# Research on the Selection of Innovative Solutions Based on SAO Structure: A Case Study of Air Purification Technology (Postprint)

**Authors:** Fu Yun, Wang Xuefeng, Li Jia, Hou Yujia

**Date:** 2023-07-26T00:00:00+00:00

## Abstract

[Purpose/Significance] Faced with an increasingly competitive social environment, innovation is the cornerstone of enterprise establishment and survival. The scope of innovation includes not only creating new products and technologies within one's own field, but also introducing new products and technologies that have emerged in other fields into one's own field, with the latter being relatively easier. However, as the degree of specialization in academic disciplines continues to increase, researchers lack sufficient capacity to master knowledge outside their own fields; therefore, scientific methods and technologies are needed to explore the deep-level connections between knowledge in different fields. [Method/Process] Drawing on the analytical process of the LRDI method, we propose an innovative solution selection method based on SAO structure. Starting from specific research questions in the target research field, we search for potential solutions across all fields and evaluate these potential solutions from two aspects: technical feasibility and expected effectiveness, thereby forming a prioritized recommendation of innovative solutions applicable to the target research field. [Results/Conclusion] An empirical study was conducted using air purification technology as an example. The research results show that some innovative solutions have been effectively applied in the field of air purification technology, further confirming the feasibility and effectiveness of the proposed research method.

## Full Text

# Research on the Selection of Innovative Solutions Based on SAO Structure: A Case Study of Air Purification Technology

**Fu Yun**[1], **Wang Xuefeng**[2], **Li Jia**[2], **Hou Yujia**[2] [1] National Science Library, Chinese Academy of Sciences, Beijing 100190 [2] School of Management and Economics, Beijing Institute of Technology, Beijing 100081

### Abstract

[**Purpose/Significance**] In today's increasingly competitive social environment, innovation is fundamental to enterprise survival and development. Innovation encompasses not only creating new products and technologies within a field, but also introducing existing technologies and products from other domains—a process that is often more feasible. However, as academic disciplines become increasingly specialized, researchers lack the time to master knowledge outside their own fields, necessitating scientific methods and technologies to explore deep connections between different domains.

[**Method/Process**] Drawing on the analytical process of the Literature-Related Discovery and Innovation (LRDI) methodology, this study proposes a method for selecting innovative solutions based on Subject-Action-Object (SAO) structure. Starting from specific research problems in the target field, the method searches for potential solutions across all domains and evaluates these candidates from both technical feasibility and expected effectiveness perspectives, ultimately generating prioritized recommendations for innovative solutions applicable to the target field.

[**Result/Conclusion**] An empirical study on air purification technology demonstrates that several selected innovative solutions have already been effectively applied in the air purification domain, confirming the feasibility and validity of the proposed method.

**Keywords:** innovative solutions; SAO structure; Literature-Related Discovery and Innovation (LRDI); technology roadmap; air purification technology

---

## 1. Introduction

Although scientific and technological development has led to increasingly fine-grained disciplinary specialization, practical analysis reveals high degrees of intersection and integration among research problems across different fields. Therefore, focusing on specific research questions without restricting their domain of origin may yield more potential solutions. However, researchers already expend considerable effort simply to master knowledge within their own fields, making it unrealistic to comprehensively grasp knowledge from other domains. Consequently, researchers typically engage in selective or targeted reading, which can easily lead to omissions of important information and prevents comprehensive analysis of deep knowledge relationships. Scientific analytical methods and tools are thus needed to connect target domain knowledge with other fields, drawing inspiration and providing more innovative solutions for target domain problems.

Comprehensive analysis of domestic and international research on innovative solutions reveals two main categories: (1) predicting new technologies and products, and (2) applying existing technologies and products. The former refers to technologies or products not yet present in the target field but predicted to emerge in the future, generally applicable to domain-leading enterprises. The latter aligns more closely with open innovation concepts, focusing on leveraging existing corporate technologies or products to analyze potential technological innovation opportunities. For technology-based companies, especially small and medium-sized enterprises, this approach enables R&D capability expansion with minimal resources. The present research focuses on utilizing existing technologies and products.

Current research on finding technological innovation opportunities through existing technologies and products falls into two streams: some concentrate within the target research domain itself, while others extend current technologies and products to broader application fields, such as the work of Korean scholar J. Yoon and colleagues. Research on applying technologies or products from other domains to solve target domain problems is particularly mature in biology and medicine, where the Literature-Related Discovery and Innovation (LRDI) method has been widely adopted and extended by scholars.

The LRDI methodology evolved from D.R. Swanson's Literature-Based Discovery (LBD) approach, proposed in 1986. The LBD method emerged in response to excessive disciplinary specialization and the resulting "knowledge fragmentation" phenomenon. Isolated knowledge units, though independent in origin, often contain overlooked logical relationships—what Swanson termed "Undiscovered Public Knowledge." Only by reconnecting these knowledge fragments can new associations emerge, potentially leading to novel discoveries. Swanson used Medline database titles as his data source, conducting word frequency analysis to identify subject term sets from two independent domains: one containing themes A and B, and another containing B and C, thereby establishing A-C connections through the common term set B. His work primarily focused on biology and medicine and yielded many important findings.

Subsequent researchers extended this approach. M.D. Gordon and R.K. Lindsay analyzed Medline abstracts and expanded the analytical unit from words to phrases. M. Weeber used MetaMap to map natural language to the Unified Medical Language System (UMLS), reducing terminology inconsistencies and significantly decreasing the workload of subject term selection. J. Stegmann and G. Grohmann incorporated co-word analysis to identify domain themes through

MeSH term clustering. In 2008, R.N. Kostoff introduced citation analysis, using expert judgment to identify core keywords, then exploring relationships among literature sets A (references), B (citing A), and C (citing core literature). Kostoff later formalized this approach as Literature-Related Discovery and Innovation (LRDI) in 2012, emphasizing the combination of discovery with innovation to solve research problems.

However, existing LRDI methods primarily rely on keyword analysis, with keyword relationships identified through expert opinion or co-word analysis. The former is time-consuming and labor-intensive, while the latter often yields ambiguous semantic relationships. To address this limitation, this study introduces SAO (Subject-Action-Object) structure analysis. SAO analysis is fact-oriented, originating from TRIZ theory, where Subject (S) represents the Solution, Action-Object (A-O) represents the specific research Problem, forming a Problem-Solution (P&S) approach. Compared to keywords, SAO structures build semantic relationships based on research content, better characterizing specific relationships between elements.

SAO-based semantic analysis has been widely applied in various contexts: mining patent data for technology monitoring; constructing patent maps to identify competitive trends; building technology roadmaps for forecasting and opportunity analysis; constructing technology trees for planning; calculating patent similarity to identify partners; assessing patent infringement risk; building functional databases to find solutions for new problems; identifying technology application fields; and analyzing R&D partners.

Building on previous research, this study proposes a systematic approach for selecting innovative solutions: guided by LRDI methodology and combined with SAO semantic analysis, it starts from specific problems in the target domain, mines solutions across all fields, and after comprehensive evaluation, identifies high-scoring solutions as innovative recommendations. The research framework is illustrated in Figure 1 [Figure 1: see original paper].

---

## 2. Innovative Solution Selection Methodology

The proposed method consists of two main steps: (1) identifying core technologies and problems in the target domain's research hotspots, and (2) selecting innovative solutions by using these problems as starting points to search for solutions across all domains, which are then evaluated and filtered to obtain top-ranked recommendations. This approach employs SAO semantic analysis rather than keyword-based methods to better reveal relationships between research problems and solutions, and to identify connections among solutions from different domains.

The detailed process is shown in Figure 2 [Figure 2: see original paper]:

**Step 1: Core Technologies and Problems in Target Domain**

- **Data acquisition and preprocessing:** Develop search strategies for the target domain, import patent data into VantagePoint (VP) software, integrate title and abstract information, and perform segmentation to obtain an initial term list. Using thematic clustering methods and after cleaning, merging, and noise removal, generate a high-frequency term list. Incorporate expert opinion to finalize the high-frequency core term list. Cluster these terms based on co-occurrence relationships to identify research themes and hotspots. Simultaneously, import patent data into GoldFire Innovator software to extract SAO structures corresponding to research hotspots.
- **Technology development path and problem analysis:** Analyze the S component in SAO structures to categorize and filter them, presenting current development status as technology roadmaps to identify core technologies and specific problems addressed by research hotspots. These problems are then summarized as search keywords for the next step.

**Step 2: Innovative Solution Selection**

- Use the summarized specific problems as search terms to retrieve initial data from full-domain literature databases.
- Extract SAO structures from this data using GoldFire Innovator software.
- Integrate, clean, and categorize the obtained SAO structures.
- Evaluate solutions from technical feasibility and expected effectiveness perspectives, selecting top-ranked solutions as potential innovative solutions for the target domain.

---

## 3. Case Study: Air Purification Technology

In recent years, "haze" has become a major public concern. According to the 2014 China Environmental Status Bulletin, only 16 of 161 cities met air quality standards under new monitoring criteria. Deteriorating air pollution seriously threatens public health, making air purifiers increasingly popular. With rising living standards and environmental demands, the air purifier market shows promising prospects. This study uses air purification technology to explore innovative solutions.

### 3.1 Core Technologies and Problems in Air Purification

**3.1.1 Data Preprocessing** **(1) Data acquisition:** The search strategy for air purification was: TS=("air clean*" OR "air purifier*"), using Derwent Innovations Index from 2012-2016 (following widespread PM2.5 attention after the U.S. Embassy's 2011 Beijing data release), yielding 9,517 patents.

**(2) Thematic analysis:** This step identifies current research hotspots. - Import raw patent data into VantagePoint, merge titles and abstracts, and perform natural language processing to obtain 120,122 segmented terms. - Using thematic clustering and after cleaning and merging, obtain 2,102 subject terms. With expert input, select 229 high-frequency core terms reflecting the domain's core research. - For visualization, conduct co-occurrence analysis: generate a co-occurrence matrix in VP, import to Gephi to build a network (Figure 3 [Figure 3: see original paper]), where node size reflects betweenness centrality (research 热度) and colors represent modularity classes (relationship strength).

Figure 3 shows that the primary research focus is effective purification (largest nodes). Closely associated nodes include energy consumption and product cost (measuring effectiveness), and purification technologies: activated carbon filters, negative ions, HEPA (High Efficiency Particulate Air) filters, and photocatalysis. Other nodes represent air detection devices (PM2.5 sensors, formaldehyde sensors), volatile organic compounds (VOCs) as key removal targets, and particulate pollutants addressable by HEPA.

The analysis reveals that air purification research centers on improving efficiency through various technologies. Among four key technologies, this study focuses on photocatalysis due to its development prospects.

**(3) SAO structure extraction:** Using GoldFire Innovator software on photocatalysis patents, with "photocatalytic degradation" as the key term to identify solutions (S) and pollutants (O), and verbs (A) like "improve" or "reduce," 236 valid SAO structures were obtained after manual correction.

**3.1.2 Technology Roadmap and Specific Problems** Photocatalysis involves carriers, active components, and co-catalysts. Active components (primarily $TiO_2$) are most critical. Co-catalysts enhance activity without being active themselves. Carriers support these materials.

To detail active material development, we categorize photocatalytic materials into three types: $TiO_2$ structural layers, $TiO_2$-mixed photocatalysts, and non-$TiO_2$ materials (Table 1 ). Figure 4 [Figure 4: see original paper] visualizes the development trajectory, showing: - Anatase $TiO_2$ dominates current research - $TiO_2$ modification (doping with non-metals like carbon or metals like Zn, Cu, Al) enables indoor light activation - $TiO_2$-mixed catalysts (with graphene, ZnO, CuO) are also common - Research continues on $TiO_2$ alternatives, though with limited success

Analysis of SAO structures' O component reveals three problem categories: (1) eliminating bacteria/viruses/fungi, (2) decomposing VOCs (formaldehyde, benzene, ethers), and (3) decomposing inorganic pollutants (CO, NOx, SOx). These problems serve as the basis for cross-domain solution searching.

### 3.2 Innovative Solution Selection Process

**3.2.1 Data Preprocessing** **(1) Data acquisition:** Using the identified problems as keywords, search SCI-EXPANDED (2015-2016, articles/reviews).

Table 2 shows the search strategy, which retrieved relevant literature while excluding air conditioner-related records.

**(2) SAO extraction:** From the retrieved literature, GoldFire Innovator extracted 168 valid SAO structures.

**(3) SAO categorization:** Focusing on $TiO_2$-related photocatalysis, we classify solutions into five categories: $TiO_2$ structure, $TiO_2$-nonmetal doping, $TiO_2$-metal doping, $TiO_2$-metal/nonmetal co-doping, and $TiO_2$ composites (Table 3 ).

**3.2.2 Potential Innovative Solutions for Air Purification Photocatalysis** Solutions are evaluated on two dimensions: technical feasibility (maturity, operability) and expected effectiveness (applicability, purification efficiency, duration) (Table 4 ).

Using expert interviews and scoring (10 experts: 7 questionnaires + 3 interviews), we obtained weights (W) for each indicator and scores (P) for each technology category. The comprehensive score S is calculated as:

$$S_i = \sum_{j=1}^{5} W_j \cdot P_{ij}, \quad i = 1, 2, \dots, 11$$

Table 5 shows the weights and scores. Table 6 presents the top 16 solutions (top two scores per category). Notably, some solutions (marked in bold italics) already appear in air purification, such as P25+Ag+graphene ternary composites also used in wastewater treatment, validating our approach.

The high-scoring solutions primarily come from water purification, indicating strong translatability to air purification and confirming our evaluation metrics' validity. Solutions from other domains may require adaptation for air purification applications.

These potential solutions provide researchers with alternatives and directions, saving literature review time and helping newcomers quickly enter the field.

---

## 4. Discussion and Limitations

**(1)** Due to space constraints, this study focused only on photocatalysis in air purification. Future work could examine other technologies to provide more comprehensive innovative solutions.

**(2)** This study evaluated solutions only on technical feasibility and expected effectiveness. Future research could incorporate economic and social factors for more comprehensive assessment, and employ advanced analytical methods to reduce dependence on expert evaluation.

**(3)** SAO structures were extracted only from titles and abstracts. With more sophisticated software, full-text extraction could yield more comprehensive results.

---

## References

[1] Yoon B, Park I, Coh B-Y. Exploring technological opportunities by linking technology and products: application of morphology analysis and text mining. *Technological Forecasting & Social Change*, 2014, 86: 287-303.

[2] Daim TU, Rueda G, Martin H, et al. Forecasting emerging technologies: use of bibliometrics and patent analysis. *Technological Forecasting & Social Change*, 2006, 73(8): 981-1012.

[3] Chesbrough H. *Open Services Innovation: Rethinking Your Business to Grow and Compete in a New Era.* San Francisco: Jossey-Bass Wiley, 2010.

[4] Zhu D, Porter AL. Automated extraction and visualization of information for technological intelligence and forecasting. *Technological Forecasting & Social Change*, 2002, 69(5): 495-506.

[5] Barker D, Smith DJH. Technology foresight using roadmaps. *Long Range Planning*, 1995, 28(2): 21-28.

[6] Wang XF, Mapp P, Huang Y, et al. Combining SAO semantic analysis and morphology analysis to identify technology opportunities. *Scientometrics*, 2017, 111(1): 3-24.

[7] Yoon J, Park H, Seo W, et al. Technology opportunity discovery (TOD) from existing technologies and products: a function-based TOD framework. *Technological Forecasting & Social Change*, 2015, 100: 153-167.

[8] Kostoff RN. Literature-related discovery and innovation-update. *Technological Forecasting & Social Change*, 2012, 79: 789-802.

[9] Swanson DR. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine*, 1986, 30(1): 7-18.

[10] Tian Ruiqiang, Yao Changqing, Pan Yuntao. Research progress on literature-related knowledge discovery and innovation. *Information Studies: Theory & Application*, 2013, 36(8): 117-123.

[11] Swanson DR. Migraine and magnesium: II. Neglected connections. *Perspectives in Biology and Medicine*, 1988, 31(4): 526-557.

[12] Swanson DR. Somatomedin-C and arginine: Implicit connections. *Perspectives in Biology and Medicine*, 1990, 33(2): 157-186.

[13] Swanson DR, Smalheiser NR. Information discovery from complementary literatures: categorizing viruses as potential weapons. *Journal of the American*

*Society for Information Science & Technology*, 2001, 52(7): 548-557.

[14] Gordon MD, Lindsay RK. Toward discovery support systems: a replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil. *Journal of the American Society for Information Science & Technology*, 1996, 47(2): 116-128.

[15] Weeber M, Klein H, Berg L, et al. Using concepts in literature-based discovery: simulating Swanson's Raynaud-fish oil and migraine-magnesium discoveries. *Journal of the American Society for Information Science & Technology*, 2001, 52(7): 548-557.

[16] Stegmann J, Grohmann G. Hypothesis generation guided by co-word clustering. *Scientometrics*, 2003, 56(1): 111-135.

[17] Kostoff RN. Literature-related discovery (LRD): Methodology. *Technological Forecasting and Social Change*, 2008, 75: 186-202.

[18] Kostoff RN, Patel U. Literature-related discovery and innovation: chronic kidney disease. *Technological Forecasting & Social Change*, 2015, 91: 341-351.

[19] Kostoff RN, Block JA, Stump JA, et al. Literature-related discovery (LRD): potential treatments for Raynaud's phenomenon. *Technological Forecasting & Social Change*, 2008, 75(2): 203-214.

[20] Kostoff RN. Literature-related discovery (LRD): potential treatments for cataracts. *Technological Forecasting and Social Change*, 2008, 75(2): 215-225.

[21] Kostoff RN, Briggs MB. Literature-related discovery (LRD): potential treatments for Parkinson's disease. *Technological Forecasting & Social Change*, 2008, 75(2): 226-238.

[22] Kostoff RN, Briggs MB, Lyons TJ. Literature-related discovery (LRD): potential treatments for multiple sclerosis. *Technological Forecasting & Social Change*, 2008, 75(2): 239-256.

[23] Kostoff RN. Literature-related discovery: potential treatments and preventatives for SARS. *Technological Forecasting & Social Change*, 2011, 78: 1164-1173.

[24] Kostoff RN, Solka JL, Rushenberg RL, et al. Literature-related discovery (LRD): water purification. *Technological Forecasting & Social Change*, 2008, 75(2): 256-275.

[25] Vitavini I, Katsuhide F, Yuya K, et al. Finding linkage between technology and social issues: a literature-based discovery approach. *Journal of the American Society for Information Science & Technology*, 2001, 52(10): 797-812.

[26] Cao Zhijie, Leng Fuhai. Application analysis of non-interactive literature knowledge discovery in aerospace science and technology intelligence research. *Information Studies: Theory & Application*, 2008(4): 569-572.

[27] Huang Shuiqing, Ma Junling. Empirical research on non-interactive literature knowledge discovery in Chinese social science literature: a case study of agricultural economics literature. *Journal of Library Science in China*, 2009(4): 31-38.

[28] Moehrle MG, Walter L, Geritz A, et al. Patent-based inventor profiles as a basis for human resource decisions in research and development. *R&D Management*, 2015, 35(5): 513-524.

[29] Gerken JM, Moehrle MG. A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis. *Scientometrics*, 2012, 91(3): 645-670.

[30] Yoon J, Park H, Kim K. Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. *Scientometrics*, 2013, 94(1): 313-331.

[31] Guo Junfang, Wang Xuefeng, Qiu Pengjun, et al. Research on technology roadmap construction based on SAO analysis. *Studies in Science of Science*, 2014, 32(7): 976-1002.

[32] Wang X, Qiu P, Zhu D, et al. Identification of technology development trends based on subject-action-object analysis: the case of dye-sensitized solar cells. *Technological Forecasting & Social Change*, 2015, 98: 24-46.

[33] Choi S, Yoon J, Kim K, et al. SAO network analysis of patent for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells. *Scientometrics*, 2011, 88(3): 863-883.

[34] Yoon J, Kim K. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics*, 2011, 88(1): 213-228.

[35] Choi S, Park H, Kang D, et al. An SAO-based text mining approach to building a technology tree for technology planning. *Expert Systems with Applications*, 2012, 39(13): 11443-11455.

[36] Wen Liang, Qiu Pengjun, Ma Pingping, et al. Identifying potential technology collaboration partners based on SAO semantic analysis. *Journal of Beijing Institute of Technology (Social Sciences Edition)*, 2017(4): 91-96.

[37] Bergmann I, Butzke D, Walter L, et al. Evaluating the risk of patent infringement by means of semantic patent analysis: the case of DNA chips. *R&D Management*, 2008, 38(5): 550-562.

[38] Park H, Yoon J, Kim K. Identifying patent infringement using SAO-based semantic technological similarities. *Scientometrics*, 2012, 90(2): 515-529.

[39] Choi S, Kang D, Lim J, et al. A fact-oriented ontological approach to SAO-based function modeling of patents for implementing function-based technology database. *Expert Systems with Applications*, 2012, 39(10): 9129-9140.

[40] Wu Feifei, Li Qian, Huang Lucheng. Research on technology application field identification method based on patent SAO structure. *Science Research Management*, 2014, 35(6): 1-7.

[41] Wang Xuefeng, Fu Yun, Qiu Pengjun, et al. Research on R&D partner identification based on SAO analysis. *Science Research Management*, 2015, 36(10): 19-27.

[42] Zhang Yi, Wang Xuefeng, Zhu Donghua, et al. Research on "thematic clustering" method: cleaning, merging and clustering of English scientific literature subject terms. *Studies in Science of Science*, 2013(11): 1615-1622.

[43] Vicente-Gomila JM. The contribution of syntactic-semantic approach to the search for complementary literatures for scientific or technical discovery. *Scientometrics*, 2014, 100(3): 659-677.

[44] Liu Hui. Study on visible light-responsive photocatalysts and their water splitting applications [D]. Shanghai: Shanghai Jiao Tong University, 2008.

## Author Contributions

**Fu Yun:** Conceptualized research, designed methodology, collected and analyzed case data, wrote manuscript.

**Wang Xuefeng:** Determined research direction, designed methodology, revised manuscript (corresponding author).

**Li Jia:** Participated in case data processing.

**Hou Yujia:** Participated in case data processing.

## English Title and Abstract

### Research on Selection of Innovative Solutions Based on SAO Structure: A Case Study on Air Purification Technology

**Abstract:** [Purpose/significance] Facing the increasingly competitive social environment, innovation is the foundation and way of existence for enterprises. The scope of innovation includes not only creating new products and technologies in the target research field, but also introducing new technologies and products from other research fields into the target field, with the latter being much easier to accomplish. However, with the increasingly high degree of specialization in every disciplinary field, researchers have little time to grasp knowledge besides their own research field. Therefore, it is necessary to use scientific methods and technology to explore the deep relationships between knowledge from different research fields. [Method/process] Using the analytical process of LRDI methodology, this paper proposes research on selection of innovative solutions based on SAO structure, seeking potential solutions in the whole research field based on specific problems from the target research field. The paper evaluates these potential solutions from aspects of technical feasibility and expected results, and gives priority to recommend the solutions as innovative solutions for the target research field. An exploratory study is conducted on air purification technology for this systematic process. [Result/conclusion] The research shows that some of the selected innovative solutions have been effectively used in the air purification field, which also verifies the proposed research method is feasible and valid.

**Keywords:** innovative solutions; SAO structure; LRDI; technology roadmap; air purification technology

*Note: Figure translations are in progress. See original paper for figures.*