ISSN:1738-6764, eISSN:2093-7504

# Function-based Patent Landscape Analysis for Diabetes Technologies: Insights from the U.S. and China

**Zhuo Wu [1] and Jarinya Limpanadusadee [2,*]**

[1] Department of Information System, Hanyang University; Master Student; takooo@hanyang.ac.kr
[2] Department of Information System, Hanyang University; Phd Candidate; jarinyalim@hanyang.ac.kr
* Correspondence

https://doi.org/10.5392/IJoC.2025.21.4.116

Manuscript Received 11 July 2025; Received 3 December 2025; Accepted 9 December 2025

***Abstract:*** *Current cross-country patent structural approaches show that comparative studies utilizing SAO (subject–action–object) structures are limited in specific technical domains. This paper analyzes the functional types of expressions in diabetes-related patent texts from the United States and China by examining verb usage and the semantic categories of subject and object phrases. It proposes a method to identify and classify patent functions based on SAO structures, aiming to uncover linguistic preferences and technological characteristics in functional expression across the two countries. The findings reveal significant differences between U.S. and Chinese diabetes patents regarding functional type, structural composition, and semantic expression style, indicating tendencies toward system integration-oriented and pharmacological mechanism-oriented expressions, respectively. Additionally, the analysis of time evolution and field distribution highlights distinct technical rhythms and layout characteristics in each country. The framework introduced in this paper offers methodological support and an empirical foundation for patent semantic modeling and cross-country technology comparison research.*

**Keywords:** Patent Analysis; SAO (Subject-action-object); Functional Structure; Functional Trend; Natural Language Processing; Semantic Modeling

## 1. Introduction

With the continuous growth of diabetes-related medical technologies, global patent applications in this field have increased significantly. Patents serve as a key carrier of innovation records, technologies function and reveal differences in national R&D strategies. Thus, structured identification and semantic modeling of functional expressions have become essential for analyzing technology trends and guiding innovation strategies.

In recent years, Natural Language Processing (NLP) technology has been applied to patent mining research. For example, Turchin [1] analyzed electronic health records (EHR) based on NLP, which improved the efficiency of the formulation of treatment strategies. Another study applied NLP techniques for topic modeling and both unsupervised and semi-supervised approaches to design AI tools that assist clinicians in providing personalized care to diabetic patients [2]. In addition, some researchers [3] identified the R&D hotspots and potential application scenarios of diabetes management technology by mining patents related to global continuous glucose monitoring (CGM).

Although the above studies have demonstrated the multifaceted value of NLP in extracting semantic information from patent texts, most of them rely on shallow methods such as keyword frequency, topic modeling, or semantic matching, which struggle to capture the complex and structured functional expression logic in patent documents. To address this limitation, many recent studies have adopted the SAO (Subject-Action-Object) structures [4] as a core framework for patent semantic analysis.

Moreover, utilizing patents from the literature data, SAO semantic structures have been used to extract elements and construct corresponding semantic networks to identify products' technologies [5]. On this basis, some work has proposed functional importance metrics, such as Function Score [6], to quantify the influence of a function in a specific period or technical domain, providing a basis for high-value technology identification and strategic patent planning.

However, the existing SAO-based patent function analysis method focuses mainly on engineering technology, a systematic structured semantic modeling framework has not been established for medical technology, especially for diabetes patents. Meanwhile, existing studies lack a research framework for systematically comparing the differences between semantic expression styles and technology strategies from a cross-country perspective. Therefore, to address the limitation, this paper proposes a set of extraction and classification methods based on the SAO structures of diabetes patents to reveal the differences and characteristics of the logic and linguistic style of technological function expression in different countries.

Currently, China has the largest number of diabetes patients in the world, the prevention and treatment of diabetes has become a key direction of the national public health system. The number of Chinese patent applications in this field has also continued to grow, especially in the areas of traditional Chinese medicine compounding and synthetic drugs, showing a diversified development trend, reflecting the characteristics and advantages of China in the field of traditional medicine) [7]. Besides, the U.S. has long been a global leader in diabetes monitoring technology, smart devices and innovative drugs. The patent layout is highly focused in the core areas of continuous glucose monitoring (CGM) technology [3] and new insulin delivery systems [8]. The two countries show significant differences in terms of patent applications, technology path selection and language expression preferences, so a systematic approach can be constructed to analyze them comparatively at the semantic level.

Based on this, this paper selects the diabetes patents of China and the United States as research objects in the empirical study and systematically compares their functional language expression features and functional type evolution trends. The results are expected to provide a semantic modeling framework and methodological support for the study of technology trends in multinational patents and the formulation of medical technology strategies.

The remainder of this paper is organized as follows. Section 2 reviews the literature on semantic analysis of patent texts and SAO-based methodologies. Section 3 presents the methodological framework, including the data sources, retrieval strategy, SAO extraction process, and functional classification of verbs and key phrases. Section 4 reports the empirical findings of the cross-national comparison of diabetes patents from China and the United States. Section 5 discusses the implications of these findings, summarizes the main conclusions, and outlines the study’s limitations and directions for future research.

## 2. Literature Review

### 2.1 SAO Structure Extraction Methods

Current SAO extraction methods can be broadly categorized into rule-based linguistic approaches and deep learning-based embedding approaches.

First, the rule-based approaches typically rely on syntactic analysis tools such as Stanford Parser [9], which analyzes sentence dependencies to identify subject, verb, and object constituents. In addition, tools such as Alchemy API, Antelope, and Knowledgist [10] have been applied to extract SAO structures from scientific and technical literature and patent texts. More recently, Python NLP libraries such as spaCy have been widely used for structured processing of patent corpora due to their lightweight design, high efficiency, and robust syntactic analysis capabilities, making them particularly suitable for large-scale technical document processing. These libraries typically incorporate part-of-speech (POS) tagging [11] and noun chunking [12], which are essential to identify SAO structures.

The second category includes semantic modeling approaches. Deep learning-based models such as SciBERT [13], a pre-trained language model trained on biomedical and scientific corpora, provide strong semantic modeling capabilities for domain-specific terms. Some studies have proposed hybrid frameworks that incorporate predefined rules and deep learning semantic and syntactic parsing to balance accuracy and efficiency [14].

### 2.2 Semantic Similarity and Functional Metrics

SAO-based semantic similarity analysis has demonstrated strong potential in patent analysis, technology assessment, and information retrieval. Early work applied the Word2Vec model [15] to represent keywords in

SAO triples and calculate word-level semantic similarity, providing semantic metrics for patent comparison. Subsequent patent similarity studies further combined Skip-gram–based word embeddings with additional signals such as image features to improve semantic generalization and retrieval performance, especially for patents that combine complex text and imagery [16].

Other researchers utilized knowledge graphs and ontologies such as WordNet [17] to calculate semantic distances based on hierarchical structures, supporting more refined conceptual generalization and synonym replacement [18]. More recent efforts have proposed hybrid models combining SAO embeddings with pre-trained patent vector models, yielding strong performance in infringement detection, patent litigation support, and semantic retrieval [19]. Enhanced similarity measures, such as those based on the extended Sørensen-Dice coefficient, have further strengthened the discriminatory power between SAO structures [20].

In recent years, some studies have further embedded SAO structures into vector spaces and combined them with advanced mapping and topic modeling techniques. For instance, the SAO2Vec algorithm embeds SAO structures as document vectors and shows superior performance compared with Doc2Vec and simple SAO frequency counts in patent grouping and similarity analysis [21]. Other studies have proposed integrating SAO representations with generative topographic mapping (GTM) for technology opportunity discovery based on patent analysis [22]. In addition, a topic modeling framework based on SAO structures has been developed to automatically annotate technological topics within patent corpora [23]. Furthermore, Ko-SAO extends SAO-based functional analysis to Korean energy patents, demonstrating the adaptability of SAO across different languages and domain-specific functional analyses [24]. These recent studies indicate that SAO is receiving increasing attention as a functional representation in patent analysis and provide the methodological context for adopting the SAO framework in this study.

Beyond SAO-based analysis, common methodologies in transnational patent research include metrics based on simple counts or IPC classifications, citation and co-classification networks, and keyword co-occurrence or topic-modeling approaches [16], [25].While these methods are effective for characterizing which technological domains and thematic clusters are prominent in each country, they do not explicitly represent subject–action–object relationships and therefore offer limited insight into how technological entities and processes are functionally linked within patent texts. Given the focus of this study, we adopt SAO structures as the core analytical unit, enabling comparisons between U.S. and Chinese diabetes patents at the level of functional pathways and expression habits—for example, process-oriented therapeutic mechanisms versus compositional device/system structures—rather than only at the level of technological domains.

### 2.3 Sub-TD Classification in Patent Analysis

To more precisely portray the evolution of technical functions and content, many studies use the sub-Technology Domain (sub-TD) as the core unit of patent analysis. The International Patent Classification (IPC) system is one of the most authoritative and widely adopted classification systems, commonly used to divide patent data based on main or multi-level IPC codes [25]. Data-driven methods such as patent citation networks and topic modeling are also used to identify latent sub-TD clusters [26]. Some researchers combine IPC classifications with industry standards or product categories to construct multidimensional classification schemes for industry chain analysis and technology iteration tracking [27].

Sub-TD classification enhances the precision of patent function analysis and improves the explanatory power of technology trend assessments, thereby supporting informed decisions in technology management and industrial strategy.

## 3. Method

This section presents the methodological framework of the study, which integrates syntactic parsing with semantic functional categorization for patent texts. The framework consists of four main steps. First, extract SVO (Subject-Verb-Object) structures to obtain functional grammatical units. Second, cluster the subject and object phrases into specific classes. Third, semantically categorize the verbs to establish the functional type mapping and assign a functional label to each pair of VOs, thus forming an SAO (subject-action-object) structure labeled with functional types. Finally, structured modeling and type determination of functional expressions in patent text is realized. Figure 1 shows the framework.

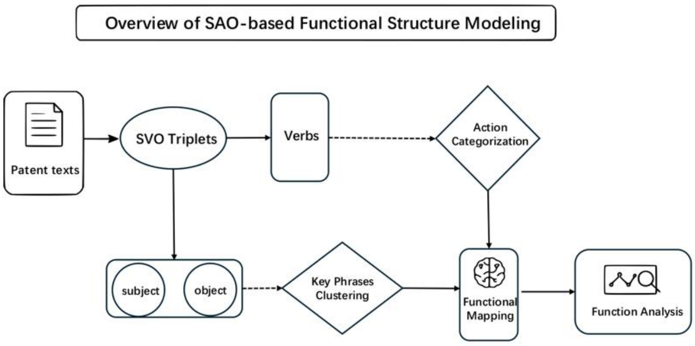



**Figure 1.** SAO-based functional expression analysis framework

3.1 Data collection

In this study, the WIPSON Patent Database was chosen as the data source platform (www.wipson.com). WIPSON is a professional patent database that integrates standardized records from major patent offices, providing unified bibliographic fields, IPC/CPC classification systems, and searchable titles, abstracts, and claims. This enables us to search patents under consistent conditions and to export large-scale text data for analysis. From WIPSON, we extracted the patent number, IPC classification, country, publication year, and English abstract for each patent. These fields constitute the dataset used for subsequent SVO extraction and longitudinal analysis. For the retrieval strategy, patent sets are retrieved using a joint query of topics (Title, Abstract, Claims) and a Boolean logic expansion search query [28].

3.2 Extraction of Subject-Verb-Object triplets

3.2.1 Preprocessing

This study begins by applying part-of-speech (POS) tagging [11] to perform a basic syntactic analysis. POS tagging assigns each word a grammatical category based on its contextual usage, enabling the identification of core syntactic components such as subjects, verbs, and objects.

3.2.2 Construction of SVO candidate structures

To extract functional expressions that are semantically complete and syntactically coherent, this study implements a rule-based method to identify Subject–Verb–Object (SVO) structures from patent texts. Based on part-of-speech (POS) tagging and dependency parsing, the method aims to isolate linguistic patterns that correspond to core technical functions. Since verbs in patent texts usually denote system behaviors, SVO structures built around verbs can effectively capture the core semantic content of functional languages. Therefore, in this step, the verb is regarded as the core unit of extraction.

First, verbs that are the core predicates in a sentence are identified, and lexical items with ROOT or conj dependencies are prioritized as candidate central predicates. Next, subjects (nsubj, nsubjpass) and objects (dobj, obj, attr, and prepositional objects obtained via the prep-to-pobj path) are extracted from the dependent child nodes of this verb. In addition, supplementary consideration is given to subjects in control structures (xsubj) and to doer structures in passive voice (agent) to cover a wide range of common expression variants.

Second, to improve the semantic integrity of the structure, this paper also introduces the noun_chunks mechanism, which replaces the recognized core nouns with complete compound phrases, to improve the boundary accuracy and clarity of expression of the extraction results. For example, in the sentence "The non-invasive blood glucose sensor measures the blood glucose level," the subject "sensor" can be expanded into a complete phrase "non-invasive blood glucose sensor," which is closer to the technical expression of the original text.

Finally, a syntactic constraint rule [14] is introduced to improve the accuracy of candidate structures by retaining only the typical 'noun-verb-noun' structures. The rules are shown in Table 1.

**Table 1.** Syntactic rules to identify SVO triplets

| Case | Dependency Rule | Example Sentence | SVO Triplet | POS Constraints |
|---|---|---|---|---|
| 1 | nsubj (verb, subject), dobj (verb, object) | The device monitors blood glucose. | (device, monitor, blood glucose) | NN*-VB*-NN* |
| 2 | xsubj (verb, subject), dobj (verb, object) | The device is used to monitor blood glucose. | (device, monitor, blood glucose) | NN*-VB*-NN* |
| 3 | agent (verb, subject), nsubjpass (verb, object) | Blood glucose is monitored by the device. | (device, monitor, blood glucose) | NN*-VB*-NN* |
| 4 | nsubj (verb, subject), prep (verb, prep), pobj (prep, object) | The device collects signals from the skin. | (device, collect, signals) | NN*-VB*-NN* |

#### 3.2.3 Lemmatization of SVO Structures

After the initial extraction of SVO candidate structures, there are still several expression redundancies and inconsistencies caused by writing form differences, lexical deformation and stop word interference. Therefore, word form reduction and semantic normalization of extraction results are necessary steps to achieve structural alignment and functional classification. To this end, this study introduces the Natural Language Toolkit (NLTK) to clean and normalize each component of the SVO candidates through the following steps.

First, regular expressions are applied to remove special characters, numbers, and overly short tokens from each field. Then, all text is converted to lowercase to eliminate the effect of case differences on semantic modeling.

Second, filtering words that are common in SVO candidate structures but have no real semantic contribution based on the StopwordList of patented technology categories summarized and constructed by Luo [29]. It covers words such as 'a,' 'the,' 'such,' 'system,' 'apparatus' and other generalized or templated terms that typically appear in patents, which can effectively improve the efficiency and accuracy of structure cleaning.

Third, the NLTK-WordNetLemmatizer is applied to lemmatize the SVO candidates separately. The process harmonizes the various types of morphological changes (e.g., detects / detecting to detect, devices to device), enhancing the consistency and comparability of functional expressions.

As a result, a set of normalized SVO structures is generated, which serves as the core semantic unit for subsequent functional semantic classification and comparative analysis.

### 3.3 SAO-Based Patent Functional identification

In technical texts, functional information is often reflected through the “action” in the language structure. Therefore, to identify “who performs what functional action on whom,” we construct an SAO (Subject-Action-Object) structure and map the actions into explicit functional types, which is the point in this section.

#### 3.3.1 Semantic Classification of Key Phrases

The first step is to categorize the key phrases. This paper extracts Key Phrase (KP) [19] based on subject and object in the SVO structure and utilizes semantic embedding and clustering methods to realize KPs' categorization. The semantic labels of the key phrases will be used as an important basis for the subsequent functional categorization rules.

1. Key Phrase Embedding Generation.
   Key phrases are phrases extracted from structured texts to represent technical elements. In patent texts, these phrases usually represent technical entities such as devices, materials, signals, properties, or physiological targets (e.g., sensor, blood glucose level, polymer coating). Here, we adopt the SciBERT model [14] that can capture the semantic differences between scientific terms and composite technical expressions to generate a semantic embedding vector containing all key phrases. SciBERT similarity is defined as the cosine similarity between the embedding vector of a candidate key phrase and the centroid vector of a key phrase category. Specifically, let $\vec{k}$ denote the embedding of a candidate key phrase, and

$\vec{c_j}$ represent the semantic centroid of the key phrase category $K_j$, which is computed as the mean of all seed key phrase embeddings within that category. The similarity between a key phrase $k$ and category $K_j$ is then calculated as follows.

$$\mathrm{Sim}(k, K_j) = \frac{\vec{k} \cdot \vec{c_j}}{\left|\vec{k}\right| \left|\vec{c_j}\right|} \quad (1)$$

By selecting the category $K_j$ with the highest similarity score, each key phrase can be automatically assigned to the most semantically appropriate functional class.

2. Construction of the Similarity Matrix.
   Based on the KPs' embedding vectors, the similarity matrix of key phrases is constructed by calculating the semantic similarity between any two phrases. With this matrix, semantic similarity between functional entities can be discovered systematically without relying on artificial rules. Table 2 shows an example of the similarity matrix of KPs.

3. Clustering and classification of key phrases.
   To realize the semantic clustering of key phrases, based on the similarity matrix, this study adopts the K-Means clustering method [30] to cluster and analyze the embedded representations of all KPs. Since K-means can automatically identify the semantic structure among phrases under unsupervised conditions and group semantically similar functional entities into the same category. After the clustering has been completed, the most representative key phrases in each cluster are combined and categorized for labeling. Eventually, all the clustering results are mapped into several semantic categories. Thus, each key phrase obtains a unified and semantically interpretable category label KP-class.

**Table 2.** KP similarity matrix

| | **salt form** | **fagopyritol** | **aerosol form** | **oblong needle** | **intradermal inj** | **inhibitory act** | **disease cond** | **surface area** |
|---|---|---|---|---|---|---|---|---|
| **salt form** | 1.000 | 0.841 | 0.857 | 0.820 | 0.811 | 0.758 | 0.824 | 0.785 |
| **fagopyritol** | 0.841 | 1.000 | 0.753 | 0.805 | 0.814 | 0.795 | 0.745 | 0.741 |
| **aerosol form** | 0.857 | 0.753 | 1.000 | 0.762 | 0.790 | 0.731 | 0.749 | 0.761 |
| **oblong needle** | 0.820 | 0.805 | 0.762 | 1.000 | 0.871 | 0.758 | 0.803 | 0.776 |
| **intradermal inj** | 0.811 | 0.814 | 0.790 | 0.871 | 1.000 | 0.815 | 0.783 | 0.751 |
| **inhibitory act** | 0.758 | 0.795 | 0.731 | 0.758 | 0.815 | 1.000 | 0.722 | 0.745 |
| **disease cond** | 0.824 | 0.745 | 0.749 | 0.803 | 0.783 | 0.722 | 1.000 | 0.782 |
| **surface area** | 0.785 | 0.741 | 0.761 | 0.776 | 0.751 | 0.745 | 0.782 | 1.000 |

3.3.2 Verb Semantic Expansion and Classification

In the second step, we map the action verbs extracted from SAO structures into three functional fields. Following previous studies that annotate actions with functional types by establishing functional categories through action words [31], we adopt a three-type scheme that distinguishes Inclusion, Attribute, and Effect verbs. The basic idea is that Inclusion-type verbs express compositional or structural relations between components (e.g., include, have, composed of), Attribute-type verbs express states, measurements, or control conditions (e.g., stabilize, measure, increase), and Effect-type verbs express functional or physical effects on materials, devices, or the human body (e.g., absorb, cool, destroy, detect). Representative seed verbs for each type are summarized in Table 3.

Based on this, we construct a semantic vector space for verbs and extend the systematic categorization of verb functions defined by the three Functional Structure Types and their representative seed verbs, using the SciBERT pre-trained language model. For each functional type, we compute the mean SciBERT embedding of all its seed verbs, which serves as the semantic center of that category. We then calculate the cosine similarity between the embedding of each verb and the center of every category. Each verb is assigned to the functional type whose semantic center has the highest cosine similarity to that verb embedding; in other words, the classification rule is to label each verb with the functional category that is most similar to it in the SciBERT-based vector space.

In practice, however, some high-frequency verbs are polysemous and may play different functional roles depending on the context. For example, the verbs “control” and “regulate” can describe process-oriented therapeutic actions (e.g., “regulate immune response”), which correspond to Effect-type expressions, but they can also describe measurement or maintenance operations (e.g., “control blood glucose level within a target range”), which are closer to Attribute-type expressions. Similarly, “provide” can act as an Inclusion-type verb when it denotes the presence or composition of components (e.g., “provide a sensing module”), yet in some sentences it refers to a more abstract service-level effect. Such cases constitute sources of classification uncertainty even when the SciBERT-based similarity rule is applied, and they are explicitly acknowledged in our analysis.

**Table 3.** Functional structure types and representative seed action words

| Functional structure types | Action words |
|---|---|
| **Inclusion type** | Include, have, composed of, comprise, supply, assemble. |
| **Attribute type** | Stabilize, vibrate, intensify, increase, measure, stabilize. |
| **Effect type** | Absorb, accumulate, clean, condense, cool, destroy, detect, erode, evaporate, extract, freeze, boil, hear, hold, mix, move, polish, produce, remove, separate. |

3.3.3 Mapping VO Structures to Functional Types

After classifying verbs into functional types and assigning semantic labels to KPs, this study further constructs a rule-based framework to map each verb–object (VO) structure to one of three functional types: Inclusion, Attribute, or Effect. This rule logically considers the “functional intent of the verb” and the “technical attributes of the object” and can more accurately classify the VO structure into the appropriate functional type, which is described in the case study.

## 4. Empirical Analysis

### 4.1 Data collection

To illustrate the proposed method, we selected the diabetes-related patents as an empirical case. The United States and China were selected as representative countries for comparative analysis, and the dataset was constructed from the WIPSON patent database using the diabetes-focused retrieval strategy in Table 4, which combines diabetes terms with treatment, diagnosis, monitoring and device keywords. This query yields 16,278 U.S. patents (1970s–2024) and 12,910 Chinese patents (1990s–2024). For each patent, we use the patent number, IPC codes, country, publication year and English abstract for SAO extraction and analysis and restrict cross-country comparisons to the overlapping period from the mid-1990s onward, where both countries have comparable patent volumes.

**Table 4.** Boolean search query used for retrieving diabetes patents

| Boolean search query |
|---|
| (diabetes OR diabetic OR “blood glucose” OR “blood sugar”) AND (treatment OR therapy OR diagnosis OR detection OR monitoring OR sensor OR device OR insulin OR medication OR drug OR “glucose control” OR CGM OR “continuous glucose monitoring” OR “artificial pancreas”) |

### 4.2 Identification of Functional Structures

4.2.1 Extraction of Subject-Verb-Object triplets

For the analysis of functional structures, we use only the English abstract of each patent. After preprocessing and lexical tagging, 45,171 U.S. patent clauses and 42,022 Chinese patent clauses were obtained for SVO extraction.

4.2.2 SAO-based Function identification

The functional structures within these clauses are identified using the SAO-based pipeline described in Section 3. In brief, key phrases are first embedded with SciBERT and clustered into semantic groups, which are then consolidated into five KP classes—Technology, Product/Device, Material/Chemical, Signal/Technical Attribute and Physio—as defined with examples in Table 5. Verbs are then assigned to one of three action classes (Inclusion, Effect or Attribute) on the basis of their SciBERT-derived similarity to the functional category centers; representative similarity scores are reported in Table 6. Finally, verb–object pairs are mapped to functional types according to the combination of verb class and KP class, and Table 7 gives examples of how sentences are classified into Inclusion, Effect and Attribute types together with the corresponding reasoning.

To analyze functional expressions across technical domains, we further use International Patent Classification (IPC) codes to construct sub-technical domains. The first four digits of each IPC code are extracted and aggregated into six sub-TDs, following the scheme proposed by medical experts [32]: Medical Devices, Pharmaceutical Use, Chemical Compounds, Health Care IT, Regenerative Medicine Technologies and Other. Figure 2 illustrates the distribution of patent counts across these sub-TDs for China and the United States.

**Table 5.** KP classes with definitions and examples

| Class of KP | Definition and Examples |
|---|---|
| **Technology** | Processes, methods, components (e.g., analysis, coding, algorithm) |
| **Product / Device** | Technology-related products (e.g., sensor, patch, capsule, device) |
| **Material / Chemical** | Drugs, molecules, chemical substances (e.g., peptide, compound, ligand) |
| **Signal / Technical Attribute** | Technical status, values, or parameters (e.g., blood glucose, level, temperature) |
| **Physio** | Physiological structures or biological targets (e.g., beta cell, pancreas) |

**Table 6.** Similarity scores between verbs and functional categories

| Verb | Action Class | Inclusion | Effect | Attribute |
|---|---|---|---|---|
| disclose | **Inclusion** | 0.7974 | 0.7964 | 0.7724 |
| encompass | **Inclusion** | 0.9122 | 0.9044 | 0.8884 |
| contain | **Inclusion** | 0.9343 | 0.8420 | 0.8059 |
| rise | **Attribute** | 0.8772 | 0.8934 | 0.8997 |
| decrease | **Attribute** | 0.8154 | 0.8717 | 0.9163 |
| improve | **Attribute** | 0.8365 | 0.8761 | 0.8849 |
| mitigate | **Effect** | 0.9119 | 0.9195 | 0.9056 |
| disperse | **Effect** | 0.8218 | 0.9029 | 0.8920 |
| condense | **Effect** | 0.7437 | 0.8088 | 0.7841 |

**Table 7.** Examples of function type classification based on verb and key phrase class

| Sentence | Function Type | Reason |
|---|---|---|
| The device includes a blood glucose sensor. | Inclusion | include + Product/Device KP |
| The system detects a glucose spike. | Effect | detect + Any KP |
| The algorithm improves insulin dose accuracy. | Attribute | improve + Attribute KP |

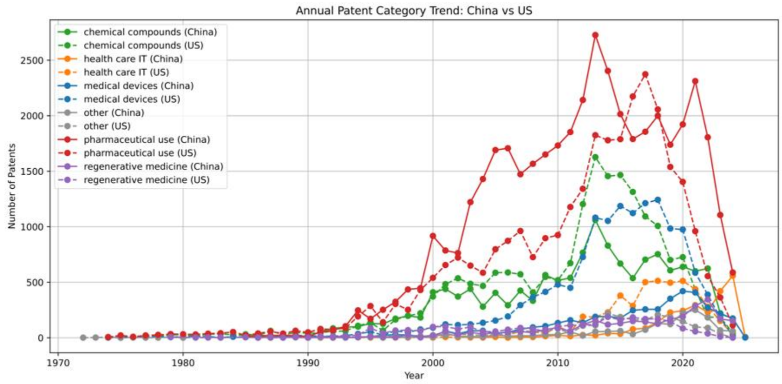



**Figure 2.** Sub-TD trend

4.3 Results

Based on the above studies, a systematic analysis of the evolutionary characteristics of functional expression was performed in the diabetes patent set in China and the United States.

To this end, the analysis is conducted across multiple dimensions, including:

Function Type with Year: Trends in the temporal evolution of function types.
Function Type with Sub-Technical Domain (sub-TD): Functional focus under different sub-TDs.
Function Type with Key Phrases: Differences in the object of functional expression.

Through statistical and visualization analysis of these dimensions, it is possible to reveal the technological focus, differences in R&D paths, and linguistic expression strategies of the US-CN in the layout of diabetes patents at the level of language structure.

4.3.1 Analysis of the Evolutionary Trend of Function Type over Time

In order to reveal the long-term evolutionary trend of diabetes technologies in terms of functional expressions, this study used functional types (Attribute, Effect, and Inclusion) as the core functional classification dimensions, counted the frequency of occurrence of each type of functional structure in patents in China and the United States in different time periods, and plotted the annual evolution curves between 1975 and 2024. The top20 functional structures are shown in Table 8 and Table 9.

1. Evolutionary Trends in U.S. diabetes Patents.
   As shown in Figure 3(a), the number of functional structures in U.S. diabetes patents increases markedly from the early 1990s onward. Effect-type structures are still relatively modest in the early 1990s but begin to grow continuously after 1992 and rise much more sharply after 1994–1995. They exhibit a first pronounced surge in the late 1990s, and a second surge from the early 2000s to the mid-2010s, with the highest values appearing around 2015. This type of expression, such as “kill activated CD4 cell”, “induce specific cytotoxic lymphocyte”, and “produce sensor signal”, reflects the significant increase in attention to treatment mechanisms, efficacy processes, and the functional behavior of medical devices in the U.S. during these periods.
   Inclusion-type structures display a similar two-stage pattern. They start to increase in the mid-1990s and experience an early jump in the late 1990s (e.g., more than one hundred structures in 1998), indicating that compositional descriptions of device combinations and system layouts became important relatively early. From the late 2000s to the mid-2010s, Inclusion structures remain at high levels and reach a distinct peak in 2013. Typical examples include “include infusion pump”, “include glucose testing measurement module housing”, and “facilitate precise metering”.
   The total number of Attribute-type structures is comparatively low but shows a gradual increase after 2000 with moderate fluctuations. They are mostly used to describe system parameters, control indicators, or physiological states, such as “modulate immune response” and “control blood glucose level”.
2. Evolutionary Trends in China diabetes Patents.
   As shown in Figure 3(b), Chinese diabetes patents exhibit a different rhythm in terms of functional types. Effect-type structures start to grow continuously from the late 1990s, form an initial consolidation in the

early 2000s, and then increase rapidly after 2010, reaching a sharp peak around 2013. This pattern indicates that China placed strong emphasis on process-centered therapeutic solutions during this period, with linguistic expressions focusing on drug action pathways and physiological regulatory mechanisms, such as “promote blood circulation”, “exhibit anti-obesity effect”, and “promote wound healing”.

Inclusion-type structures remain relatively low before 2010 but already show a gradual upward trend from the late 1990s and then grow rapidly after 2010, reaching a historical high around 2020. This suggests that compositional expressions such as device combinations and component nesting have gradually become mainstream, especially in patents related to medical devices and sensing systems, such as “include continuous discrete acquisition”, “include handheld medical device”, and “disclose blood glucose detection device”.

Similar to the U.S. case, Attribute-type structures account for a relatively small proportion in Chinese patents but display a moderate upward trend with a local maximum in the early 2010s. They are mainly used to describe local attribute-type functions such as detection parameters and stability maintenance, for example “measure patient pain” and “regulate blood sugar level”.

3. Comparison of Evolutionary Trends between the U.S. and China.

   Overall, the United States enters a stage in which Inclusion-type expressions become as prominent as Effect-type expressions earlier than China. In the U.S. data, Inclusion-type functional structures start to increase from the mid-1990s, show an early jump in the late 1990s, and then remain at high levels from the late 2000s to the mid-2010s, with a distinct peak in 2013. By contrast, in China, Inclusion-type structures stay relatively low before 2010 and only begin to grow rapidly thereafter, reaching their historical maximum around 2020.

   For Effect-type expressions, both countries exhibit strong growth, but with different temporal profiles: U.S. Effect-type structures show two main surges, culminating in a peak around 2015, whereas Chinese Effect-type structures rise sharply after 2010 and reach a sharp peak around 2013. In both countries, Attribute-type expressions remain marginal in terms of volume. They never form a dominant functional pattern, yet they persist as a supporting expression type in technological stages where parameters and physiological states need to be explicitly described.

4. External validation and added value of the SAO-based analysis

   These SAO-based patterns are consistent with prior studies on diabetes technologies and diabetes-related innovation. Reviews of diabetes technology and continuous glucose monitoring report that major advances in CGM devices, insulin pumps, and closed-loop “artificial pancreas” systems are driven by device- and system-oriented manufacturers, particularly in the United States, and emphasize modularized designs and standardized interfaces between sensors, pumps, and control units [33]. At the same time, patent and clinical analyses of antihyperglycemic therapies and traditional Chinese medicine prescriptions for diabetes show that a large share of diabetes-related innovation in China focuses on drug formulations, herbal compositions, and pharmacological mechanism–oriented treatments rather than medical devices [7], [35, 36]. These external studies provide literature-based support for our inference that U.S. diabetes patents are more closely related to devices and systems, whereas Chinese patents place greater emphasis on drug mechanisms and therapeutic pathways.

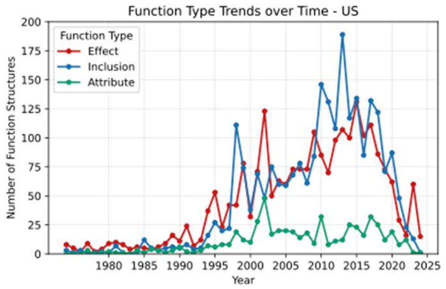
   However, while such broad tendencies may already be partly familiar to domain experts, our SAO-based framework adds value by quantifying and structuring these patterns. By classifying SAO structures into Effect, Inclusion, and Attribute types and tracking their temporal evolution, we can measure the relative prevalence of different functional expressions in each country and to reveal specific action–object combinations and the timing of shifts from process-centered to composition-centered expressions. These finer-grained functional patterns would be difficult to obtain from IPC distributions or simple keyword statistics alone.

**Table 8.** Top 20 functional structures by functional type in U.S. diabetes patents

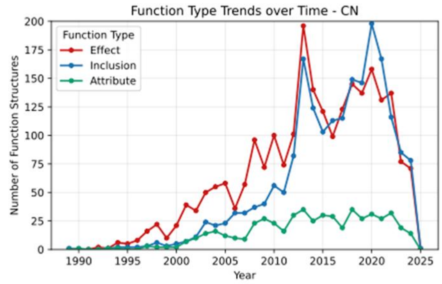
| Effect | Inclusion | Attribute |
|---|---|---|
| produce sensor signal | include conductive contact | regulate flux |
| induce psychosis | include sample chamber | improve glucose tolerance |
| mediate effect | include plurality | catalyze formation |
| play role | comprise administration | cause meter |
| reset hormonal timing | determine analyte concentration | cause detectable change |
| display rate | include measurement | cause acute drop |
| inhibit expression | include test strip | control actuator |
| detect analyte | include infusion pump | limit alzheimer disease |
| play pathophysiological role | comprise first residue | implement feature |
| inhibit hydroxysteroid dehydrogenase type | comprise plurality | reduce risk |
| exhibit effect | involve administration | increase insulin secretion |
| contemplate method | describe novel method | improve accuracy |
| input data | include electrode | reduce biological organism |
| develop script program | include treatment | varicose vein |
| extend duration | include visualization | feature pharmaceutical composition |
| induce specific cytotoxic lymphocyte | facilitate precise metering | reduce side effect |
| inhibit development | include health care provider apparatus | modulate immune response |
| kill activated cd4 cell | include glucose testing measurement module housing | cause susceptibility |
| inhibit ige response | include medical device | reduce hematocrit effect |
| inject bolus dose | comprise chain polypeptide | neutralize activity |

**Table 9.** Top 20 functional structures by functional type in Chinese diabetes patents

| Effect | Inclusion | Attribute |
|---|---|---|
| solve problem | disclose soybean milk | reduce blood sugar |
| inhibit activity | disclose preparing method | reduce content |
| play role | include glucose sensor | reduce blood glucose |
| overcome defect | include continuous discrete acquisition | reduce level |
| reach percent | contain active component | control blood sugar |
| meet requirement | comprise active component | reduce blood fat |
| say compound | disclose making method | reduce blood glucose level |
| promote healing | disclose synthesis method | reduce body weight |
| play important role | disclose blood glucose detection device | reduce blood sugar concentration |
| induce psychosis | disclose noninvasive blood glucose detection method | feature pharmaceutical composition |
| treat diabetes | comprise bed body | reduce fasting blood glucose |
| wolfberry fruit powder | comprise light unit | reduce fasting blood glucose level |
| promote proliferation | include heater | reduce blood glucose concentration |
| exhibit effect | include infusion pump | reduce blood sugar content |
| mediate effect | contain coffee powder | reduce side effect |
| promote secretion | comprise effective component | regulate blood sugar |
| treat disease | comprise effective amount | lower blood sugar |
| accelerate secretory function | comprise power supply | reduce blood sugar level |
| treat symptom | disclose portable automatic injection device | decrease blood sugar |
| inhibit activation | disclose detection method | concern method |




(**a**) United States




(**b**) China

**Figure 3.** Trends of function types in diabetes patents by country: **(a)** Trends of function types in diabetes patents of U.S.; **(b)** Trends of function types in diabetes patents of China.

4.3.2 Analysis of the distribution of function types under sub-TDs

This paper further examines the distribution characteristics of functional expressions in various types of technology directions and the differences from the perspective of sub-technology domains between China and the U.S. The results are shown in Figure 4.

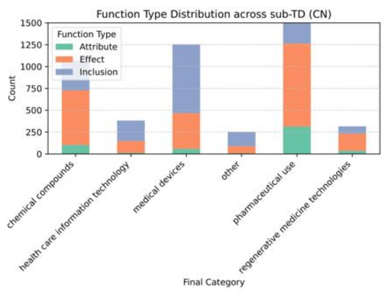
In the field of medical devices, U.S. patents show a very high density of functional structures, and the number of inclusion type structures is particularly prominent, reflecting a more systematic language expression in terms of device combination and module composition. For example, common structures include “compare feedback health condition data” and “comprise electrode layer” focusing on technology integration and device composition. China, on the other hand, also focuses on the Effect and Inclusion types in this domain, but the total number is slightly lower, indicating that it is still in a growing stage in terms of system integration.

In the direction of pharmaceutical use, effect type structures are active in both China and the United States, especially in China, with functional expressions related to the mechanism of efficacy, therapeutic targets and response processes, such as “promote vascular regeneration,” “reverse partial dysfunction type,” “inhibit diabetes muscular atrophy” and “inhibit lipid synthesis pathway.” This suggests that this area is the main direction in which the functional expression of diabetes therapeutic technology is most concentrated in 2 countries.

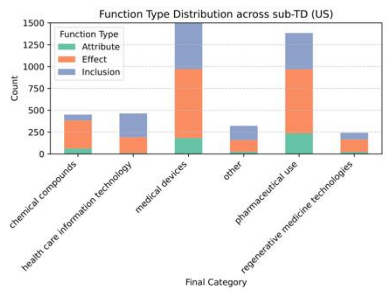
In the fields of chemical compounds and regenerative medicine technologies, both China and the United States predominantly use Effect-based expressions, such as “promote vascular regeneration,” “reverse partial dysfunction type,” “inhibit aberrant protease activity” and “diagnose autoimmune disease.” It reflects a linguistic preference for describing therapeutic mechanisms and biological pathways in these fields.

It should be noted that in both 2 countries, the number of functional structures in the field of healthcare information technology is relatively small, which suggests that diabetes-related patents have paid limited attention to this area and that functional modeling in this field remains underdeveloped. The use of Include-type structures was particularly prominent in the United States, suggesting a focus on component-level description and system integration. In contrast, China relies more on Effect-type representations, suggesting a greater focus on drug development and modeling of therapeutic effects.

These differences not only reveal different patterns of functional expression but also reflect underlying technology development strategies: the U.S. tends to favor system-level integration and structural modeling, while China emphasizes functional effects and mechanisms of action of diabetes-related technologies.




(**a**) China




(**b**) United States

**Figure 4.** Function type distribution across sub-TDs

4.3.3 Functional type combined with key phrases structures analysis

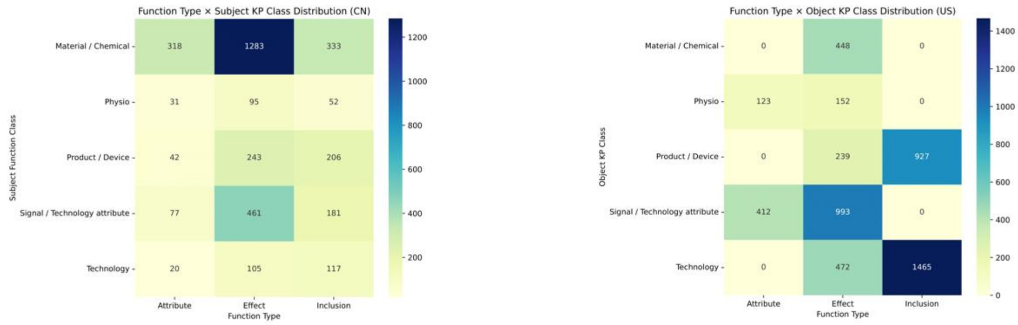
To further reveal the semantic structural differences of functional expressions, this paper also counts the distribution of key phrases (KPs) of different functional types in diabetes patents and draws a group of heat maps of China and the United States.

1. KP Class of Subject and Function Type.
   From the perspective of subject matter, the subject setting in the functional expression structure of China and the United States reflects different technical expression tendencies, which are shown in Figure 5. When “Material / Chemical” is used as the subject of Chinese diabetes patents, it is mainly used to describe the function of Effect type, which indicates that China pays more attention to the result-oriented language pattern of “a certain substance produces a certain effect” in the field of diabetes. In contrast, the U.S. has a significantly smaller number of such constructions and prefers to use technical concepts such as “Technolog” or “Signal / Technical Attribute” as the subject and combine them with the function of “Inclusion” to form structural expressions such as “a system includes a module” or “a signal has a certain characteristic.”
   It is noteworthy that the number of U.S. subject matter of “Product/Device” in combination with a generalized function is much higher than that of China, which is further evidence that the U.S. places more emphasis on systematic descriptions of devices and components in the expression of diabetes patents.

2. KP Class of Object and Function Type.
   In Figure 6, China and the United States also show significant differences in semantic expressions in terms of the distribution of functional types and object key phrase categories.
   In Chinese diabetes patents, there is a high frequency of functional expressions with “Material/Chemical” as the object and combined with the type of “Effect.” This reflects that Chinese diabetes patent texts tend to emphasize more on the effect of the drug and the regulation of physiological processes by the chemical composition. In contrast, the lower number of U.S. patents under the same structure suggests that the U.S. places relatively less emphasis on such expressions.
   On the other hand, in the class of “Technology” and “Signal/Technical Attribute,” there are significantly more functional expressions in the U.S. patents than in the Chinese patents in terms of both Attribute and Effect, showing that the U.S. patents are more inclined to take the technical terms as the functional objects in terms of the system composition and signal response.
   In addition, in the class of “Product / Device,” both China and the U.S. focus on the functional pairing with the Inclusion type, indicating that equipment or device as part of the system composition is a more common expression logic in multinational patent drafting.

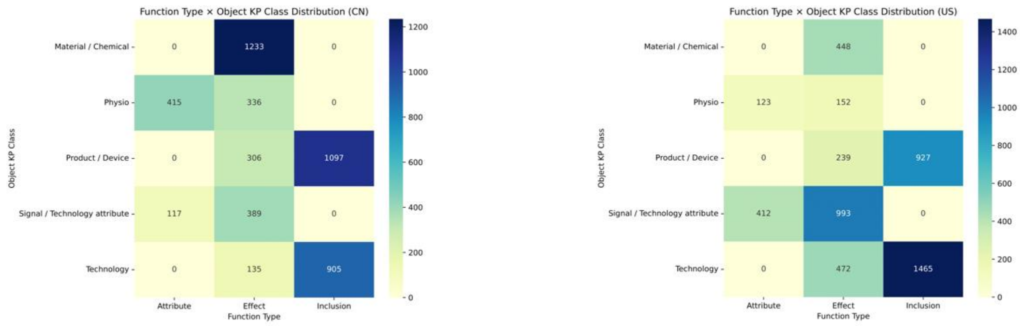
3. Summary of overall trends and differences in presentation styles.
   Chinese patents focus more on “Effect” expressions centered on drug components and physiological mechanisms, emphasizing the therapeutic effects of blood glucose regulation and physiological improvement triggered by a certain chemical substance, therapeutic means, or component of traditional Chinese medicine. The U.S., on the other hand, is more inclined to adopt the “structural composition” expression with equipment, system and signal as the subject or object, and pays more attention to the systematic composition of diabetes management technology, the logic of monitoring and regulation of physiological parameters (e.g., blood glucose, insulin level).

In addition, although the use of products or devices in the Inclusion type of structure is more consistent between the two countries, there are systematic differences in the subject and object categories relied upon in the Effect and Attribute types of expression. This difference not only reflects the different emphasis on diabetes R&D and patent writing styles between China and the United States, but also suggests that the diversity of semantic styles, expression logics and technological directions should be considered when analyzing diabetes technology trends and functional evolution studies between different countries in the future.




(**a**) China - subject (**b**) United States - Subject

**Figure 5.** Function type with kp - subject class distribution




(**a**) China - object (**b**) United States - object

**Figure 6.** Function type with kp - object class distribution

## 5. Discussion and Conclusion

This study proposes a method for extracting and categorizing patent functional expressions based on the SAO (Subject-Action-Object) structures and performs a semantic comparative analysis with the diabetes patent sets of China and the United States. Through SVO extraction algorithm, verb semantic clustering based on SciBERT, key phrase embedding and clustering analysis, and VO to function type rule mapping, a set of unified semantic modeling framework for functional expression is constructed, which can mine the core functional structure in patent text in a more systematic way. In terms of domain delineation, this paper uses the IPC as the basis for the delineation of sub-TDs, which has a strong semantic generalization ability.

The results show that U.S. patents are more related to device and system, with standardized language structure and modularized functional organization, whereas Chinese patents pay more attention to the expression of drug mechanism and therapeutic pathway, with drug driven as the core of the language style. These trends broadly align with prior research on diabetes technologies and treatments in China. Manufacturers of medical devices and systems play a central role in developing continuous glucose monitoring (CGM) and insulin pump systems, while treatment approaches centered on drugs and traditional Chinese medicine occupy a significant share of China's diabetes innovation landscape. Our analysis further refines and quantifies these differences by revealing the relative importance and temporal evolution of three functional categories (Effect, Inclusion, Attribute) in the two countries. This perspective uncovers patterns of functional expression within technological strategies, going beyond mere differences in technological domains.

We also found that in terms of time dimension, the U.S. has entered the expression phase dominated by "Effect to Inclusion" at an early stage, while China has rapidly increased the complexity of functional structures since 2010, forming its own evolutionary rhythm. The functional structure distribution of sub-TD further reveals the systematic advantage of the United States in the field of "medical devices," while China shows active expression density in the directions of "chemical drugs" and "regenerative medicine."

Several key patterns therefore emerge from the U.S.–China comparison. U.S. diabetes patents enter an "Effect-to-Inclusion" stage earlier and consistently emphasize modular device and system architectures, especially in medical device sub-TDs. Chinese patents are more strongly oriented toward drug mechanisms and therapeutic pathways and show a later but rapid increase in functional complexity after 2010, with dense expression in chemical drug and regenerative medicine sub-TDs. These patterns suggest complementary national strengths—device- and system-centered solutions in the U.S. versus drug- and mechanism-centered solutions in China. The empirical results are specific to diabetes-related medical patents in the WIPSON database and should be viewed as indicative functional tendencies in this domain rather than definitive characterizations of all medical or pharmaceutical innovation, whereas the SAO-based comparative framework itself is general and can be applied to other disease areas and technological fields in future research.

Methodologically, this study extends prior SAO-based patent analyses by developing a comparative SAO framework for cross-country analysis and by introducing a domain-specific functional classification of verbs into Inclusion, Attribute, and Effect types, operationalized with SciBERT-based embeddings and a cosine-similarity decision rule, which enables scalable and reproducible assignment of functional categories to actions in medical patents. Theoretically, our functional indicators provide a complementary perspective on national medical innovation to traditional measures such as IPC distributions, citation counts and topic modelling, by showing how diabetes-related problems are addressed in practice and how these pathways evolve over time.

However, although this study provides a systematic exploration at the methodological and empirical levels, several limitations still exist. First, the VO-to-function type mapping rule relies on a predefined semantic framework, which is interpretable, but still suffers from insufficient adaptability when dealing with polysemous verbs or boundary fuzzy structures. As discussed in Section 3.3.2, verbs such as ‘control’, ‘regulate’, and ‘provide’ may reasonably belong to different functional types depending on their local context, and relying solely on similarity to category centers may therefore lead to classification errors. In this study, we partially mitigated this issue by manually reviewing high-frequency verbs and refining the seed verb list, although more systematic validation is still needed. Future work could further refine and expand the functional seed sets, incorporate contextual features such as adjacent nouns and objects into the classifier, support multi-label assignments for highly polysemous verbs, and conduct large-scale manual annotation coupled with inter-annotator reliability analysis to enhance the robustness of the model. Second, some potentially useful SAO structures do not match the defined functional categories, which is ignored in statistical and trend analyses. In the future, it can be further deepened in domain extension, multilingual adaptation, and contextual chain relationship modeling to support a wider range of patent semantic intelligence analysis tasks. Third, the time span of the U.S. and Chinese datasets is not perfectly aligned. In the WIPSON database, U.S. diabetes-related patents matching our query appear from the 1970s, whereas Chinese patents become substantial only from the 1990s onward. Although we focus our interpretation of cross-country differences on the overlapping period from the mid-1990s onward, this temporal asymmetry may still influence the observed long-term trends and should be kept in mind when generalizing the results. Finally, in this study we rely on targeted literature-based validation rather than conducting new interviews with medical professionals to corroborate the SAO-based patterns; incorporating systematic expert interviews would be a valuable extension for future work.

In summary, the SAO-based functional expression modeling framework constructed in this paper provides structural support in cross-national patent comparison, semantic structure modeling and trend evolution analysis. By systematically extracting and categorizing functional structures in medical patents, key therapeutic mechanisms and device action modes can be effectively identified, providing semantic support and direction guidance for disease monitoring, drug release, and smart device design.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

[1] A. Turchin and L. F. Florez Builes, “Using natural language processing to measure and improve quality of diabetes care: a systematic review,” J. Diabetes Sci. Technol., vol. 15, no. 3, pp. 553–560, 2021, doi: https://doi.org/10.1177/19322968211000831.

[2] J. Kim, M. L. Chen, S. J. Rezaei, T. Hernandez-Boussard, J. H. Chen, F. Rodriguez, and E. Linos, “Artificial intelligence tools in supporting healthcare professionals for tailored patient care,” npj Digit. Med., vol. 8, p. 210, 2025, doi: https://doi.org/10.1038/s41746-025-01604-3.

[3] M. Litvinova, M. Eitenberger, A. Bilir, A. W. K. Yeung, E. D. Parvanov, A. MohanaSundaram, and H. Willschke, “Patent analysis of digital sensors for continuous glucose monitoring,” Front. Public Health, vol. 11, p. 1205903, 2023, doi: https://doi.org/10.3389/fpubh.2023.1205903.

[4] J. Hirtz, R. B. Stone, D. A. McAdams, S. Szykman, and K. L. Wood, “A functional basis for engineering design: reconciling and evolving previous efforts,” Res. Eng. Des., vol. 13, pp. 65–82, 2002, doi: https://doi.org/10.1007/s00163-001-0008-3.

[5] C. Yang, C. Huang, and J. Su, “An improved SAO network-based method for technology trend analysis: A case study of graphene,” J. Informetr., vol. 12, no. 1, pp. 271–286, 2018, doi: https://doi.org/10.1016/j.joi.2018.01.006.

[6] C. Mun, S. Yoon, N. Raghavan, D. Hwang, S. Basnet, and H. Park, “Function score-based technological trend analysis,” Technovation, vol. 101, p. 102199, 2021, doi: https://doi.org/10.1016/j.technovation.2020.102199.

[7] W. Zhu, W. Huang, Z. Xu, M. Cao, Q. Hu, C. Pan, and H. Yuan, “Analysis of patents issued in China for antihyperglycemic therapies for type 2 diabetes mellitus,” Front. Pharmacol., vol. 10, p. 586, 2019, doi: https://doi.org/10.3389/fphar.2019.00586.

[8] E. O. Olorunsola, M. F. Alozie, K. G. Davies, and M. O. Adedokun, “Advances in the science and technology of insulin delivery: a review,” J. Appl. Pharm. Sci., vol. 11, pp. 184–191, 2021. [Online] Available: https://japsonline.com/abstract.php?article_id=3437&sts=2

[9] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, “The Stanford CoreNLP natural language processing toolkit,” in Proc. 52nd Annu. Meeting Assoc. Comput. Linguist.: Syst. Demonstr., pp. 55–60, 2014, doi: https://doi.org/10.3115/v1/P14-5010.

[10] J. Lim, S. Choi, C. Lim, and K. Kim, “SAO-based semantic mining of patents for semi-automatic construction of a customer job map,” Sustainability, vol. 9, p. 1386, 2017, doi: https://doi.org/10.3390/su9081386.

[11] M. Marcus, B. Santorini, and M. A. Marcinkiewicz, “Building a large annotated corpus of English: the Penn Treebank,” Comput. Linguist., vol. 19, pp. 313–330, 1993, doi: https://doi.org/10.21236/ADA273556.

[12] K. W. Church, “A stochastic parts program and noun phrase parser for unrestricted text,” in Proc. Int. Conf. Acoust., Speech, Signal Process., pp. 695–698, 1989, doi: https://doi.org/10.1109/ICASSP.1989.266522.

[13] I. Beltagy, K. Lo, and A. Cohan, “SciBERT: a pretrained language model for scientific text,” arXiv preprint arXiv:1903.10676, 2019, doi: https://doi.org/10.18653/v1/D19-1371.

[14] H. Cheong, W. Li, A. Cheung, A. Nogueira, and F. Iorio, “Automated extraction of function knowledge from text,” J. Mech. Des., vol. 139, p. 111407, 2017, doi: https://doi.org/10.1115/1.4037817.

[15] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” Adv. Neural Inf. Process. Syst., vol. 26, 2013, doi: https://doi.org/10.48550/arXiv.1310.4546.

[16] W. Lin, W. Yu, and R. Xiao, “Measuring patent similarity based on text mining and image recognition,” Systems, vol. 11, p. 294, 2023, doi: https://doi.org/10.3390/systems11060294.

[17] J. B. Gao, B. W. Zhang, and X. H. Chen, “A WordNet-based semantic similarity measurement combining edge-counting and information content theory,” Eng. Appl. Artif. Intell., vol. 39, pp. 80–88, 2015, doi: https://doi.org/10.1016/j.engappai.2014.11.009.

[18] Y. Cai, Q. Zhang, W. Lu, and X. Che, “A hybrid approach for measuring semantic similarity based on IC-weighted path distance in WordNet,” J. Intell. Inf. Syst., vol. 51, pp. 23–47, 2018, doi: https://doi.org/10.1007/s10844-017-0479-y.

[19] S. Choi, H. Park, D. Kang, J. Y. Lee, and K. Kim, “An SAO-based text mining approach to building a technology tree for technology planning,” Expert Syst. Appl., vol. 39, pp. 11443–11455, 2012, doi: https://doi.org/10.1016/j.eswa.2012.04.014.

[20] X. Li, C. Wang, X. Zhang, and W. Sun, “Generic SAO similarity measure via extended Sørensen–Dice index,” IEEE Access, vol. 8, pp. 66538–66552, Mar. 2020, https://doi.org/10.1109/ACCESS.2020.2984024.

[21] S. Kim, I. Park, and B. Yoon, “SAO2Vec: development of an algorithm for embedding the subject–action–object (SAO) structure using Doc2Vec,” PLOS ONE, vol. 15, p. e0227930, 2020, doi: https://doi.org/10.1371/journal.pone.0227930.

[22] J. Wang and Z. Ding, “Technology opportunity discovery based on patent analysis: a hybrid approach of subject–action–object and generative topographic mapping,” Technol. Anal. Strateg. Manag., vol. 36, no. 9, pp. 2070–2083, 2024, doi: https://doi.org/10.1080/09537325.2022.2126306.

[23] M. Park, S. Kim, and B. Yoon, “Developing automatic-labeled topic modeling based on SAO structure for technology analysis,” PLOS ONE, vol. 20, no. 8, p. e0330275, 2025, doi: https://doi.org/10.1371/journal.pone.0330275.

[24] S. Kim and H. Jang, “Ko-SAO structure for functional analysis of Korean patents in energy field,” J. Korean Inst. Ind. Eng., vol. 51, no. 2, pp. 194–200, 2025, doi: https://doi.org/10.7232/JKIIE.2025.51.2.194.

[25] X. Xi, J. Zhao, L. Yu, and C. Wang, “Exploring the potentials of artificial intelligence towards carbon neutrality: technological convergence forecasting through link prediction and community detection,” Comput. Ind. Eng., vol. 190, p. 110015, 2024, doi: https://doi.org/10.1016/j.cie.2024.110015.

[26] A. J. Trappey, R. J. Pa, N. K. Chen, A. Z. Huang, K. A. Li, and L. P. Hung, “Digital transformation of technological IP portfolio analysis for complex domain of satellite communication innovations,” Adv. Eng. Inform., vol. 55, p. 101879, 2023, doi: https://doi.org/10.1016/j.aei.2023.101879.

[27] A. J. Trappey, F. Li, C. H. Lee, J. P. Mo, J. Stjepandic, and R. Jiao, "How to manage and balance uncertainty by transdisciplinary engineering methods focusing on digital transformations of complex systems," Adv. Eng. Inform., vol. 59, p. 102330, 2024, doi: https://doi.org/10.1016/j.aei.2023.102330.

[28] G. G. Chowdhury, "Introduction to Modern Information Retrieval", Facet Publishing, pp. 1–528, 2010. [Online] Available: https://sigir.org/files/museum/introduction_to_modern_information_retrieval/frontmatter.pdf

[29] S. Sarica and J. Luo, "Stopwords in technical language processing," PLOS ONE, vol. 16, p. e0254937, 2021, doi: https://doi.org/10.1371/journal.pone.0254937.

[30] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proc. 5th Berkeley Symp. Math. Statist. Probab., Univ. California Press, vol. 5, pp. 281–298, 1967. [Online] Available: https://scispace.com/pdf/some-methods-for-classification-and-analysis-of-multivariate-4pswti19oz.pdf

[31] H. Jang, S. Park, and B. Yoon, "Exploring technology opportunities based on user needs: application of opinion mining and SAO analysis," Eng. Manag. J., vol. 35, pp. 209–222, 2023, doi: https://doi.org/10.1080/10429247.2022.2050130.

[32] D. L. Mayfield, "Medical patents and how new instruments or medications might be patented," Missouri Med., vol. 113, p. 456, 2016. [Online] Available: https://pubmed.ncbi.nlm.nih.gov/30228529/

[33] A. M. Almurashi, E. Rodriguez, and S. K. Garg, "Emerging diabetes technologies: continuous glucose monitors/artificial pancreases," J. Indian Inst. Sci., vol. 103, pp. 205–230, 2023, doi: https://doi.org/10.1007/s41745-022-00348-3.

[34] A. Daly and R. Hovorka, "Technology in the management of type 2 diabetes: Present status and future prospects," Diabetes Obes. Metab., vol. 23, pp. 1722–1732, 2021, doi: https://doi.org/10.1111/dom.14418.

[35] B. Pang, F.-M. Lian, X. Y. Zhao, X. M. Zhao, D. Jin, Y. Q. Lin, Y. J. Zheng, Q. Ni, and X. L. Tong, "Prevention of type 2 diabetes with the traditional Chinese patent medicine: a systematic review and meta-analysis," Diabetes Res. Clin. Pract., vol. 131, pp. 242–259, 2017, doi: https://doi.org/10.1016/j.diabres.2017.07.020.

[36] L. Jiang, S. Wang, J. Zhao, C. Chien, Y. Zhang, G. Su, X. Chen, D. Song, Y. Chen, W. Huang, Y. Xiao, Y. Cao, and Z. Hu, "Treatment options of traditional Chinese patent medicines for dyslipidemia in patients with prediabetes: a systematic review and network meta-analysis," Front. Pharmacol., vol. 13, p. 942563, 2022, doi: https://doi.org/10.3389/fphar.2022.942563.

© 2025 by the authors. Copyrights of all published papers are owned by the IJOC. They also follow the Creative Commons Attribution License (https://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.