Extraction of Keywords of Novelties From Patent Claims

Shoko Suzuki
IBM Research - Tokyo
19-21 Nihonbashi, Hakozaki-cho,
Chuo-ku, Tokyo, Japan
e30126@jp.ibm.com

Hiromichi Takatsuka
IBM Japan
19-21 Nihonbashi, Hakozaki-cho,
Chuo-ku, Tokyo, Japan
e50508@jp.ibm.com

Abstract

There are growing needs for patent analysis using Natural Language Processing (NLP)-based approaches. Although NLP-based approaches can extract various information from patents, there are very few approaches proposed to extract those parts what inventors regard as novel or having an inventive step compared to all existing works ever. To extract such parts is difficult even for human annotators except for well-trained experts. This causes many difficulties in analyzing patents. We propose a novel approach to automatically extract such keywords that relate to novelties or inventive steps from patent claims using the structure of the claims. In addition, we also propose a new framework of evaluating our approach. The experiments show that our approach outperforms the existing keyword extraction methods significantly in many technical fields.

1 Introduction

Recently there are growing needs for analyzing patents. Many companies want to analyze large amount of patents for various purposes like patent retrieval or analyzing technical trends, etc. For searching and analyzing large amount of patents, NLP-based approaches are adequate, and many approaches are developed (for example (Abbas et al., 2014)). Several keyword extraction methods are proposed in the context of patent retrieval or information extraction from patents. Most of them use traditional unsupervised approaches like BM25 (Robertson and Zaragoza, 2009) or supervised approaches like CRF (Lafferty et al., 2001). While BM25 tends to extract keywords that are characteristic to each patent, CRF is applied to extract various kinds of Named Entities such as technologies, effects, and attributes. (for example in (NTCIR, 2010),(Nishiyama et al., 2010)).

However, considering the original purpose of submitting patents, patents must contain rich information not limited to the above examples. Especially, every patent must contain what the inventors think as novel or having an inventive step compared to the all existing works ever. There is no doubt that extracting such keywords is quite important and applicable to all other patent analysis like patent retrieval or analyzing technical trend, etc. To our knowledge there are very few works that explicitly try to extract those keywords that relate to the novelties or the inventive steps of each patent (we will call these keywords as keywords of novelties in this paper).

In general patent retrieval task, various kind of weights are calculated for keywords/keyphrases. But these weights don't necessarily reflect the degree of the novelties. Several approaches seem to extract novel parts of each patent implicitly, but they don't go so far as to extract keywords of novelties. Besides, in patent retrieval, similar patents often have different surface expressions especially on the novel parts, resulting in the situation that the performance in patent retrieval task is not necessarily related to the performance of extracting the novel part. This means that the extracting keywords of novelties cannot be evaluated directly in patent retrieval task.

In this paper, we propose a new approach of extracting keywords of novelties from patent claims. Among various parts in patent applications, patent claims are the most crucial parts that define the scope

of protection and contain all of the important parts of the invention. Since patent claims are written in a specific manner, usual NLP approaches may fail to extract useful information. Therefore, we assume several underlying structural rules in patent claims and utilize these assumptions in extracting keywords.

In addition, we also propose a new framework of evaluating our approach.

In section 2, we first explain patent claim structures and related work to extract those structures. Then we briefly introduce some approaches of keyword extraction from patents. In section 3 we propose a novel approach to extract keywords of novelties. And in section 4, we introduce a new framework of evaluating the performance of our approach. Section 5 shows the experimental setting and the evaluation results of our proposed approach compared to other keyword extraction approaches. Section 6 describes some concluding remarks and future application using our approach.

2   Related Work

Several attempts are made to implicitly extract keywords of novelties for each patent, mainly in the context of patent retrieval. Patent retrieval is a task to extract similar patents when a target patent is given.(Patent retrieval includes those concepts such as invalidity search and prior art search.) Among naive approaches that use common techniques in information retrieval (IR), some utilize the features that are specific to patent claims, which result in better performance than using usual IR techniques.

Although each patent claim is a plain text without apparent sections, it is built under some rules. There is no doubt that utilizing these rules achieves better performance in extracting information from patent claims.

2.1   Patent Claim Structures

Patent claims have specific structures, and it is difficult to understand the contents except for experts. Each patent has one or more claims. In case a patent has multiple claims, these claims are divided into two kinds of claims; independent claims and dependent claims. Each claim can be decomposed to several structural elements. There are also several types for claim structure like Jepson type or Markush type. For example, Jepson type has two parts in a claim, preamble and body. In addition, each claim must declare a subject matter, which corresponds to a noun phrase in a specific position of the claim.

To extract those structures automatically, several approaches exist. For example, Sheremetyeva et al. have decomposed each claim to structural elements using POS-tags (Sheremetyeva et al., 1996), Parapatics et al. have categorized claims into several types using cue phrases (Parapatics and Dittenbach, 2009), and Shinmori et al. have proposed to apply Rhetorical Structure Theory (RST) for parsing structural elements (Shinmori et al., 2003).

2.2   Keyword Extraction from Patent Documents

Keyword extraction is an important task for every kind of documents, and many approaches are suggested. The most known approaches are tf-idf, BM25 and TextRank. These approaches are also applicable to patent claims, but these are not designed to extract keywords of novelties, and tend to achieve low performance in various patent analysis.

For example, TextRank approach is applied (Verma and Varma, 2011) to the whole patent application or to a specific part of a patent like patent claim, abstract, or detailed description in order to extract important keywords for invalidity search. But they did not intend to extract such keywords of novelties. Besides, they did not utilize claim structure.

Word age of each term is introduced to measure the degree of the novelties for each patent (Hasan et al., 2009) in order to score novelty of each patent. Word age represents how long the term exists in a corpus (time span from the time when it first appeared to the time when the target patent submitted), and it represents novelty to some extent. But it is obvious that word age highly depends on the corpus, technical field or surface expressions, which is not related to what inventors think as novel part. Moreover, inventors don't necessarily use new keywords to describe the novel part of of their inventions.

Other than those approaches that directly extract important keywords from patent claims, there are some approaches that patent-specific keyword extraction are used.

Shinmori et al.(2003) proposed an approach to extract such keywords/keyphrases that are representative in each structural element in patent claims by using some morphological patterns after decomposation of claims to structural elements(Shinmori et al., 2003).

Takaki et al.(2004) first decomposed each claim into structural elements, and calculated the importance of each element by a measure of how each term in the element is locally distributed in those elements (Takaki et al., 2004). Their approach seems to grasp one side of patent claim structure, and their approach is effective in patent retrieval task to some extent.

Lin et al.(2010) tried more systematic approach utilizing patent claim structure for patent retrieval (Lin et al., 2010). They built a claim tree by extracting the relations between independent/dependent claims and the structural elements for all patents, then search the similarities of the claim trees. They extracted relevant keywords of each structural element from patent specifications (not from claims) using term frequency in the specifications and calculating mutual information of terms selected from claim and specification.

These patent-specific methods extract keywords from various aspects, but none of them directly tried to extract keywords of novelties for each patent.

There are other approaches to extract keywords from patent claims using claim structures, but the focuses are only limited to discriminations like preambles/bodies (Mase et al., 2005) or the subject matters.

3 Extraction of Keywords Related to Novelty from Patent Claims

Compared to the related work, our approach is built under some assumptions of claim structures. Using these assumptions we propose a new approach to extract keywords of novelties from patent claims.

3.1 Assumptions of Claim Structures

One main reason that a patent is constructed from multiple claims is that an applicant want to protect the scope as broad as possible after registration by a patent office. Patent offices in many countries examine patent claims, and if one idea already exists in prior art or easily imaginable from prior art, they refuse the claim. However, if there is a claim that describes a bit smaller scope of the rejected claim, and the scope is invalidated by any prior art, this claim is granted. Therefore, applicants tend to submit multiple claims in one patent with some hierarchical structures of dependencies. In general, a dependent claim describes smaller scope than that of the "parent" claim. This means that a dependent claim focuses on an important part of the parent claim to secure this part. In many cases, this important part (that the applicant wants wider scope as possible) describes the novelties or the inventive steps that the inventor wants to claim, or describes the area the patent is applicable.

Besides, each claim usually has multiple structural elements. Since a claim must contain every element that is necessary, part of the elements are necessary to construct the subject matter but not directly related to the novelties or the inventive steps of the patent. This also suggests that elements are related to each other somehow. These relations might represent some process flows or adding functions/features to other elements. Then we can interpret these relations as some kind of hierarchies. Therefore, just like dependent claim structures, we can rebuild structural elements in a claim into some hierarchical structures.

Figure 1 shows an example of hierarchical structures. The element 1-4 is in the claim 1, and the dependent claim 2 and claim 3 depends on the element 2 and 4 respectively.

Now there is another rule in writing a claim; there must be no unnecessary element appeared in a claim. This is because an unnecessary element narrows the scope of the patent, and applicants never want this situation. Suppose that a claim has an additional new element which depends on an key element containing novelties of the patent (this new element might use the output of the key element's process or add some features to the key element), then the claim has only a limited scope compared to the claim without the additional element. This means that a properly written patent claims contain only a few key elements and these key elements should be placed in a lower level of the hierarchical structure.

Based on these observations, we assume two things;

- the keywords of novelties tend to exist in the elements that the dependent claims depend on.
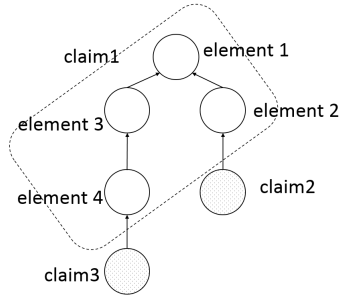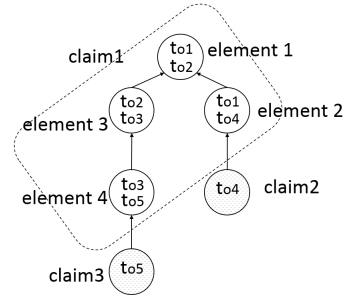
Figure 1: Example of Claim Structure

Figure 2: with Overlapping terms

- the keywords of novelties tend to exist in a lower level of the element hierarchical structure.

3.2 Outline of New Approach to Extract Keywords

In this subsection, we describe the outline of our proposed approach to extract keywords of novelties in the first claim using patent claim structures under the assumptions in the previous section. (The extension from the first claim to all independent claims is straightforward.) The notation is listed in Table 1.

Notation	Description
$e(w)$	the first element in $\{e_i\}$ where term w appears
$d(e_i)$	the depth of the element e_i
t_o	overlapping term that connects element-element dependency or element-dependent claim dependency
$ET(e_i)$	map from child element e_i to t_o
$DT(c_i)$	map from dependent claim c_i to t_o
$parent(e_i)$	the parent element of e_i
T_o	the set of all t_o
$ncl(t_o)$	the number of dependent claims $DT^{-1}(t_o)$
$Mod(k)$	map to $\{t_o\}$ that k modifies

Table 1: Notation

Algorithm 1 extracting element dependency structure

$d(e_1) \leftarrow 0$
for $i = 2$ **to** $i = |\{e_i\}|$ **do**
    search $e(w)$ in $\{e_j | j \leq i\}$ for $\forall w \in e_i$
    $d(e_i) \leftarrow -1$
    $d(e_i) \leftarrow \max_{w \in e_i} d(e(w)) + 1$
    if $d(e_i) \neq 0$ **then**
        $t_o \leftarrow \arg\max_{w \in e_i} d(e(w))$
        $parent(e_i) \leftarrow e(t_o)$
        $ET(e_i) \leftarrow t_o$
    end if
end for

Step1:extracting element structure: The first claim of a patent is parsed so that each structural element is decomposed, and dependencies within the elements are extracted. Dependencies between the elements are extracted in the following way. First, the decomposition of each element is done by using cue phrases (Shinmori et al., 2003) or using line breaks. Then, the depth for each element is calculated by following the procedure in Algorithm 1. Note that $\{e_i\}$ is a set of elements sequentially derived from the first claim. Those terms that appear in the subject matter are removed from this analysis. Figure 2 represents the example of dependencies using overlapping terms t_o. The term t_{o3} first appears in e_3, then in e_4. Therefore $parent(e_4) = e_3$, and $ET(e_4) = t_{o3}$. Other types of structure extraction is also applicable.

Step2:extracting claim dependency structure: The claims that depend on the first claim are parsed. The procedure to attach depth for all dependent claims is the same as that of step 1. This means that overlapping term t_o is attached for each dependent claim c_i, i.e. $DT(c_i) \leftarrow t_o$. In figure 2, t_{o5} first appeared in e_4. t_{o5} also appears in the claim 3, then $DT(c_3) = t_{o5}$.

Step3:calculating the score of the representative keywords: For every overlapping term t_o stored in the set T_o, the score s_o is calculated using the depth of element $e(t_o)$ and the number of dependent claims $ncl(t_o)$ attached to t_o. The definition of score s_o is explained in the next subsection 3.3.

Step4:calculating the score of the keywords of novelties: Keywords k modifying each t_o are searched in the first claim. Modification of t_o by k is defined by satisfing at least one of the following conditions;
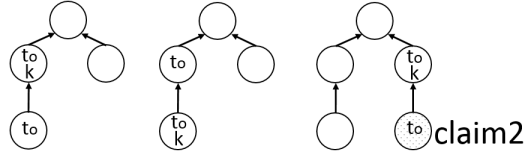
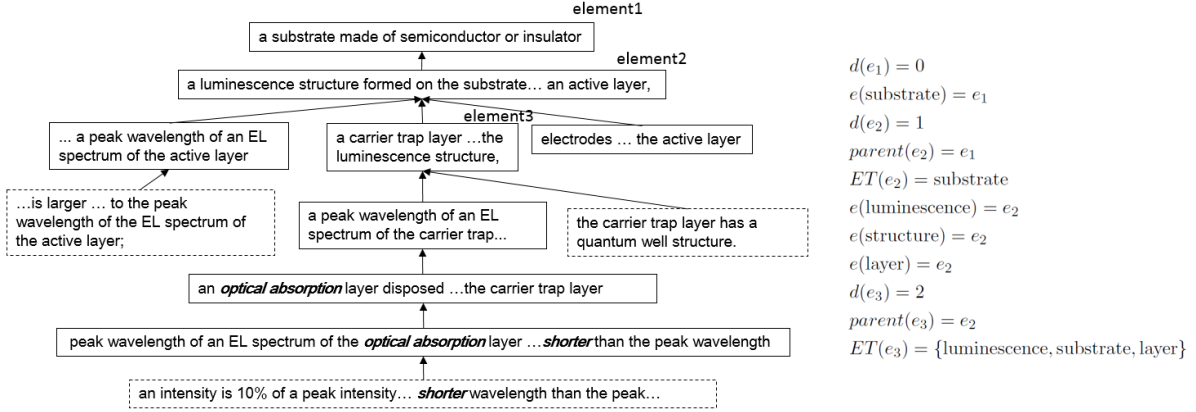Figure 3: Examples of Modification: k modifying t_o

Figure 4: Actual Example of Claim Structure

1. k appears in the element $ET^{-1}(t_o)$.

2. k appears in the element $e(t_o)$.

t_o can be also regarded as k. A keyword k modifying t_o is represented as $t_o \in Mod(k)$ and $k \in Mod^{-1}(t_o)$. Note that the mapping function Mod and Mod^{-1} is many-to-many mapping. Figure 3 shows examples of modifications. In every example in figure 3, k modifies t_o. Then the score $S(k)$ for each $k \in \{k| \cup_{t_o \in T_o} Mod^{-1}(t_o)\}$ is calculated using the definition in the next subsection 3.3.

The actual example of claim structures is shown in figure 4. The boxes enclosed by solid line represent elements in the first claim and those of dashed line represent dependent claims. The bold italic keywords are the keywords of novelties annotated automatically in the framework introduced in the section 4. The right side of figure 4 shows the actual procedure of Algorithm 1 for the first 3 elements.

3.3 Definition of Scores

During the whole process, the locality score for each term w is calculated by

$$loc(w) = \frac{|\{e_i\}|}{|\{e_i|e_i \ni w\}|}, \tag{1}$$

that is, locality represents how much w is localized among all elements.

The score for overlapping terms is defined as

$$s_{o1}(t_o) = d(e(t_o))*(ncl(t_o) + 1) \tag{2}$$
$$s_{o2}(t_o) = d(e(t_o))*(log(ncl(t_o)+1) + 1), \tag{3}$$

and using the score of overlapping terms t_o, the score of k that modifies t_o is defined in the equation

$$S_l(k) = \max_{t_o \in Mod(k)} loc(k) * s_{ol}(t_o) \tag{4}$$

with $l = 1, 2$ and s_{ol} represents eq.(2) or eq.(3), respectively. Note that $S_l(k) = 0$ if $Mod(k) = \emptyset$.

4 Evaluation Method Using Rejected and Granted Patents

In the previous section, we proposed a new approach of how to calculate the scores for each term k. This score $S_l(k)$ represents the degree of novelties of term k in each patent. But the evaluation of this approach is another difficult task. To evaluate the performance of extracting such keywords, we need annotated data. But manual annotation needs domain knowledge of each technical field and expertise in reading patent claims. Therefore, preparing large amount of manual annotations for various technical fields is quite difficult. Since extracting the keywords of novelties is not a conventional task in patent analysis, there is no commonly available shared task set for evaluation. In this section, we propose a general framework to obtain annotated data automatically from the examination result of Patent Offices.

4.1 Process of Patent Examination

Patent Offices in some countries publish all examination processes of each patent. Basically, a submitted patent follows a process like this; **1**. A patent is examined and if an insufficient part exists, the examiner reject the patent. **2**. The applicant may withdraw submission, or modify the claim to overcome the reason of the rejection. **3**. After repeated examination and possible rejection/modification, a patent is decided to be granted or rejected.

One of the major and critical reasons of the rejection is the existence of prior art that invalidates the submitted patent. In other words, if this type of rejection is overcome by modifications of the claims, this modified part must contain the keywords of novelties.

4.2 Framework of Evaluation

Based on the assumption, we build a framework to evaluate the extraction of keywords of novelties for each patent. Each patent document used in this framework must be rejected only by the reason of existence of prior art and then be granted after modifications. For each patent satisfying the above condition, two types of claims are extracted for each patent; the claims before the patent is examined and the claims after the patent is granted. Then those keywords, which appear in the granted first claim but not in the rejected first claim, are extracted. Such keywords are regarded as the positive-labeled set of keywords of novelties. The reason of extracting keywords only from the first claim is, the correspondence between the two types is clear. (For example, the third claim in the rejected patent corresponds to the second claim in the granted patent, which is difficult to guess. But usually, the first claim of the rejected patent is corrected by adding keywords from the proceeding dependent claims or from the description in the application.)

In order to examine the extracted set of keywords are truly positive, we randomly pickup and check some of the granted first claims manually by referencing the arguments in response to the notices of reasons for rejection (i.e. Office Actions). In the arguments, applicants explain how the corrected claims differ from the prior art. In this preliminary analysis, the average f-measure of 26 patents becomes 0.8339. This result ensures that our proposed framework is adequate for preparing positive-labeled set of keywords of novelties at least as an approximation.

The positive-labeled set is compared to those keywords extracted by various approaches for evaluation.

5 Experiments

In this section, we evaluate our approach proposed in section 3 using the framework proposed in section 4.

5.1 Corpus

The Japanese patents submitted during Jun 1 to March 31 in 2005 are used in the evaluation. We selected patents for each technical field corresponding to International Patent Classification (IPC) from A to H. The IPC Section title and the definition list is shown in table 2 (More detailed explanation can be found in (WIPO Guide, 2016)). For calculating document frequency in BM25 (one of the baseline approaches we applied), we use other corpus containing the patent data submitted before Jun 1, 2005. This corpus contains around 3 million patents.

Section	Definition
A	HUMAN NECESSITIES
B	PERFORMING OPERATIONS; TRANSPORTIN
C	CHEMISTRY; METALLURG
D	TEXTILES; PAPER
E	FIXED CONSTRUCTIONS
F	MECHANICAL ENGINEERING; LIGHTING; HEATING; WEAPONS; BLASTING
G	PHYSICS
H	ELECTRICITY

Table 2: IPC Section Title

Approach	Definition
Proposed1	$S_1(k)$ in eq.(4)
Proposed2	$S_2(k)$ in eq.(4)
BM25	BM25 score $BM(w)$
BM25perEle	$S_{base2}(w)$ in eq.(7)
Locality	$loc(w)$ in eq.(1)
Loc*Ele1	$\sum_{e_i \ni w} loc(w) * \tilde{S}_{e1}(e_i)$
Loc*Ele2	$\sum_{e_i \ni w} loc(w) * \tilde{S}_{e2}(e_i)$

Table 3: Definition of Approaches

5.2 Approaches to be Evaluated

As a baseline, we try BM25, the traditional keyword extraction approach. Besides, we also try several naive approaches that are easily conceivable from the related work in section 2. One is using locality as a score defined in eq.(1). This is similar to Takaki(2004)'s approach while they also used element-wise score as well. Here we prepare two types of element-wise score referring to Takaki's approach.

$$\tilde{S}_{e1}(e_i) = \frac{1}{|e_i|} \sum_{w \in e_i} loc(w) \tag{5}$$

$$\tilde{S}_{e2}(e_i) = \frac{1}{log(|e_i| + 1)} \sum_{w \in e_i} loc(w) \tag{6}$$

Moreover we prepare one simple extension of BM25:

$$S_{base2}(w) = \max_{e_i \ni w} \tilde{S}_{e2}(e_i) * BM_i(w) \tag{7}$$

where $BM_i(w)$ represents the value of BM25 of w regarding the element e_i as a document.

The table 3 shows the definition of scores of tested approaches. **Proposed1** and **Proposed2** are our new proposed approaches introduced in section 3. **BM25** is the baseline approach. **BM25perEle**, **Loc*Ele1** and **Loc*Ele2** are the approaches easily conceivable from the related work.

5.3 Evaluation Results

We evaluate the performance of keyword extraction using each score by Mean Average Precision (MAP) which is often used to evaluate the performance of information retrieval. Since several types of scores like our proposed approaches or locality tend to have the same value for multiple terms, we calculate all the orders for those tie ranks and averaged the MAP value of each order.

The result is in the table 4. This shows that in every technical field our proposed approaches **Proposed1**, **Proposed2** outperform the baselines using BM25 or those approaches that are easily conceivable from the relate work. Especially the approach **Proposed2** significantly outperforms the baselines in most of the tehcnical fields except for IPC=C (chemistry). One reason for relatively low performance of our proposed approach in the field of chemistry is, that patent claims in the field are often the type of Markush Claim which doesn't fit our current structure parsing method in step1 of subsection 3.2 (The current method fits Jepson Claim better).

6 Conclusion

We propose a new approach to extract keywords of novelties from patent claims. It is a challenging task partly because the problem setting itself is rather unique, and partly because there has been no framework to evaluate the performance directly. We show that the existing keyword extraction techniques don't work well when analyzing patent claims, since patent claims have a specific format that a non-expert finds difficult. Although analyzing patent claims is quite important in many industries, understanding the contents of claims using NLP techniques is still in a developing phase.

Approach	MAP							
	IPC=A	IPC=B	IPC=C	IPC=D	IPC=E	IPC=F	IPC=G	IPC=H
Proposed1	0.5947	0.5526	0.4874	0.5338	0.5685	0.5783	0.5825	0.5520
Proposed2	**0.6109**	**0.5741**	**0.4925**	**0.5471**	**0.5880**	**0.6066**	**0.5984**	**0.5779**
BM25	0.3751	0.3329	0.3855	0.3464	0.3373	0.3033	0.3394	0.3333
BM25perEle	0.4106	0.3689	0.4042	0.3709	0.3579	0.3355	0.3825	0.3708
Locality	0.4997	0.4579	0.4475	0.4458	0.4534	0.4390	0.4847	0.4613
Loc*Ele1	0.5155	0.4736	0.4753	0.4739	0.4685	0.4597	0.5082	0.4820
Loc*Ele2	0.5376	0.4925	0.4578	0.4635	0.4795	0.4806	0.5239	0.4943
number of documents	356	501	278	198	370	351	338	501

Table 4: MAP result

While there are works on extracting novelties from trend analysis of news, SNS or other documents, these are mainly aiming to extract only keywords from those documents with brand new contents. But in patent claims, novelties are often represented in general expressions. This is one main reason that we focus on extracting what inventors think as novel for every patent using the structures of patent claims.

In this paper we apply only preliminary approaches, but additional attempts such as using key phrases, using dependencies between terms, or using other fields in patent specifications will surely increase the performance. Moreover, since annotated data is available using our proposed framework, supervised approaches are applicable. Combining features like eq.(4) with other features derived from patent claim structures, supervised approaches may give us some knowledges on what kind of structure may be effective for extracting keywords of novelties.

A simple expansion of the method in step 1 of subsection 3.2 to Markush Claim will also increase the performance in the field of chemistry.

Future work also includes applying our approach to patent retrieval, patent summarization, or detecting important technologies.

References

Akihiro Shinmori, Manabu Okumura, Yuzo Marukawa, and Makoto Iwayama. 2003. *Patent Claim Processing for Readability: Structure Analysis and Term Explanation.* Proceedings of the ACL-2003 Workshop on Patent Corpus Processing, 20:56-65.

Assad Abbas, Limin Zhang, and Samee U. Khan. 2014. *A literature review on the state-of-the-art in patent analysis.* World Patent Information, 37:3-13.

Fu-ren Lin and Feng-mei Huang. 2010. *The Study of Patent Prior Art Retrieval Using Claim Structure and Link Analysis.* Pacific Asia Conference on Information Systems:9-12.

Hidetsugu Nanba, Atsushi Fujii, Makoto Iwayama, and Taiichi Hashimoto. 2010. *Overview of the Patent Mining Task at the NTCIR-8 Workshop.* NTCIR-8 Workshop Meeting.

Hisao Mase, Tadataka Matsubayashi, Yuichi Ogawa, Makoto Iwayama, and Tadaaki Oshio. 2005. *Proposal of Two-stage Patent Retrieval Method Considering the Claim Structure.* ACM Transactions on Asian Language Information Processing, 4(2):190-206.

John D. Lafferty, Andrew McCallum and Fernando C. N. Pereira. 2001. *Conditional random fields: Probabilistic models for segmenting and labeling sequence data.* Proceedings of the 18th International Conference on Machine Learning:282-289.

Manisha Verma and Vasudeva Varma. 2011. *Applying Key Phrase Extraction to Aid Invalidity Search.* Proceedings of the 13th International Conference on Artificial Intelligence and Law:249-255.

Mohammad Al Hasan, W. Scott Spangler, Thomas Griffin, and Alfredo Alba. 2009. *COA: Finding Novel Patents Through Text Analysis.* Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining:1175-1184.

Peter Parapatics and Michael Dittenbach. 2009. *Patent Claim Decomposition for Improved Information Extraction.* Proceedings of the 2nd International Workshop on Patent Information Retrieval:33-36.

Risa Nishiyama, Yuta Tsuboi, Yuya Unno and Hironori Takeuchi. 2010. *Feature-Rich Information Extraction for the Technical Trend-Map Creation*. Proceedings of NTCIR Workshop 8 Meeting.

Stephen Robertson and Hugo Zaragoza. 2009. *The Probabilistic Relevance Framework: BM25 and Beyond*. Foundations and Trends in Information Retrieval, 3(4):333-389.

Svetlana Sheremetyeva, Sergei Nirenburg, and Irene Nirenburg. 1996. *Generating patent claims from interactive input*. Proceedings of the 8th. International Workshop on Natural Language Generation:61-70.

Toru Takaki, Atsushi Fujii, and Tetsuya Ishikawa. 2004. *Associative Document Retrieval by Query Subtopic Analysis and Its Application to Invalidity Patent Search*. Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management:399-405.

World Intellectual Property Organization. *INTERNATIONAL PATENT CLASSIFICATION (Version 2016) GUIDE*. `http://www.wipo.int/export/sites/www/classifications/ipc/en/guide/guide_ipc.pdf`.