, M., Sorce, S., Harrison, D. and Malizia, A. (2018), "Design for invention: a framework for identifying emerging design–prior art conflict", Journal of Engineering Design, Vol. 29 No. 10, pp. 596–615.

Kim, Y.G., Suh, J.H. and Park, S.C. (2008), "Visualization of patent analysis for emerging technology", Expert Systems with Applications, Vol. 34 No. 3, pp. 1804–1812.

Koh, E.C.Y. (2020), "Read the full patent or just the claims? Mitigating design fixation and design distraction when reviewing patent documents", Design Studies, Elsevier Ltd, No. xxxx, pp. 1–24.

Koster, B. (2015), "Topic 10 : Introduction and Theory of the Patent Claim".

Lee, S.H., Jiang, P. and Childs, P.R. (2013), "Design for functional requirements enabled by a mechanism and machine element taxonomy", Proceedings of the International Conference on Engineering Design, ICED, Vol. 7 DS75-07 No. August, pp. 387–396.

Li, Z., Atherton, M. and Harrison, D. (2014), "Identifying patent conflicts: TRIZ-Led Patent Mapping", World Patent Information, Elsevier Ltd, Vol. 39, pp. 11–23.

Li, Z., Tate, D., Lane, C. and Adams, C. (2012), "A framework for automatic TRIZ level of invention estimation of patents using natural language processing, knowledge-transfer and patent citation metrics", CAD Computer Aided Design, Elsevier Ltd, Vol. 44 No. 10, pp. 987–1010.

Liu, D.-R. and Shih, M.-J. (2011), "Hybrid-Patent Classification Based on Patent-Network Analysis", Journal of the American Society for Information Science and Technology, Vol. 62 No. 2, pp. 246–256.

Michalakoudis, I., Childs, P., Aurisicchio, M. and Harding, J. (2016), "Using functional analysis diagrams to improve product reliability and cost", Advances in Mechanical Engineering, Vol. 9 No. 1, pp. 1–11.

Pahl, G., Beitz, W., Feldhuson, J. and Grote, K.H. (2007), Engineering Design: A Systematic Approach, 3rd ed., Springer-Verlag London Limited, London.

Sorce, S., Malizia, A., Jiang, P., Atherton, M. and Harrison, D. (2018), "A Novel Visual Interface to Foster Innovation in Mechanical Engineering and Protect from Patent Infringement", Journal of Physics: Conference Series, Vol. 1004 No. 1, available at: https://doi.org/10.1088/1742-6596/1004/1/012024.

Trappey, A.J.C., Trappey, C. V., Wu, C.Y. and Lin, C.W. (2012), "A patent quality analysis for innovative technology and product development", Advanced Engineering Informatics, Elsevier Ltd, Vol. 26 No. 1, pp. 26–34.

Tseng, Y.H., Lin, C.J. and Lin, Y.I. (2007), "Text mining techniques for patent analysis", Information Processing and Management, Vol. 43 No. 5, pp. 1216–1247.

Ulrich, K.T. and Eppinger, S.D. (2016), Product Design and Development, 5th ed., Mc Graw Hill India.

WIPO. (2019), World Intellectual Property Indicators 2019, World Intellectual Property Indicators Series, available at: http://www.wipo.int/publications/en/details.jsp?id=4234&plang=EN.

Wodehouse, A., Vasantha, G., Corney, J., Jagadeesan, A. and MacLachlan, R. (2018), "Realising the affective potential of patents: a new model of database interpretation for user-centred design", Journal of Engineering Design, Vol. 29 No. 8–9, pp. 484–511.

Wodehouse, A., Vasantha, G., Corney, J., Maclachlan, R. and Jagadeesan, A. (2017), "The generation of problem-focussed patent clusters: A comparative analysis of crowd intelligence with algorithmic and expert approaches", Design Science, Vol. 3, pp. 1–31.